International Journal of Forecasting 7 (1991) 199–208
North-Holland

Seasonality, non-stationarity and the forecasting of monthly time series

Philip Hans Franses *

Econometric Institute and Tinbergen Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, Netherlands

Abstract: We focus on two forecasting models for a monthly time series. The first model requires that the variable is first order and seasonally differenced. The second model considers the series only in its first differences, while seasonality is modeled with a constant and seasonal dummies. A method to distinguish empirically between these two models is presented. The relevance of this method is established by simulation results as well as empirical evidence, which show, first, that conventional autocorrelation checks are often not discriminative and, second, that considering the first model while the second is more appropriate yields a deterioration of forecasting performance.

Keywords: Monthly time series, Non-stationarity, Seasonality, Seasonal unit roots, Seasonal differencing, Forecasting performance.

1. Introduction and summary

In this paper the focus is on two forecasting models for monthly time series. The first is the well-known multiplicative seasonal model advocated by Box and Jenkins (1970), which requires that the variable is transformed to annual differences of the monthly growth rates. The second is an autoregressive-moving average model for the variable in its first differences, in which seasonality is modeled with a constant and 11 seasonal dummy variables. The primary motive of the present study is the observation that the forecasts for the number of airline passengers from the first model, as it is applied in box and Jenkins (1970), are all too high. This might indicate that the model may be misspecified. We will argue here, on the basis of simulation results and of empirical evidence, that this can be caused by considering the first model while the second would have been more appropriate. It will be shown that the conventional autocorrelation checks are often not discriminative, but that the method described in Franses (1990), which is an extension of the one in Hylleberg et al. (1990), allows to distinguish empirically between the two models.

In Section 2, the two competing forecasting models will be introduced, and a small simulation experiment will illustrate the impact on forecasting of using one model while the alternative is correct. In Section 3, a brief account is given of a method to test for seasonal unit roots in monthly data, being a method to choose between the models. It will be applied to three empirical series, one of which is the aforementioned airline data. In Section 4, both forecasting schemes will be used for the three series. From an extensive forecasting performance evaluation it will emerge that indeed the first model yields far worse results when the second model is appropriate. In Section 5, some concluding remarks will be given.

* Helpful comments from an Associate Editor, and from Lourens Broersma, Teun Kloek and Erno Kuiper are gratefully acknowledged. Copies of unpublished papers in the list of references are available from the author.

2. Two forecasting models for monthly time series

Consider the following forecasting models for a monthly time series y_t. The first is the multiplicative seasonal model, to be denoted as MSBJ in the sequel, which is advocated in Box and Jenkins (1970) and which is often used in practice, or

$$\Delta_1\Delta_{12}y_t = \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-12} + \beta_3\varepsilon_{t-13}, \qquad (1)$$

where

$$\Delta_k y_t \equiv (1 - B^k)y_t \equiv y_t - y_{t-k},$$

and where ε_t is assumed to be a white noise process with

$$E(\varepsilon_t) = 0, \qquad E(\varepsilon_t^2) = \sigma^2,$$

and

$$E(\varepsilon_s\varepsilon_t) = 0 \quad \text{for } s \neq t.$$

This interpretation for ε_t will be used throughout the paper. Arguments to be discussed below may naturally apply to more complicated autoregressive-moving average models for $\Delta_1\Delta_{12}y_t$, but eq. (1) suffices for the present purposes.

The second model consists of an autoregressive-moving average model for the variable y_t in first differences, a constant and 11 seasonal dummy variables, or

$$\phi_p(B)\Delta_1 y_t = \alpha_0 + \sum_{i=1}^{11}\alpha_i D_{it} + \theta_q(B)\varepsilon_t, \qquad (2)$$

where D_{it} are seasonal dummies with a "1" in the corresponding month, and a "0" in other months, with D_{1t} representing January, etc. The $\phi_p(B)$ and $\theta_q(B)$ are polynomials in the backward shift operator B, for which the usual assumptions apply (see, e.g., Granger and Newbold, 1986). In the sequel, model (2) with deterministic seasonality will be labeled the FDSD model.

