BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-84/32

A NOTE ON THE EQUIVALENCE OF MODELS WITH MONTHLY MEANS AND MODELS WITH $(0,1,1)_{12}$ SEASONAL PARTS WHEN $\Theta = 1$

by

William R. Bell
Statistical Research Division
Bureau of the Census
Room 3524, F.O.B. #3
Washington, D.C. 20233 U.S.A.

Recommended by: Nash J. Monsour

Report completed: November 29, 1984

Report issued: December 10, 1984

A Note on the Equivalence of Models with Monthly Means and Models with $(0,1,1)_{12}$ Seasonal Parts when $\Theta = 1$

William Bell
Statistical Research Division

November 12, 1984

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

$$\text{Model I :} \quad z_t = \sum_{1}^{12} \alpha_i M_{it} + u_t$$

$$\text{Model II:} \quad (1-B^{12})z_t = w_t = (1-\Theta B^{12})u_t$$

(We use monthly time series for concreteness. All that follows applies immediately to time series with other seasonal periods.) 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

$$M_{1t} = \begin{cases} 1 & t \sim \text{Jan.} \\ 0 & \text{otherwise} \end{cases}$$

$$\vdots$$

$$M_{12,t} = \begin{cases} 1 & t \sim \text{Dec.} \\ 0 & \text{otherwise} \end{cases} .$$

If $E(u_t) = 0$, then under Model I $\alpha_1 = E(z_t)$ when time point t is a January, $\alpha_2 = E(z_t)$ when time point t is a February, etc., so that $\alpha_1,\ldots,\alpha_{12}$ are then the monthly means. We actually do not need any assumptions about u_t for what follows other than later assuming it has continuous probability distributions. Still the most common application of our result would be to the case where u_t follows an ARIMA (autoregressive-integrated-moving average) model. We could have replaced u_t in Model II by another time series v_t, as long as it was assumed that u_t and v_t have the same probability structure.

Our result states that Model I and Model II can be regarded as equivalent when $\Theta = 1$. By this we mean that the joint probability distribution of any set of z_t's is the same under both models. For convenience, we may demonstrate this for the set $(z_1,\ldots,z_n) = \underset{\sim}{z}$. 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 I yields

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

(since $(1-B^{12})M_{it} = 0$) and this is Model II with $\Theta = 1$. However, Model I applies to all $t=1,\ldots,n$ whereas Model II, as specified, applies only to $t=13,\ldots,n$ and does not say anything about $z_1,\ldots,z_{12}$. A little further thought makes it clear that equivalence of the two models also requires that $z_1,\ldots,z_{12}$ have the same distribution (joint with all other random variables involved) under both models. This means that for Model II we must have $z_t = \alpha_t + u_t$ $t=1,\ldots,12$ (we shall hereafter assume t=1 is a January). This leads us to our result.

<u>Theorem</u>: Models I and II given above are equivalent if and only if in Model II we have (i) $\Theta = 1$, and (ii) $z_t = \sum_1^{12}\alpha_i M_{it} + u_t$ $t=1,\ldots,12$.

<u>Proof</u>: The argument above amounts to proving that Model I implies Model II with $\Theta = 1$ and $z_t = \sum_1^{12}\alpha_i M_{it} + u_t$ $t=1,\ldots,12$. To prove the reverse im-

plication, and to gain further insight into the relationships between the models, we solve the difference equation given by Model II for any Θ. Letting $t = i+12k$ for $i=1,\ldots,12$ and $k \geq 0$ we easily see that

$$z_{i+12k} = z_i + \sum_{j=1}^{k} w_{i+12j} \qquad k \geq 1$$

$$= z_i + \sum_{j=1}^{k} (1-\Theta B^{12}) u_{i+12j}$$

$$= z_i + \sum_{j=1}^{k} [u_{i+12j} - \Theta u_{i+12(j-1)}]$$

$$= z_i + u_{i+12k} + (1-\Theta) \sum_{j=1}^{k-1} u_{i+12j} - \Theta u_i \ . \tag{1}$$

