Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: http://www.tandfonline.com/loi/ubes20

A Note on Overdifferencing and the Equivalence of Seasonal Time Series Models With Monthly Means and Models With $(0, 1, 1)_{12}$ Seasonal Parts When $\Theta = 1$

William Bell

Journal of Business & Economic Statistics, July 1987, Vol. 5, No. 3

A Note on Overdifferencing and the Equivalence of Seasonal Time Series Models With Monthly Means and Models With $(0, 1, 1)_{12}$ Seasonal Parts When $\Theta = 1$

William Bell
Statistical Research Division, U.S. Bureau of the Census, Washington, DC 20233

Two general models for monthly seasonal time series are considered, one in which seasonality is modeled with monthly means and another in which seasonality is modeled with a $(0, 1, 1)_{12}$ ARIMA structure. The models are shown to be equivalent if the seasonal moving average parameter (Θ) is 1 and if the same assumptions about the 12 initial observations are made for both models. The role of the assumptions about the initial observations is analyzed, and it is argued that for practical purposes the two models can be regarded as equivalent when $\Theta = 1$. It is observed that the result extends easily to more general models involving overdifferencing.

KEY WORDS: Seasonal ARIMA model; Starting values.

1. INTRODUCTION

Consider the following two models for a monthly time series z_t:

Model 1. $z_t = \Sigma_1^{12} \alpha_i M_{it} + u_t$.

Model 2. $(1 - B^{12})z_t = w_t = (1 - \Theta B^{12})u_t$.

(I use monthly time series for concreteness. All that follows applies immediately to time series with other seasonal periods. Extensions to more general models will also be discussed.) B is the backshift operator ($B^j z_t = z_{t-j}$); $\alpha_1, \ldots, \alpha_{12}$ and Θ are parameters; and the monthly indicator variables, M_{it}, are defined by

$$\begin{aligned} M_{1t} &= 1, \quad t \sim \text{January} \\ &= 0, \quad \text{otherwise} \\ &\vdots \\ M_{12,t} &= 1, \quad t \sim \text{December} \\ &= 0, \quad \text{otherwise.} \end{aligned}$$

For convenience, assume that $t = 1$ is a January and that $E(u_t) = 0$ for all t, so under Model 1 $\alpha_1, \ldots, \alpha_{12}$ are the monthly means and $E(z_t) = \alpha_t (t = 1, \ldots, 12)$. We actually do not need any assumptions about u_t for what follows other than later assuming that it has continuous probability distributions. Still the most common application of our result would be to the case in which u_t follows an autoregressive integrated moving average (ARIMA) model. I could have replaced u_t in Model 2 by another time series, v_t, as long as it was assumed that u_t and v_t have the same joint probability distributions.

Models 1 and 2 have been considered extensively in the literature for modeling, forecasting, and seasonal adjustment. Model 1 provides for deterministic (fixed) seasonality and Model 2 for one form of stochastic (moving) seasonality (see, e.g., Cleveland and Tiao 1979; Pierce 1978). In practice, the question arises as to which model is more suitable for a given time series. My result states that Model 1 and Model 2 are equivalent if $\Theta = 1$ and if the same assumptions about the initial observations $(z_1, \ldots, z_{12})$ are made under both models. By "equivalent," I mean that the joint probability distribution of any set of z_t's is the same under both models. I shall argue that distinguishing between the two models by making different assumptions about the initial observations is generally not appropriate or important in practice, so in practice the two models can be regarded as equivalent when $\Theta = 1$. Model 2 with $\Theta = 1$ thus provides a link between stochastic and deterministic seasonality, and for $\Theta = 1$ there is no choice to be made between the two models.

In Section 2 I state my result, discuss the practical relevance of assumptions about the initial observations, and prove the result to illustrate its implications. The result is actually a corollary of a more general result on models with cancellation of operators, or "overdifferencing," which is stated and proved in Section 3. Results such as these were given by Abraham and Box (1978) and Harvey (1981), but with little or no consideration given to assumptions about initial observations. I also discuss this work in Section 3.