The MSBJ model is often used in forecasting exercises. A phenomenon which is sometimes encountered in practice is that its forecasts may all be too low or too high – see, e.g., the example of forecasting the number of airline passengers in Box and Jenkins (1970), where all 36 monthly forecasts are too high. This may suggest that model (1) is misspecified. This may be caused by the fact that the appropriate model for y_t is eq. (2), while using eq. (1) results in overdifferencing and misspecification. Transforming a series with the $\Delta_1\Delta_{12}$ filter assumes the presence of 13 roots on the unit circle (see also eq. (4) below), two of which are at the zero frequency. Hence, in case only the Δ_1 filter is sufficient to remove non-stationarity, the incorrect assumption of the presence of the other roots implies overdifferencing. The misspecification originates from treating deterministic seasonality incorrectly as being stochastic. In Osborn (1990) it is empirically demonstrated that this type of misspecification often occurs. In Section 3, a procedure will be described to test for the presence of unit roots in monthly data. Next, we will show with a small experiment that using the MSBJ model while the FDSD model is the appropriate data generating process may indeed explain the observed empirical forecast error patterns, although the usual autocorrelation checks often do not cause alarm.

For an artificial sample, ranging from 1950.01 to 1970.12, observations on y_t are generated from the model

$$y_t = y_{t-1} + \alpha_0 + \sum_{i=1}^{11}\alpha_i D_{it} + 0.3\Delta_1 y_{t-1} + \varepsilon_t - 0.6\varepsilon_{t-1}, \qquad (3)$$

where, in case (a), the α_0 through α_{11} have been set equal to −1, −4, −3, −1, 2, 5, 7, 9, 4, 2, 1, −2, yielding a time series resembling the airline data and, in case (b), the α's are −1, −1, 1, 2, 3, −5, 6, 8, −6, 4, 2, −2. Furthermore, ε_t is drawn from a standard normal distribution, and $y_0 = 0$ and $y_1 = 0$. From this large sample, the first eight years are deleted to reduce starting-up effects, and the last three years will be used for out-of-sample forecasting. To the remaining 120 observations, model (1) is fitted, after which the residuals are checked for autocorrelation with the usual portmanteau test statistic (see Box and Jenkins, 1970; Granger and Newbold, 1986). This exercise has been carried out for 100 replications, where all calculations have been performed with TSP version 6.53 (1989). The results for the autocorrelation tests are summarized in Exhibit 1.

Suppose that a 10% level of significance is used, and also that the strategy is adopted that models where too much autocorrelation is left in the residuals will not be used in a forecast evaluation, for they are already misspecified; then it can be seen that for cases (a) and (b) there remain 69 and 64

Exhibit 1
Number of times the null hypothesis of no autocorrelation is rejected when an MSBJ model is fitted to observations generated by an FDSD model (based on 100 simulations).

Case	Size	Test statistic [a]	
		BP(12)	BP(24)
(a)	0.05	26	17
	0.10	31	22
(b)	0.05	26	13
	0.10	36	17

[a] The Box–Pierce test statistic for autocorrelation of order 12 and 24. Under the null it is χ^2 distributed with 9 and 21 degrees of freedom, respectively.

replications for forecasting exercises, respectively. For each of these repetitions, forecasts for 36 months out-of-sample are calculated and compared with the true observations. Denoting M as the number of times that the true value exceeds the forecasted value, the distributions of M are given in Exhibit 2(a), (b). In the ideal situation, one would theoretically expect that M is symmetrically distributed with mean 18 and with standard deviation equal to 3. Or, it would be expected that about 95% of the observations is within the interval 12–24.

