MODELING AND MODEL SELECTION FOR MOVING HOLIDAYS

David F. Findley and Raymond J. Soukup, Census Bureau
David F. Findley, Census Bureau, Washington, DC 20233-9100

1. Introduction

Certain kinds of economic activity, and their associated time series, are affected significantly by holidays. When the date of a holiday changes from year to year, the effects of the holiday can impact two or more months in a way that depends on the date. In this case, the effect of the holiday is not confined to the seasonal component of the series. If such holiday effects are ignored, then models fit to the time series will often have reduced forecasting ability. Also the model residuals may show a lack of fit. Similarly, seasonal adjustments of the series may provide misleading signals in months affected by the holidays. In the U.S., the major moving holidays are Easter, Labor Day, and Thanksgiving, the dates of which vary over March 22–April 25, September 1-7, and November 22-28, respectively. Elsewhere, holidays tied to a lunar calendar, such as the Chinese New Year, Passover, and Ramadan have an economic impact. Although we only apply the modeling and model selection procedures described in this paper to model U.S. holidays, they are applicable more generally.

Disclaimer. This paper reports results of research undertaken by Census Bureau staff. It has undergone a more limited review by the Census Bureau than do official publications. It is released to inform and to encourage discussion.

1.1 The Bell-Hillmer interval model

In the context of modeling Easter effects, Bell and Hillmer (1983) introduced a simple type of regressor that has proven to be versatile for modeling effects of a variety of moving holidays. Its underlying premise is that there is an interval of length τ days over which the effect of the holiday can be regarded as being the same for each day. With τ_t denoting the number of days in month t that belong to this interval, the value in month t of the holiday regressor $H(\tau, t)$ associated with this interval is defined to be the proportion of the interval contained within the month,

$$H(\tau, t) = \frac{\tau_t}{\tau}. \qquad (1)$$

Several such regressors can be used to model a more complicated effect, either in the way that a step function can be used to approximate a non-constant function, or to model the situation in which there are different intervals over which the effect of the holiday is different. For example, in some countries, around the Chinese New Year there is usually a surge in economic activity before the holiday, followed by a period of little activity, and then by a period of increasing activity back to normal levels. However, for the U. S. series considered in this article, a single interval seems adequate, an interval whose last day is the day before the holiday in the case of Easter and Labor Day. To illustrate, shoe store sales in the U.S. increase before Easter. Suppose an 8-day interval is considered for modeling this increase. Then if month t is March, 1994, a year in which Easter fell on April 3rd, we have $H(8, t-1) = 0$, $H(8, t) = 6/8$, $H(8, t+1) = 2/8$, $H(8, t+2) = 0$, etc.

With the Census Bureau's X-12-ARIMA program, holiday effects for one or several holidays can be estimated by means of a regARIMA model, i.e. a time series model specifying a regression mean function for the time series Y_t (or, more commonly, for its logarithms) that includes appropriate $H(\tau, t)$'s and specifying an ARIMA model for its autocovariance function. We indicate such a model with a formula of the form

$$\log Y_t = \boldsymbol{\beta}' \mathbf{X}_t + z_t, \qquad (2)$$

in which $\mathbf{X}_t$ denotes the mean function's regressor, which could also include trading day and outlier variables, and z_t denotes a mean zero ARIMA process. With B denoting the backshift operator, $BZ_t = Z_{t-1}$, and ε_t a white noise process, a more explicit formula is

$$\delta(B)^d \delta\left(B^{12}\right)^D \phi(B)\, \Phi\left(B^{12}\right)\left(\log Y_t - \boldsymbol{\beta}' \mathbf{X}_t\right) = \theta(B)\, \Theta\left(B^{12}\right) \varepsilon_t, \qquad (3)$$

where $\delta(x) = 1 - x$, ϕ,θ,Φ, and Θ are polynomials, and $d, D \geq 0$. To focus on a single holiday regressor (1) included in $\mathbf{X}_t$, we rewrite (2) as