2. RESULT, DISCUSSION, AND PROOF

For convenience, I may state and prove my result for $(z_1, \ldots, z_n)' = \mathbf{z}$. The set of time points $1, \ldots, n$ might correspond to times at which z_t is observed, or to some observed and some future times, for example.

At first the equivalence result may appear obvious, since applying $(1 - B^{12})$ to Model 1 yields

$$(1 - B^{12})z_t = (1 - B^{12})u_t,$$

[since $(1 - B^{12})M_{it} = 0$] and this is Model 2 with $\Theta = 1$. Since Model 2 contains the seasonal difference $1 - B^{12}$, however, it requires additional assumptions about 12 starting values not required under Model 1 (see Bell 1984). [If $w_t = (1 - B^{12})z_t$ requires differencing, assumptions about additional starting values would be required, but these need not concern us here because I assume that they would be the same under both models.] These assumptions can be made about the 12 initial observations, $z_1, \ldots, z_{12}$. A little thought then makes it clear that equivalence of the two models is achieved if $z_1, \ldots, z_{12}$ have the same distribution (joint with all other random variables involved) under both models. This means that for Model 2 I must have $z_t = \alpha_t + u_t$ $(t = 1, \ldots, 12)$. This leads us to my result, which I shall prove later.

Result. Models 1 and 2 given previously are equivalent iff in Model 2 we have (a) $\Theta = 1$ and (b) $z_t = \Sigma\alpha_i M_{it} + u_t = \alpha_t + u_t$ $(t = 1, \ldots, 12)$. (Summations involving M_{it} are over $i = 1, \ldots, 12$.)

To examine the implications of the models, I consider the joint probability density for $\mathbf{z} = (z_1, \ldots, z_n)'$ under the two models. Let $p(\cdot)$ denote the joint density for any given set of random variables—that is, the appropriate density for the given arguments, which can vary, but which will always be specified. Since the transformation

$$\begin{bmatrix} z_1 \\ \cdot \\ \cdot \\ \cdot \\ z_{12} \\ w_{13} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ w_n \end{bmatrix} = \begin{bmatrix} 1 & & & & & & \\ 0 & \cdot & & & & & \\ \cdot & \cdot & \cdot & & & & \\ \cdot & \cdot & \cdot & \cdot & & & \\ 0 & \cdot & \cdot & 0 & 1 & & \\ -1 & 0 & \cdot & \cdot & 0 & 1 & \\ & \cdot & & & \cdot & \cdot & \\ & & \cdot & & & \cdot & \cdot \\ & & & \cdot & & \cdot & \cdot \\ & & & & \cdot & \cdot & \cdot \\ & & & & -1 & 0 \cdot \cdot & 0 \;\; 1 \end{bmatrix} \begin{bmatrix} z_1 \\ \cdot \\ \cdot \\ \cdot \\ z_{12} \\ z_{13} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ z_n \end{bmatrix}$$

has a Jacobian of 1, we have

$$\begin{aligned} p(\mathbf{z}) &= p(z_1, \ldots, z_{12}, w_{13}, \ldots, w_n) \\ &= p(w_{13}, \ldots, w_n) \\ &\qquad \times p(z_1, \ldots, z_{12} \mid w_{13}, \ldots, w_n). \end{aligned} \tag{1}$$

Under Model 1, $w_t = (1 - B^{12})u_t$ and under Model 2, $w_t = (1 - \Theta B^{12})u_t$, so $p(w_{13}, \ldots, w_n)$ in (1) is the same under Models 1 and 2 iff $\Theta = 1$. Under Model 1,

$$\begin{aligned} &p_z(z_1, \ldots, z_{12} \mid w_{13}, \ldots, w_n) \\ &\quad = p_z(\alpha_1 + u_1, \ldots, \alpha_{12} + u_{12} \mid w_{13}, \ldots, w_n) \\ &\quad = p_u(u_1, \ldots, u_{12} \mid w_{13}, \ldots, w_n). \end{aligned} \tag{2}$$

I use $p_z(\cdot)$ and $p_u(\cdot)$ in (2) to emphasize when I mean the density of the z_t's and when I mean that of the u_t's. Condition (b) that $z_t = \alpha_t + u_t$ $(t = 1, \ldots, 12)$ is the same as saying that (2) holds. So under Model 2, $\Theta = 1$ and condition (b) imply that both terms on the right side of (1) are the same as under Model 1 and hence that $p(\mathbf{z})$ is the same under both models—the models are equivalent.