From Exhibit 2 it is obvious that this situation

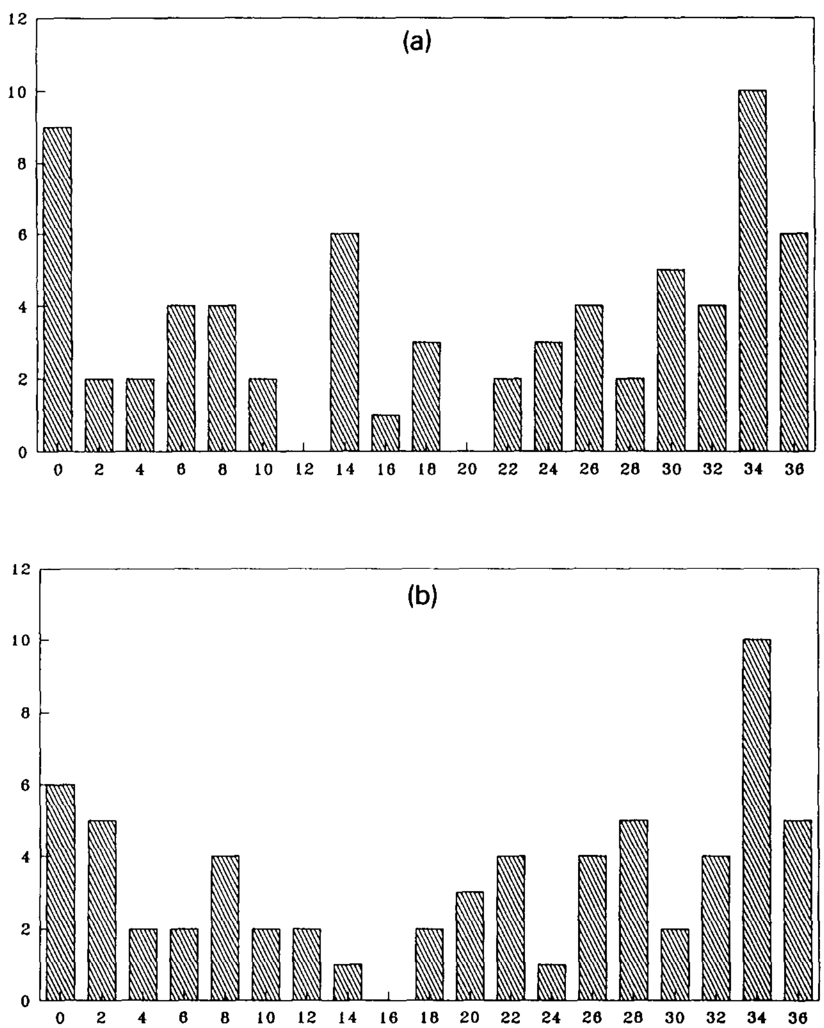

Exhibit 2. Forecast performance evaluation (horizon = 36) of MSBJ model when an FDSD is the data generating process (based on 69 and 64 simulations for case (a) and (b), respectively); the columns show M, the number of times the true value exceeds the forecasted value.

is certainly not the case here. Furtheremore, it can be seen that the forecasts can be too high or too low about equally well.

These simulation experiments strongly suggest that considering the incorrect model can yield biased forecasts. Furthermore, it emerges that the usual specification checks are often not discriminative enough to reject this incorrect model. This calls for a method to empirically distinguish between the MSBJ and the FDSD model, which will be briefly described in the next section.

3. Testing for seasonal unit roots

The differencing operator Δ_{12} assumes the presence of 12 roots on the unit circle, which becomes clear from noting that

$$\begin{aligned}1 - B^{12} &= (1-B)(1+B)(1-iB)(1+iB)\\ &\times\left[1+(\sqrt{3}+i)B/2\right]\left[1+(\sqrt{3}-i)B/2\right]\\ &\times\left[1-(\sqrt{3}+i)B/2\right]\left[1-(\sqrt{3}-i)B/2\right]\\ &\times\left[1+(i\sqrt{3}+1)B/2\right]\left[1-(i\sqrt{3}-1)B/2\right]\\ &\times\left[1-(i\sqrt{3}+1)B/2\right]\left[1+(i\sqrt{3}-1)B/2\right],\end{aligned} \tag{4}$$

where all terms other than $(1-B)$ correspond to seasonal unit roots. In Hylleberg et al. (1990), a method has been developed for testing for the presence of seasonal unit roots in quarterly data. In Franses (1990), this method has been extended to time series consisting of monthly observations. To save space only the final test equation will be presented to ensure that the reader can verify some of the claims made here.

Testing for unit roots in monthly time series is equivalent to testing for the significance of the parameters in the auxiliary regression

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

where $\varphi^*(B)$ is some polynomial function of B for which the usual assumption applies, and where

$$\begin{aligned}y_{1,t} &= (1+B)(1+B^2)(1+B^4+B^8)y_t,\\ y_{2,t} &= -(1-B)(1+B^2)(1+B^4+B^8)y_t,\\ y_{3,t} &= -(1-B^2)(1+B^4+B^8)y_t,\\ y_{4,t} &= -(1-B^4)(1-\sqrt{3}B+B^2)(1+B^2+B^4)y_t,\\ y_{5,t} &= -(1-B^4)(1+\sqrt{3}B+B^2)(1+B^2+B^4)y_t,\\ y_{6,t} &= -(1-B^4)(1-B^2+B^4)(1-B+B^2)y_t,\\ y_{7,t} &= -(1-B^4)(1-B^2+B^4)(1+B+B^2)y,\\ y_{8,t} &= (1-B^{12})y_t.\end{aligned}$$

Furthermore, the μ_t in eq. (5) covers the deterministic part and might consist of a constant, seasonal dummies, or a trend. This depends on the hypothesized alternative to the null hypothesis of 12 unit roots.