$$\log Y_t = \alpha H(\tau, t) + y_t, \tag{4}$$

with $y_t = \boldsymbol{\beta}'\mathbf{X}_t - \alpha H(\tau, t) + z_t$. From (4), one sees that the model identifies the holiday's effect in Y_t as the factor $\exp(\alpha \tau_t / \tau)$, or approximately $1 + \alpha \tau_t / \tau$. To this approximation, 100α can be interpreted as the percentage effect of the holiday for a month that completely contains the interval associated with $H(\tau, t)$, and $100\alpha/\tau$ can be interpreted as the percentage effect of a single day in the holiday interval in any month that intersects the interval.

If each holiday interval of length τ is contained within a single calendar year, then the proportions $H(\tau, t)$ sum to one over each calendar year, a property we express by $\sum_{year} H(\tau, t) = 1$. In case holiday adjustment is the goal of modeling, we shall explain in Section 4 how $H(\tau, t)$ can be modified in a natural way to obtain a regressor $\tilde{H}(\tau, t)$ having the same coefficient but satisfying $\sum_{year} \tilde{H}(\tau, t) = 0$.

2. A Basic Model Selection Problem: Determining τ

Often, for a given holiday and series, even if it is known that a holiday effect is present and where an endpoint of its interval should be, an appropriate interval length τ is not known in advance. Competing models with different τ's are non-nested, i.e. none is a special case of another. Consequently, standard statistical tests are not available for determining τ. We shall consider two alternative approaches to this model selection problem, (1) the use of a variant of Akaike's AIC criterion, and (2) the comparison of out-of-sample forecast errors. Results from applying these approaches to choose holiday effect models for series from the U.S. Retail Trade Survey are given in Section 3. Both approaches can also be used with the more complex models of Morris and Pfeffermann (1984).

For this article, the parameters of every regARIMA model are estimated from data $Y_t, 1 \leq t \leq T$ by maximizing the associated Gaussian likelihood function. With $\hat{L}_{T-12D-d}$ designating the maximized log likelihood of a regARIMA model for Y_t with d, D as in (3), Hurvich and Tsay (1989)'s sample-size corrected version of the model's AIC, which we shall denote by AICC, is

$$AICC_{T-12D-d} = -2\hat{L}_{T-12D-d} + 2p\left\{\frac{1}{1 - \frac{p+1}{T-12D-d}}\right\} \tag{5}$$

where p denotes the number of estimated parameters in the model. AICC's can be compared for models with the same values of d and D and the same outlier regressors. Among such models, the one with the smallest AICC value is preferred. Although the value of p in (5) will be the same for models (4) that differ only in the value of τ, we shall also consider a model without the regressor $H(\tau, t)$, to represent the possibility of no effect in the interval, a model with one parameter less.

The theory supporting AIC and AICC assumes that the models fit the data reasonably well, see Findley (1999). Our other model selection procedure, which involves withholding some recent data, reestimating the models from the remaining data, and comparing their forecasts of the withheld data, does not have this requirement, nor the other requirements concerning identical outlier regressors and identical values of d and D. Its disadvantages are that it compares models two at a time and that it can be inconclusive: neither model may have persistently smaller forecast errors than the other over the interval for which forecast errors are obtained.

For models numbered $i = 1, 2$ and some forecast lead $h \geq 1$, let $Y_{t+h|t}^{(i)}$ denote model i's forecast of Y_{t+h} from time t obtained when its parameters are estimated from $Y_1, Y_2, \cdots, Y_t$. For $t \leq T - h$, we can calculate the resulting "out-of-sample" forecast error (OSFE) $Y_{t+h} - Y_{t+h|t}^{(i)}$. These forecasts and errors are calculated for $T_0 + 1 \leq t \leq T - h$, where T_0 is chosen large enough that parameter estimates used to produce the forecasts can be expected to be of reasonable quality (e.g. as indicated by their estimated standard errors or by comparison to the parameter values obtained with all of the data). For the Easter holiday coefficients and series discussed in the next section, $T_0 = 60$ or 72 seemed adequate.