Questions about the equivalence of Models 1 and 2 arise most often in practice when a model of the form of Model 2 (say with u_t following an ARIMA model) has been fitted and the estimate of Θ, which must lie in $[-1, 1]$, is either the boundary value of 1 or close enough to it to make the model with $\Theta = 1$ seem reasonable. Cancelling $1 - B^{12}$ from both sides of Model 2 leads to Model 1, or rather it does if condition (b) of the result holds [equivalently, (2) holds]. Thus one would consider using Model 1, although there might be some question about (2) holding. It is my opinion that this should not be a concern and that, in practice, when $\Theta = 1$ Models 1 and 2 should be regarded as equivalent. There are two reasons for this.

The first reason is that I see no justification for making different assumptions about $p(z_1, \ldots, z_{12} \mid w_{13}, \ldots, w_n)$ under Models 1 and 2 when $\Theta = 1$. The assumptions typically made in practice are made really for convenience: under Model 1, (b) of the result is assumed, but under Model 2, one works only with the differenced data $w_{13}, \ldots, w_n$, by using $p(w_{13}, \ldots, w_n)$. The latter assumes that $p(z_1, \ldots, z_{12} \mid w_{13}, \ldots, w_n)$ in (1) does not involve the model parameters. In principle, either of these assumptions about $p(z_1, \ldots, z_{12} \mid w_{13}, \ldots, w_n)$ could be used in the other model (yielding equivalent results), or still other assumptions could be used. In practice, this could be difficult, and it is my impression that this is rarely considered. Whatever assumptions I might make about $p(z_1, \ldots, z_{12} \mid w_{13}, \ldots, w_n)$, however, probably could not be checked from the data anyway. Thus an argument that when $\Theta = 1$ Models 1 and 2 should be distinguished because I wish to make different assumptions about $p(z_1, \ldots, z_{12} \mid w_{13}, \ldots, w_n)$ seems unconvincing.

One way to get a $p(z_1, \ldots, z_{12} \mid w_{13}, \ldots, w_n)$ not depending on model parameters is to assume $z_1, \ldots, z_{12}$ independent of $w_{13}, \ldots, w_n$ (assumption A of Bell 1984) and assume a diffuse distribution for $z_1, \ldots, z_{12}$. Ansley and Kohn (1985) discussed such diffuse "priors." In this case $p(\mathbf{z}) = p(w_{13}, \ldots, w_n) = p(z_{13}, \ldots, z_n \mid z_1, \ldots, z_{12})$, and use of only the differenced data is appropriate and the same as analyzing the data conditional on $z_1, \ldots, z_{12}$. If $z_1, \ldots, z_{12}$ is not independent of $w_{13}, \ldots, w_n$, then $p(w_{13}, \ldots, w_n) \neq p(z_{13}, \ldots, z_n \mid z_1, \ldots, z_{12})$.

The second reason for assuming that Models 1 and 2 are equivalent when $\Theta = 1$ is that assumptions about $z_1, \ldots, z_{12}$, as expressed in $p(z_1, \ldots, z_{12} \mid w_{13}, \ldots, w_n)$, should usually have no effect asymptotically and thus should make little difference in practice with reasonably long time series. To see why, notice from (1) that the log-likelihood function, l, is

$$
\begin{aligned}
l = \ln p(z) &= \ln p(w_{13}, \ldots, w_n) \\
&\quad + \ln p(z_1, \ldots, z_{12} \mid w_{13}, \ldots, w_n) \\
&= \sum_{t=13}^{n} \ln p(w_t \mid w_{13}, \ldots, w_{t-1}) \\
&\quad + \ln p(z_1, \ldots, z_{12} \mid w_{13}, \ldots, w_n),
\end{aligned}
$$