Applying ordinary least squares to eq. (5) gives estimates of the π_i. In case there are (seasonal) unit roots, the corresponding π_i are zero. Due to the fact that pairs of complex unit roots are conjugates, it should be noted that these roots are only present when pairs of π's are equal to zero simultaneously, for example the roots i and $-i$ are only present when π_3 and π_4 are equal to zero (see Franses, 1990, for detailed derivations). There will be no seasonal unit roots if π_2 through π_{12} are significantly different from zero. If $\pi_1 = 0$, then the presence of root 1 can not be rejected. When $\pi_1 = 0$, π_2 through π_{12} are unequal to zero, and when, additionally, seasonality can be modeled with seasonal dummies, an FDSD model as in eq. (2) may emerge. In case all π_i, $i = 1,\dots,12$, are equal to zero, it is appropriate to apply the Δ_{12} filter, and hence the MSBJ model may be useful. Extensive tables with critical values for t-tests of the separate π's, and for F-tests of pairs of π's, as well as for a joint F-test of $\pi_3 = \cdots = \pi_{12}$ can be found in Franses (1990). Some critical values which will be of relevance later in this section are given in Exhibit 3.

In Beaulieu and Miron (1990), the Hylleberg et al. (1990) procedure is also extended to monthly data, but their test equation differs from eq. (5) and is somewhat more complicated. Furthermore, the authors do not consider the useful joint F-test for the presence of the complex unit roots.

Exhibit 3
Tables with critical values. Some critical values for testing for seasonal unit roots in monthly data; based on 5000 Monte Carlo simulations, DGP: $y = y(-12) + \varepsilon$, $\varepsilon \sim N(0, 1)$; number of observations = 120. [a]

	Axiliary regression							
	Constant, dummies and trend				Constant, dummies and no trend			
t-statistics	0.05		0.10		0.05		0.10	
π_1	−3.24		−2.92		−2.63		−2.35	
π_2	−2.65		−2.39		−2.65		−2.40	
t-statistics	0.025	0.05	0.95	0.975	0.025	0.05	0.95	0.975
π_3	−2.05	−1.71	1.72	2.10	−2.11	−1.76	1.74	2.11
π_4	−3.34	−3.12	−0.45	−0.15	−3.34	−3.12	−0.44	−0.14
π_5	−3.29	−2.99	−0.06	0.24	−3.29	−3.00	−0.05	0.25
π_6	−3.38	−3.12	−0.44	−0.11	−3.39	−3.12	−0.42	−0.09
π_7	−0.18	0.12	2.98	3.28	−0.27	0.05	3.00	3.31
π_8	−3.40	−3.15	−0.43	−0.17	−3.39	−3.14	−0.42	−0.18
π_9	−2.86	−2.54	0.81	1.12	−2.87	−2.54	0.82	1.13
π_{10}	−3.36	−3.07	−0.40	−0.09	−3.37	−3.07	−0.39	−0.07
π_{11}	−1.08	−0.73	2.55	2.80	−1.11	−0.78	2.56	2.83
π_{12}	−3.42	−3.16	−0.44	−0.17	−3.43	−3.16	−0.42	−0.14
F-statistics	0.90	0.95			0.90	0.95		
π_3, π_4	4.81	5.63			4.83	5.62		
π_5, π_6	4.86	5.84			4.89	5.86		
π_7, π_8	4.94	5.90			4.94	5.86		
π_9, π_{10}	4.76	5.71			4.79	5.75		
π_{11}, π_{12}	4.92	5.84			4.94	5.89		
$\pi_3, \ldots, \pi_{12}$	4.00	4.45			4.00	4.46		

[a] *Source*: Franses (1990, pp. 12–18). Note that the tests for π_1 and π_2 are one-sided tests, while the other t-tests are two-sided.

Exhibit 4a
Index of industrial production (the Netherlands, 1980 = 100). [a]

Month	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978
Jan.	64	71	79	81	88	93	88	95	98	98
Feb.	67	74	80	83	92	98	93	98	100	102
Mar.	68	77	80	84	89	98	95	98	100	101
Apr.	68	77	80	87	92	95	93	99	102	100
May	68	76	80	83	89	95	88	95	96	94
June	68	74	79	82	89	95	88	96	96	95
July	59	65	66	68	74	77	70	76	78	78
Aug.	63	69	74	77	82	87	77	85	83	83
Sept.	68	74	80	83	91	94	87	98	95	95
Oct.	73	81	86	89	96	99	93	100	99	101
Nov.	78	83	86	92	99	101	100	103	102	106
Dec.	77	81	83	92	98	95	102	111	110	116

Month	1979	1980	1981	1982	1983	1984	1985	1986	1987
Jan.	108	111	105	104	97	108	118	112	118
Feb.	109	107	113	104	106	111	118	121	118
Mar.	106	111	103	102	102	110	115	112	118
Apr.	107	105	103	101	100	105	107	113	108
May	98	98	94	92	96	98	102	99	104
June	95	95	92	90	91	97	102	101	102
July	78	76	78	75	77	80	82	85	87
Aug.	83	81	78	75	77	85	87	88	87
Sept.	98	90	90	89	91	96	97	102	98
Oct.	104	101	101	94	97	102	106	108	109
Nov.	112	111	105	98	105	108	120	114	118
Dec.	112	114	114	107	113	110	112	115	114

[a] *Source*: OECD Main Economic Indicators.