The OSFE diagnostic for comparing the forecasting ability of models 1 and 2 over

$Y_{T_0+1}, Y_2, \cdots, Y_{T-h}$ are the graphs of

$$\frac{\sum_{t=T_0+1}^{T_0+N}\{(Y_{t+h}-Y_{t+h|t}^{(1)})^2-(Y_{t+h}-Y_{t+h|t}^{(2)})^2\}}{\sum_{t=T_0+1}^{T-h}(Y_{t+h}-Y_{t+h|t}^{(2)})^2/(T-T_0)} \tag{6}$$

versus N for$1 \leq N \leq T-T_0-h$ for several choices of h, usually $h = 1, 12$. From this formula, it is clear that over intervals of values of N where the graph goes persistently up, the forecasting performance of model (2) is better, i.e. has smaller accumulated forecast errors; where the graph goes persistently down, model (1) is better. Where the graph is mostly level, there is no essential difference in forecast performance of the two models. The denominator in (2) enables one to interpret jumps in the graph in term of units of mean square forecast error of the second model. More details are given in Findley et al. (1998).

The graph of (6) in Figure 1 shows the improvements in forecast errors for the Retail Sales from Shoe Stores series obtained for leads $h = 1, 12$ when a model (model 2) with an Easter interval regressor with $\tau = 8$ is used in place of a model without an Easter regressor (model 1). The improvement is substantial in most Marches and Aprils, as expected.

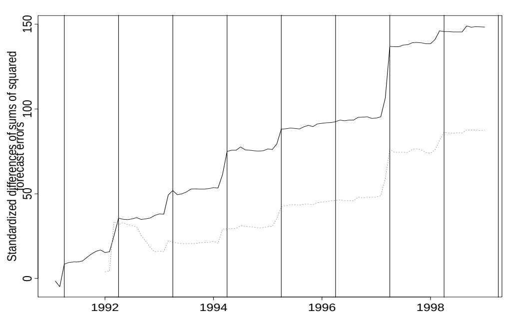

Figure 1: Plot of (6) comparing Model 1 with no Easter regressor to Model 2 with $H(8,t)$ using leads $h = 1$ (solid) and $h = 12$ (dashes). Vertical lines identify Aprils.

3. Determining Holiday Models for the U.S. Retail Trade Survey

We now summarize the results of model comparisons for determining which holiday regressors to use, if any, for Easter, Labor Day, and Thanksgiving effects, for the candidate series in the U.S. Retail Trade survey, using data ending in February, 1999 and, in most cases, starting in 1991 to avoid recession effects. We begin with Easter because its effects are largest and we did not find any series without Easter effects that required adjustment for other holidays.

3.1 Easter holiday effects

In a candidate set of forty-seven retail sales series, we found twenty-nine for which AICC supported inclusion of an Easter regressor with $\tau = 1, 8$, or 15. OSFE analysis of these twenty-nine series gave the following results: use of an Easter regressor improved forecasts for fifteen, had no clear effect on forecasts for thirteen, and produced worse forecasts for one series. Table 1 provides a summary.

Table 1. τ-Preferences

AICC	OSFE	No. of Series
1	1	1
8	8	10
8	8 or 1	1
8	8 or 15	1
8	1	2
8	inconclusive	8
15	inconclusive	5
15	no $H(\tau,t)$	1

We also considered $\tau = 2, 9, 16$, but the results were never better and usually worse.

When OSFE and AIC disagree, which can happen because the nature of the holiday effect has changed in the recent data, we favor the OSFE preference if the more recent data or forecasts have special importance for the users of the modeling results.