where the first term in the sum is understood to be $\ln p(w_{13})$. Assuming that $p(w_t \mid w_{13}, \ldots, w_{t-1})$ and $p(z_1, \ldots, z_{12} \mid w_{13}, \ldots, w_t)$ exhibit some sort of stable behavior as t grows large, we see the behavior of

$$
\begin{aligned}
n^{-1}l &= n^{-1} \sum_{t=13}^{n} \ln p(w_t \mid w_{13}, \ldots, w_{t-1}) \\
&\quad + n^{-1} \ln p(z_1, \ldots, z_{12} \mid w_{13}, \ldots, w_n)
\end{aligned}
$$

will be governed by the behavior of the first term, since the second term will approach 0. This analysis could be checked in any specific case for any particular assumptions about $z_1, \ldots, z_{12}$. [The assumptions corresponding to what is known as conditional likelihood estimation for moving average models (Box and Jenkins 1970) do not fit into this framework. These assumptions deal with initial random shocks in the model for w_t, not with $z_1, \ldots, z_{12}$, and hence affect the first term in $n^{-1}l$ rather than the second.]

The equivalence of Model 1 and 2 with $\Theta = 1$ means that computations done (correctly) with either model will yield the same results. The effort required may not be the same, however. If u_t is stationary, computations are typically much easier under Model 1 than under Model 2. To take the simplest case, if u_t is white noise, Model 1 is a simple linear regression model, whereas Model 2 with $\Theta = 1$ is a noninvertible ARIMA model. For the latter, computations such as evaluation of the likelihood function (Hillmer and Tiao 1979; Ljung and Box 1979) or computing of forecasts (Harvey 1981) are somewhat difficult. On the other hand, if u_t in fact needs seasonal differencing (and possibly also regular differencing) to achieve stationarity, then the preferred strategy would be to apply $1 - B^{12}$ to Model 1, annihilating $\sum_1^{12} \alpha_i M_{it}$ and yielding Model 2 with $\Theta = 1$. I can then work with $w_t = (1 - B^{12})z_t = (1 - B^{12})u_t$ for $t = 13, \ldots, n$. Direct use of Model 1 in this case requires explicit assumptions about the 12 initial observations $u_1, \ldots, u_{12}$ of the nonstationary series u_t. Since there is generally little basis for such assumptions anyway, one may as well assume a diffuse distribution for $z_1, \ldots, z_{12}$ (equivalently for $u_1, \ldots, u_{12}$) as discussed previously. This leads to using Model 2 with $\Theta = 1$ for $w_{13}, \ldots, w_n$. This discussion also points out that Model 2 implicitly allows for deterministic seasonality, with this allowance made explicit if explicit assumptions are made about the initial observations.

It should be kept in mind that Model 2 is more general than Model 1, because Model 1 imposes the constraint $\Theta = 1$. This constraint may affect identification and estimation of models for the stochastic structure of u_t, so Models 1 and 2 (with Θ not constrained to be 1) may lead to overall models of different form.

Proof of Result. The argument at the beginning of this section amounts to proving that Model 1 implies Model 2 with $\Theta = 1$ and $z_t = \alpha_t + u_t$ ($t = 1, \ldots, 12$). To prove the reverse implication, and to gain further insight into the relationships between the models, I solve the difference equation given by Model 2 for any Θ. Letting $t = i + 12k$ for $i = 1, \ldots, 12$ and $k \geqslant 0$, it can be easily seen that

$$
\begin{aligned}
z_{i+12k} &= z_i + \sum_{j=1}^{k} w_{i+12j}, \qquad k \geqslant 1 \\
&= z_i + \sum_{j=1}^{k} [u_{i+12j} - \Theta u_{i+12(j-1)}] \\
&= z_i + u_{i+12k} + (1 - \Theta) \sum_{j=1}^{k-1} u_{i+12j} - \Theta u_i.
\end{aligned}
\tag{3}
$$

If $\Theta = 1$, this reduces to

$$z_{i+12k} = z_i + u_{i+12k} - u_i, \qquad k \geqslant 0. \tag{4}$$

Now, by using the condition $z_i = \alpha_i + u_i$ ($i = 1, \ldots, 12$), we get

$$z_{i+12k} = \alpha_i + u_{i+12k}, \tag{5}$$

which is Model 1, thus proving the result.