Exhibit 4b
New car registrations (the Netherlands). [a]

Month	1978	1979	1980	1981	1982	1983
Jan.	65624	61720	74619	51368	43477	57005
Feb.	39004	41875	39920	35811	32975	33851
Mar.	55928	75989	45404	44507	45435	57053
Apr.	51089	62938	45791	39362	45751	47870
May	53920	54831	42023	41392	40067	43041
June	73526	51197	38875	37099	39455	49482
July	35328	37123	30909	31839	31074	33993
Aug.	33756	34858	27308	21659	23562	26720
Sept.	43344	32165	29279	24936	28074	33377
Oct.	70418	45347	33437	28098	34313	35261
Nov.	48249	38598	26084	21765	28240	27193
Dec.	14400	15962	11184	8947	10680	11508
Month	1984	1985	1986	1987	1988	
Jan.	68662	62079	76975	85519	89929	
Feb.	40007	39134	44701	42154	33771	
Mar.	53149	58685	56175	61224	52082	
Apr.	46193	53148	58748	62051	47504	
May	50648	49239	56614	53501	42885	
June	39593	44575	55460	51869	45786	
July	28684	36319	40472	42020	32933	
Aug.	27584	33753	35076	31038	28803	
Sept.	30296	33331	46107	38041	35323	
oct.	37899	40673	46667	42331	34216	
Nov.	29316	30695	30756	29119	28067	
Dec.	9360	14089	11084	15436	9350	

[a] *Source*: Central Bureau of Statistics and RAI.

The method given in eq. (5) to test for seasonal unit roots is applied to the first nine years of the airline data, *lnp*, as they are given in Box and Jenkins (1970, p. 304). Two other montly series, which are an index for industrial production and new car registrations, are also considered. The observations are displayed in Exhibit 4. In the sequel, both series will be measured in natural

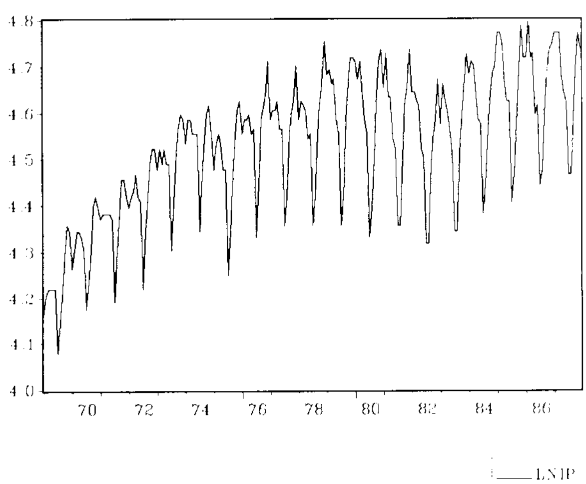

Exhibit 5. Natural logarithms of industrial production index (the Netherlands, 1969.01–1987.12, 1980 = 100).

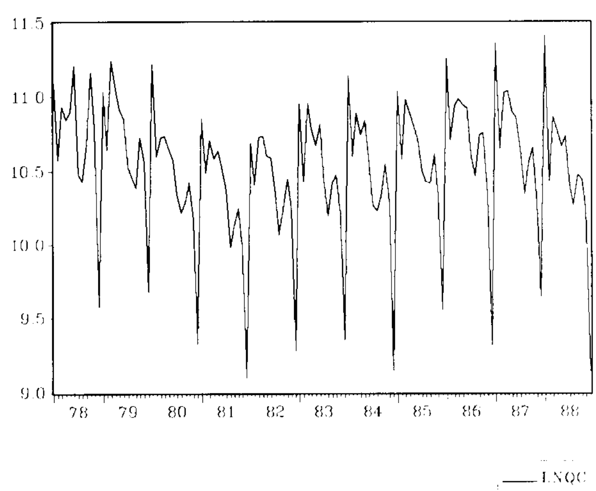

Exhibit 6. Natural logarithms of new car registrations (the Netherlands, 1978.01–1988.12).