If $\Theta = 1$ this reduces to

$$z_{i+12k} = z_i + u_{i+12k} - u_i \qquad k \geq 0 \tag{2}$$

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

$$z_{i+12K} = \alpha_i + u_{i+12k} \tag{3}$$

which is Model I, thus proving the theorem. QED

To see the necessity of conditions (i) and (ii) of the theorem for Model II notice the following:

(a) If $\Theta \neq 1$ then Models I and II are not equivalent since (1) above will not reduce to (2) - z_t will depend not just on u_t, but also on $u_{t-12}, u_{t-24}, \ldots$

(b) If condition (i) of the theorem holds but not (ii), then (2) above will not reduce to (3) and Models I and II are not equivalent. In this case Models I and II say something different about the starting values $z_1, \ldots, z_{12}$, though having the same implications for $w_t = (1-B^{12})z_t$, $t=13, \ldots, n$.

<u>Note</u>: Applying $1-B^{12}$ to Model I and setting $\Theta = 1$ in Model II leads to the same difference equation:

$$(1-B^{12})z_t = u_t - u_{t-12} \, . \tag{4}$$

The general solution to (4) is the sum of any particular solution and a solution to the homogeneous equation

$$(1-B^{12})\mu_t = 0 \, . \tag{5}$$

As (4) is a twelfth order equation a particular solution is determined by specifying z_t for twelve values of t. The particular solution given by Model I is determined by the conditions $z_i = \alpha_i + u_i$ $i=1, \ldots, 12$. Without initial conditions the solution under Model II with $\Theta = 1$ (say z_t^{II}) can differ from that under Model I (say z_t^{I}) by any solution to (5), which can be written $\mu_t = \sum_1^{12} \beta_i M_{it}$. In symbols $z_t^{II} = z_t^I + \sum_1^{12} \beta_i M_{it}$.

Condition (ii) requires $\beta_1 = \ldots = \beta_{12} = 0$; without it or other initial conditions, Model II leaves the β_i unspecified. For example, under Model II we could set $\beta_i = -(\alpha_i + u_i)$ so that $z_1 = \ldots = z_{12} = 0$.

To further examine the implications of the models we consider the joint probability density for $\underset{\sim}{z} = (z_1, \ldots, z_n)$ 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

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

has unit Jacobian we have

$$p(\underset{\sim}{z}) = p(z_1,\ldots,z_{12},w_{13},\ldots,w_n)$$
$$= p(w_{13},\ldots,w_n)p(z_1,\ldots,z_{12}|w_{13},\ldots,w_n) \tag{6}$$

Under Model I $w_t = (1-B^{12})u_t$, and under Model II $w_t = (1-\Theta B^{12})u_t$ so that $p(w_{13},\ldots,w_n)$ in (6) is the same under Models I and II if and only if $\Theta = 1$. Under Model I

$$p_z(z_1,\ldots,z_{12}|w_{13},\ldots,w_n) = p_z(\alpha_1+u_1,\ldots,\alpha_{12}+u_{12}|w_{13},\ldots,w_n)$$
$$= p_u(u_1,\ldots,u_{12}|w_{13},\ldots,w_n) \tag{7}$$

We use $p_z(\cdot)$ and $p_u(\cdot)$ in (7) to emphasize when we mean the density of the z_t's and when we mean that of the u_t's. Condition (ii) that

$z_t = \sum_1^{12}\alpha_i M_{it} + u_t$ $t=1,\ldots,12$ is the same as saying that (7) holds. So under Model II, $\Theta = 1$ and condition (ii) imply that both terms on the right hand side of (7) are the same as under Model I, and hence that $p(\underset{\sim}{z})$ is the same under both models - the models are equivalent.