To see the necessity of conditions (a) and (b) of the result for Model 2, notice the following:

1. If $\Theta \neq 1$, then Models 1 and 2 are not equivalent, because (3) will not reduce to (4); z_t will depend not just on u_t, but also on $u_{t-12}, u_{t-24}, \ldots$.
2. If condition (a) of the result holds but not (b), then (4) will not reduce to (5) and Models 1 and 2 are not strictly equivalent. In this case Models 1 and 2 say something different about the initial observations $z_1, \ldots, z_{12}$, although having the same implications for $w_t = (1 - B^{12})z_t$ ($t = 13, \ldots, n$).

As another way of looking at this, notice that applying $1 - B^{12}$ to Model 1 and setting $\Theta = 1$ in Model 2 leads to the same difference equation:

$$(1 - B^{12})z_t = u_t - u_{t-12}. \quad (6)$$

The general solution to (6) is the sum of any particular solution and a solution to the homogeneous equation

$$(1 - B^{12})\mu_t = 0. \quad (7)$$

Because (6) is a twelfth-order equation, a particular solution is determined by specifying z_t for 12 values of t. The particular solution given by Model 1 is determined by the conditions $z_t = \alpha_t + u_t$ $(t = 1, \ldots, 12)$. Without initial conditions the solution under Model 2 with $\Theta = 1$ (say $z_t^{\text{Model 2}}$) can differ from that under Model 1 (say $z_t^{\text{Model 1}}$) by any solution to (7), which can be written as $\mu_t = \sum \beta_i M_{it}$. In symbols, $z_t^{\text{Model 2}} = z_t^{\text{Model 1}} + \sum \beta_i M_{it}$. Condition (b) requires that $\beta_1 = \cdots = \beta_{12} = 0$; without it or other initial conditions, Model 2 leaves the β_i unspecified. For example, under Model 2, I could set $\beta_i = -(\alpha_i + u_i)$ so that $z_1 = \cdots = z_{12} = 0$.

3. GENERAL RESULT AND RELATED WORK

Let $\delta(B) = 1 - \delta_1 B - \cdots - \delta_d B^d$, and let $f(t; \boldsymbol{\beta})$ be the general solution to $\delta(B)f(t; \boldsymbol{\beta}) = 0$, which depends on parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)'$. We can write $f(t; \beta) = \sum_{i=1}^{d} \beta_i X_{it} = \mathbf{X}_t'\boldsymbol{\beta}$, where $\mathbf{X}_t = (X_{1t}, \ldots, X_{dt})'$ are d linearly independent solutions to $\delta(B)X_{it} = 0$. Abraham and Box (1978) noted that the X_{it} can be exponential, polynomial, or trigonometric functions of t, or any mixture of these. I consider the following two models:

$$\text{Model } 1'. \quad z_t = f(t; \boldsymbol{\beta}) + u_t = \mathbf{X}_t'\boldsymbol{\beta} + u_t,$$

$$\text{Model } 2'. \quad \delta(B)z_t = w_t = \delta(B)u_t.$$

These models specialize to the case considered before when $\delta(B) = 1 - B^{12}$.

Theorem. Models 1′ and 2′ are equivalent iff in Model 2′ we have $z_t = \mathbf{X}_t'\boldsymbol{\beta} + u_t$ $(t = 1, \ldots, d)$.

This theorem could be proved in an instructive way analogous to the proof in Section 2. It also follows from previously noted results on solutions to difference equations, since z_t in Model 1′ solves the difference equation of Model 2′, and the condition $z_t = \mathbf{X}_t'\boldsymbol{\beta} + u_t$ $(t = 1, \ldots, d)$ assures that the particular solution used in Model 2′ is that given in Model 1′.