Exhibit 7
Testing for (seasonal) unit roots.

t-statistics	Variable		
	lnp [a]	*lnip* [b]	*lnqc* [c]
π_1	−2.253	−2.471	−2.035
π_2	−2.984 **	−3.360 **	−2.638 *
π_3	−2.715 **	−2.053 *	−3.537 **
π_4	−2.329	−4.800 **	−2.943
π_5	−2.973	−3.786 **	−2.861
π_6	−3.881 **	−3.825 **	−3.292 **
π_7	0.933	−0.063 *	1.969
π_8	−2.086	−1.529	−3.454 **
π_9	−1.332	−2.338	−1.383
π_{10}	−3.626 **	−3.789 **	−2.880
π_{11}	−1.331 **	−2.577 **	−0.265
π_{12}	−2.085	−3.455 **	−3.221 *
F-statistics	*lnp*	*lnip*	*lnqc*
π_3, π_4	7.028 **	14.318 **	11.951 **
π_5, π_6	7.895 **	7.814 **	5.423 *
π_7, π_8	4.940 *	5.424 *	10.698 **
π_9, π_{10}	6.864 **	7.329 **	4.150
π_{11}, π_{12}	7.206 **	22.461 **	8.646 **
$\pi_3, \ldots, \pi_{12}$	15.348 **	24.965 **	16.083 **

[a] The auxiliary regression contains constant, trend and seasonal dummies, while $\varphi^*(B)$ is $(1-\varphi_1 B^{12})$ and the number of observations equals 84.
[b] The axiliary regression contains constant, trend and seasonal dummies, while $\varphi^*(B)$ is 1 and the number of observations equals 180.
[c] The auxiliary regression contains constant and seasonal dummies, while $\varphi^*(B)$ is 1 and the number of observations is 84.
* Significant at 10% level.
** Significant at 5% level.

logarithms. Graphs of *lnip* and *lnqc* are given in Exhibits 5 and 6.

The last 36 observations are again not used, for they will be used for forecast evaluation. From Exhibits 5 and 6, and from the graph in Box and Jenkins (1970, p. 308) it is clear that the alternatives for non-stationary stochastic seasonality, necessitating the use of a Δ_{12} filter, may be a deterministic seasonal pattern and, additionally, a trend for *lnp* and *lnip*. The test results are displayed in Exhibit 7.

Simulation evidence in Franses (1990) shows that the power of the test statistics may be low, except for the joint F-test for all complex π_i, and hence that significance levels of 10%, or even higher, may be more appropriate. Considering the results in Exhibit 7, it seems that the general result is that seasonality and non-stationarity in the three time series can be appropriately modeled with an FDSD model as in eq. (2), although the evidence for *lnqc* is not overwhelming. Anyhow, the regularly applied Δ_{12} filter, not to mention the $\Delta_1\Delta_{12}$ filter, is certainly not appropriate. This corresponds to the results in Beaulieu and Miron (1990), and also in Osborn (1990) similar findings for quarterly data are reported.

4. Forecasting

Now the type of seasonality and non-stationarity has been established, several FDSD models for *lnp*, *lnip*, and *lnqc* can be built. The models, which have been found after a brief specification search, are given in Exhibit 8, together with their estimation results and some evaluation criteria. The statistical package used is TSP version 6.53 (1989), and the estimation method is iterative least squares.

From Exhibit 8 it is obvious that the FDSD type of model gives a fairly good representation of the data for all three variables. Most parameters for the seasonal dummies are highly significant, the adjusted coefficients of determination are high and the checks on autocorrelation do not provide strong arguments to suspect misspecification.

The estimation and evaluation results of models of type (1), which will be the competitors in the forecasting exercises below, are displayed in Exhibit 9. These models also show significant estimated parameters and no significant residual autocorrelation. Hence, on the basis of these criteria, the choice for an MSBJ model might be defended.

To evaluate the FDSD and MSBJ models in Exhibits 8 and 9 with respect to their forecasting performance, forecasts for 36 months out-of-sample are generated from each of these models. The values of several forecast evaluation criteria are given in Exhibit 10.

A test to investigate whether there are significant differences between the forecasts is the Wilcoxon signed-rank test (see, e.g., Flores, 1989). The results for this test indicate that there are statistically significant differences indeed. The general result with respect to the criteria *ME* through Theil's U-statistic seems to be that the FDSD model outperforms the MSBJ model. It is also clear that for *lnip* and *lnqc* the numbers of positive forecast errors M from using an FDSD model

Exhibit 8
Estimation results of models for $\Delta_1 lnp$, $\Delta_1 lnip$, and $\Delta_1 lnqc$.