Comments

Questions about the equivalence of Models I and II arise most often in practice when a model of the form of Model II (say with u_t following an ARIMA model) has been fitted and the estimate of Θ, which by invertibility 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 II leads to Model I, or rather it does if condition (ii) of the theorem holds (equivalently, (7) holds). Thus, one would consider using Model I, although there might be some question about (7) holding. It is our opinion that this should not be a concern, and that in practice when $\Theta = 1$ Models I and II should be regarded as equivalent. There are two reasons for this.

The first reason is that we see no justification for making different assumptions about $p(z_1,\ldots,z_{12}|w_{13},\ldots,w_n)$ under Models I and II when $\Theta = 1$. The assumptions typically made in practice are made really for convenience: under Model I (7) is assumed (so the model holds for all t), while under Model II one works only with the differenced data $w_{13},\ldots,w_n$, which is the same as setting $p(z_1,\ldots,z_{12}|w_{13},\ldots,w_n) = 1$ in (6), or assuming that $z_1,\ldots,z_{12}$ are degenerate random variables, or analyzing the data conditional on $z_1,\ldots,z_{12}$. In fact, if under Model II with $\Theta = 1$ we make assumptions about $p(z_1,\ldots,z_{12}|w_{13},\ldots,w_n)$ that do not correspond to $z_i = \alpha_i + u_i$, $i=1,\ldots,12$, there is no reason we cannot use Model I by redefining it to hold only for $t \geqslant 13$, and then making the same assumptions about $p(z_1,\ldots,z_{12}|w_{13},\ldots,w_n)$ that we were making under Model II. Most assumptions we would make about $p(z_1,\ldots,z_{12}|w_{13},\ldots,w_n)$ cannot be checked from the data anyway. Thus an argument that when $\Theta = 1$ Models I and II should be distinguished by assumptions about $p(z_1,\ldots,z_{12}|w_{13},\ldots,w_n)$ seems unconvincing.

The other reason for assuming that Models I and II are equivalent when $\Theta = 1$ is that any reasonable assumptions about $z_1,\ldots,z_{12}$, as expressed in $p(z_1,\ldots,z_{12}|w_{13},\ldots,w_n)$, will have no effect asymptotically, and thus should make little difference in practice. To see why, notice from (6) that the log-likelihood function, ℓ, is

$$\ell = \ell n\ p(\underset{\sim}{z}) = \ell n\ p(w_{13},\ldots,w_n) + \ell n\ p(z_1,\ldots,z_{12}|w_{13},\ldots,w_n)$$

$$= \sum_{t=13}^{n} \ell n\ p(w_t|w_{13},\ldots,w_{t-1}) + \ell n\ p(z_1,\ldots,z_{12}|w_{13},\ldots,w_n).$$

Assuming $p(w_t|w_{13},\ldots,w_{t-1})$ and $p(z_1,\ldots,z_{12}|w_{13},\ldots,w_t)$ exhibit some sort of stable behavior as t grows large, we see the behavior of

$$n^{-1}\ \ell = n^{-1} \sum_{t=13}^{n} \ell n\ p(w_t|w_{13},\ldots,w_{t-1}) + n^{-1}\ \ell n\ p(z_1,\ldots,z_{12}|w_{13},\ldots,w_n)$$

will be governed by the behavior of the first term, since the second term will approach zero. Analogous analyses could be given for other manipulations involving the likelihood function.

The equivalence of Model I and Model II with $\Theta = 1$ means that computations done (correctly) with either model will yield the same results. However, the effort required may not be the same: computations are typically much easier under Model I than under Model II. To take the simplest case, if u_t is white noise Model I is a simple linear regression model whereas Model II with $\Theta = 1$ is a noninvertible ARIMA model. For the latter, computations such as evaluation of the likelihood function (Ljung and Box 1979, Hillmer and Tiao 1979) or computing of forecasts (Harvey 1981) are somewhat difficult.

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

REFERENCES

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

This paper demonstrates that, assuming the finite sample predictor is used and the computations are done correctly, forecasts based on Model I are the same as forecasts based on Model II with $\Theta = 1$.

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.