Abraham and Box (1978) (see also Abraham and Ledolter 1983) essentially gave the preceding result—they actually started with a constant term added to the right side of Model 2′ and thus got a polynomial of degree $r + 1$, say, in Model 1′ (along with other deterministic terms), where r is the number of times $(1 - B)$ occurs in $\delta(B)$. They also noted the special case in which $\delta(B) = 1 - B^{12}$, although expressing $f(t; \boldsymbol{\beta})$ equivalently for this case in terms of trigonometric functions rather than monthly indicator variables. Cleveland and Tiao (1979) considered models with seasonal indicator variables (Model 1) versus Model 2, noting that seasonally differencing Model 1 produces Model 2 with $\Theta = 1$. Neither Cleveland and Tiao (1979) nor Abraham and Box (1978) considered assumptions about initial observations or their implications. My arguments that differences in assumptions about initial conditions are not relevant for practice provides practical justification for their ignoring this aspect of the problem.

Harvey (1981) considered the use of Models 1′ and 2′ in forecasting, also noting the special case in which $\delta(B) = 1 - B^{12}$. He used the Kalman filter to construct a finite sample predictor and showed that if this is used Models 1′ and 2′ yield the same results for forecasting. In doing this, Harvey (1981) made initializing assumptions for the Kalman filter that are different from the assumptions about the initial observations that I consider here. His assumptions resulted in the Kalman filter when applied to Model 2′, implicitly doing generalized least squares (GLS) estimation of the parameters $\boldsymbol{\beta}$ in Model 1′ given an autoregressive moving average structure for u_t and taking this into account in computing forecasts and forecast error variances. Under his assumptions, Models 1′ and 2′ are not equivalent as stated, because the GLS estimation of $\boldsymbol{\beta}$ would still need to be performed in Model 1′. Harvey and Phillips (1979) showed how to use the Kalman filter to do this by making appropriate assumptions about the initial observations for Model 1′ that are different from mine here. Thus Harvey (1981) extended these results to Model 2′ by making assumptions about the initial observations consistent with those of Harvey and Phillips (1979) for Model 1′.

ACKNOWLEDGMENTS

I wish to thank David F. Findley, a referee, and the associate editor for useful comments.

[Received October 1985. Revised September 1986.]

REFERENCES

Abraham, B., and Box, G. E. P. (1978), "Deterministic and Forecast-Adaptive Time-Dependent Models," *Applied Statistics,* 27, 120–130.

Abraham, B., and Ledolter, J. (1983), *Statistical Methods for Forecasting,* New York: John Wiley.

Ansley, C. F., and Kohn, R. (1985), "Estimation, Filtering, and Smoothing in State Space Models With Incompletely Specified Initial Conditions," *The Annals of Statistics,* 13, 1286–1316.

Bell, W. R. (1984), "Signal Extraction for Nonstationary Time Series," *The Annals of Statistics,* 12, 646–664.

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

Cleveland, W. P., and Tiao, G. C. (1979), "Modelling Seasonal Time Series," *Economie Appliquée,* 32, 107–129.

Harvey, A. C. (1981), "Finite Sample Prediction and Overdifferencing," *Journal of Time Series Analysis,* 2, 221–232.

Harvey, A. C., and Phillips, G. D. A. (1979), "Maximum Likelihood Estimation of Regression Models With Autoregressive-Moving Average Disturbances," *Biometrika*, 66, 49–58.

Hillmer, S. C., and Tiao, G. C. (1979), "Likelihood Function of Stationary Multiple Autoregressive Moving Average Models," *Journal of the American Statistical Association*, 74, 652–667.

Ljung, G. M., and Box, G. E. P. (1979), "The Likelihood Function of Stationary Autoregressive-Moving Average Models," *Biometrika*, 66, 265–270.

Pierce, D. A. (1978), "Seasonal Adjustment When Both Deterministic and Stochastic Seasonality Are Present," in *Seasonal Analysis of Economic Time Series*, ed. A. Zellner, Washington, DC: U.S. Department of Commerce, Bureau of the Census, pp. 242–269.