Model variables [a]	Dependent variable					
	$\Delta_1 lnp$		$\Delta_1 lnip$		$\Delta_1 lnqc$	
C	0.097 **	(0.017)	0.018 **	(0.007)	−0.851 **	(0.077)
D_1	−0.038	(0.032)	−0.056 **	(0.012)	2.859 **	(0.184)
D_2	−0.092 **	(0.023)	0.001	(0.010)	−0.250	(0.473)
D_3	0.051 **	(0.021)	−0.023 **	(0.010)	1.378 **	(0.077)
D_4	−0.088 **	(0.032)	−0.022 **	(0.010)	0.607 **	(0.163)
D_5	−0.109 **	(0.019)	−0.051 **	(0.011)	0.846 **	(0.072)
D_6	0.032	(0.021)	−0.026 **	(0.010)	0.847 **	(0.080)
D_7	0.044	(0.031)	−0.137 **	(0.017)	0.547 **	(0.084)
D_8	−0.697 **	(0.029)	0.025 **	(0.010)	0.823 **	(0.063)
D_9	−0.211 **	(0.021)	0.057 **	(0.012)	1.021 **	(0.065)
D_{10}	−0.260 **	(0.017)	0.023 **	(0.010)	1.029 **	(0.108)
D_{11}	−0.263 **	(0.017)	0.009	(0.010)	0.510 **	(0.132)
AR_1	−0.273 **	(0.099)			0.396	(0.248)
AR_{12}			0.388 **	(0.064)		
MA_1			−0.401 **	(0.078)	−0.815 **	(0.274)
MA_4			−0.216 **	(0.079)		
Evaluation criteria [b]	$\Delta_1 lnp$		$\Delta_1 lnip$		$\Delta_1 lnqc$	
BP(12)	9.293		7.849		9.925	
BP(24)	22.049		30.363		23.377	
$\bar{R}^2$	0.887		0.894		0.957	

[a] The model contains a constant C, 11 seasonal dummies, $D_1, \ldots, D_{11}$, where D_1 corresponds to Jan., autoregressive terms at la, p, AR_p, and moving average terms at lag q, MA_q.
[b] The evaluation criteria are the Box–Pierce portmanteau test statics, calculated for m lags. Under the null hypothesis, this BP(m) follows a χ^2 distribution with $m - r$ degrees of freedom, where r is the sum of the number of autoregressive and moving average parameters. R^2 denotes the adjusted coefficient of determination.
** Significant at 5% level. Standard deviations are given in parentheses.

are close to what might have been expected, while those when using an MSBJ model are out of any reasonable range. These empirical results for M seem to confirm the simulation evidence in Section 2. From the results of the Wilcoxon rank sum test for squared errors and percentage errors, of a sign test, and of a percentage improvement measure, it appears that most differences between the models are significant and are in favor of the FDSD model. However, for the airline series the dif-

Exhibit 9
Estimation results of models for $\Delta_1\Delta_{12} lnp$, $\Delta_1\Delta_{12} lnip$ and $\Delta_1\Delta_{12} lnqc$.

Model variables [a]	Dependent variable					
	$\Delta_1\Delta_{12} lnp$		$\Delta_1\Delta_{12} lnip$		$\Delta_1\Delta_{12} lnqc$	
MA_1	−0.338 **	(0.104)	−0.436 **	(0.076)	−0.337 **	(0.113)
MA_{12}	−0.715 **	(0.104)	−0.571 **	(0.078)	−0.733 **	(0.103)
MA_{13}	0.322 **	(0.104)	0.363 **	(0.078)	0.359 **	(0.103)
Evaluation criteria [b]	$\Delta_1\Delta_{12} lnp$		$\Delta_1\Delta_{12} lnip$		$\Delta_1\Delta_{12} lnqc$	
BP(12)	6.606		8.813		9.661	
BP(24)	15.325		20.185		15.848	
$\bar{R}^2$	0.415		0.388		0.447	

[a] The model contains moving average terms at lag q, MA_q.
[b] The evaluation criteria are the Box–Pierce portmanteau test statistic, calculated for m lags. Under the null this BP(m) follows a χ^2 distribution with $m - r$ degrees of freedom, where r is the sum of the number of autoregressive and moving average parameters. $\bar{R}^2$ denotes the adjusted coefficient of determination.
** Significant at 5% level. Standard deviations are given in parentheses.

Exhibit 10
Evaluation of the 36 months out-of-sample forecasting performance of models for the variables *lnp*, *lnip*, and *lnqc*.

Criterion [a]	*lnp*		*lnip*		*lnqc*	
	MSBJ	FDSD	MSBJ	FDSD	MSBJ	FDSD
ME	−0.074	−0.064	0.042	0.022	−0.200	0.039
MAE	0.074	0.067	0.044	0.033	0.221	0.124
maxAE	0.179	0.196	0.109	0.112	0.607	0.346
minAE	0.003	0.000 [b]	0.002	0.000 [b]	0.009	0.017
MAPE	1.229	1.116	0.942	0.691	2.117	1.171
MSE	0.007	0.006	0.003	0.002	0.079	0.022
RMSE	0.081	0.079	0.051	0.044	0.280	0.148
M	0	4	33	24	5	22
U (×100)	1.339	1.303	1.102	0.943	2.641	1.392
PIMSE of FDSD	5.402		24.17		72.22	
SIGNSE	25 *		24 *		22	
Signed Rank	2.781 *		−4.305 *		5.232 *	
Rank Sum SE	0.248		2.106 *		3.221 *	
Rank Sum PE	−0.158		2.117 *		1.994 *	

[a] The forecast error is defined as the true value y minus the forecasted value f. Forecast evaluation criteria are the mean error, *ME*, mean absolute error, *MAE*, maximum and minimum value of absolute error, *maxAE* and *minAE*, mean average percentage error, *MAP*, and (root) mean squared error, *(R)MSE*. *M* denotes the number of times y exceeds f. *U* is Theil's test statistic. *PIMSE* denotes the precentage improvement of forecasts from the FDSD model with respect to mean squared error. *SIGNSE* refers to the sign test which reports the number of times the squared error of FDSD is smaller than that of MSBJ in pairwise comparison. The Wilcoxon *signed rank* test statistic refers to the ranks of positive differences between the forecasts. *Rank Sum* refers to the Wilcoxon test for differences in forecast performance with respect to squared error *SE* or to percentage error *PE*. Positive values for this test indicate that the FDSD model is better. Definitions and asymptotic results for the Wilcoxon tests can be found in Lehmann (1975).

[b] The rounded value is smaller than 0.001.

* Significant at 5% level.

ferences in forecasting performance between the MSBJ and the FDSD model are not that striking, although some forecasting improvement can be witnessed.

5. Concluding remarks

In this paper it has been shown that correctly taking account of the type of seasonality and non-stationarity in monthly data can improve forecasting performance. This is illustrated for the case where a moving average model is fitted to a first order and seasonally differenced variable, while an autoregressive-moving average model for the first order differenced variable together with the inclusion of a constant and seasonal dummies would have been more appropriate. A method to choose empirically between these models is also given. Of course, these results may naturally be extended to time series consisting of quarterly observations, and those which contain deterministic trends instead of stochastic trends.

The major result of the present paper is that the recognition of the presence, or better, of the absence of seasonal unit roots can have important implications for forecasting and model building. Recent additional arguments for not automatically doubly differencing a seasonal variable can be found in Bodo and Signorini (1987), where econometric models with seasonal dummies also yield better forecasts, and in Heuts and Bronckers (1988), where doubly differencing the same production index as above makes that this variable shows no correlation with other variables.

References

Beaulieu, J.J. and J.A. Miron, 1990, "Seasonal unit roots and deterministic seasonals in aggregate U.S. data", Presented at the *C.R.D.E./Journal of Econometrics Conference on Seasonality and Econometric Models*, Montreal, May.

Bodo, G. and L.F. Signorini, 1987, "Short-term forecasting of the industrial production index", *International Journal of Forecasting*, 3, 245–259.

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

Flores, B.E., 1989, "The utilization of the Wilcoxon test to compare forecasting methods: A note", *International Journal of Forecasting*, 5, 529–535.

Franses, P.H., 1990, "Testing for seasonal unit roots in monthly data", Econometric Institute Report #9032/A, Erasmus University Rotterdam.

Granger, C.W.J. and P. Newbold, 1986, *Forecasting Economic Time Series*, 2nd ed. (Academic Press, San Diego, CA).

Heuts, R.M.J. and J.H.J.M. Bronckers, 1988, "Forecasting the Dutch heavy truck market: A multivariate approach", *International Journal of Forecasting*, 4, 57–79.

Hylleberg, S., R.F. Engle, C.W.J. Granger and B.S. Yoo, 1990, "Seasonal integration and cointegration", *Journal of Econometrics*, 44, 215–238.

Lehmann, E.L., 1975, *Nonparametrics: Statistical Methods Based on Ranks* (Holden-Day, San Francisco, CA).

Osborn, D.R., 1990, "A survey of seasonality in UK macroeconomic variables", *International Journal of Forecasting*, 6, 327–336.

Biography: Philip Hans FRANSES is affiliated with the Econometric Institute and Tinbergen Institute of the Erasmus University Rotterdam, the Netherlands. He obtained his M.Sc. degree in econometrics at the University of Groningen. He has published articles in *Econometric Reviews*, *Journal of Marketing Research*, and other journals.