3.2 Thanksgiving and Labor Day effects

Thanksgiving (the fourth Thursday in November) and Labor Day (the first Monday in September) can affect a neighboring month's data in a way that depends on the holiday's date. For example, retail sales linked to the December holidays can be distributed differently between November and December from year-to-year depending on the date of Thanksgiving.

For Thanksgiving, we consider regressors for a holiday effect interval from some specified number of days ν before or after Thanksgiving through December 24. We use negative values of ν to indicate an effect that begins ν days *after* Thanksgiving. For Labor Day, the holiday effect intervals begin τ days before Labor Day.

For the Thanksgiving effect, we considered ν values of 10, 3, -1, -7, and -8 for the twenty-nine series with Easter effects according to AICC. The -7 and -8 values were clearly inferior, so we only present comparisons of models with the τ-values 10, 3 and -1. We always included the Easter regressor favored by AICC (and also trading day regressors) in the models considered for AICC and OSFE comparisons for Thanksgiving effects. The AICC comparisons for Thanksgiving suggested modeling its effects in nine series. Among these nine, OSFE analysis showed that inclusion of a Thanksgiving regressor resulted in persistent forecast improvements for six series, and that all regressors were detrimental to the forecasting of one series, indicating none should be used. For the six series for which a Thanksgiving regressor improved forecasts, there was OSFE preference for $\nu = -1$ for two series and for $\nu = 10$ for one series — for the other three series, there was no preferred value of ν among -1, 3, 10.

In the one series for which the OSFE analysis contradicted AICC's support for a Thanksgiving regressor, the AICC value for the model with no Thanksgiving regressor exceeded the minimum AICC value by less than 1.0. Distributional considerations for log-likelihood ratios suggest that AICC differences less than one in magnitude can be considered inconclusive, see Sakamoto, Ishiguro and Kitagawa (1986).

For Labor Day effects, we considered τ values of 2, 9 and 16, assuming that the Labor Day effect interval would begin on one of the three Saturdays before the holiday. Among the twenty-nine series with Easter effects according to AICC, the further AICC comparisons for Labor Day suggested modeling its effects in six series, always with $\tau = 2$. OSFE analysis favored inclusion of a Labor Day regressor in four of these six and was inconclusive for the other two. Among these four, OSFE analysis favored $\tau = 2$ for two series and was indifferent to the choice of τ for the other two.

4. Modifying the Models for Holiday Adjustment

If the goal of modeling is to remove the holiday effects from the data, along with any seasonal effects and trading day effects, then the regressors (1) may be somewhat unsatisfactory from both conceptual and practical viewpoints. Consider an example in which the estimated coefficient $\hat{\alpha}$ of $H(\tau,t)$ is positive, e.g. retail shoe sales. Then the factors $e^{\hat{\alpha}H(\tau,t)}$ exceed 1.0 in every month containing a day of the holiday interval. Removing the effects by dividing out these factors presupposes that the excess of shoes bought in the holiday interval would never have been bought had there not been a holiday, surely a largely incorrect assumption. Also, the adjusted series $Y_t/e^{\hat{\alpha}H(\tau,t)}$ will have a lower level than the original series in all months touched by the holiday interval and therefore consistently smaller annual totals,

$$\sum_{year} \frac{Y_t}{e^{\hat{\alpha}H(\tau,t)}} < \sum_{year} Y_t.$$

This will lead many users of the adjusted series to perceive it as being downwardly biased.

For some series and holidays, it may be possible to resolve this difficulty by finding another interval (or intervals) associated with the holiday whose Bell-Hillmer regressor has the opposite sign from $\hat{\alpha}$ and is such that the combined adjustment from the several regressors has annual totals that are usually close to those of the original series. But a more generally applicable strategy is needed, and one can be found in the device used by the X–11 seasonal adjustment procedure to try to insure that the seasonally adjusted series preserves levels adequately (see Appendix A of Findley et. al. (1998)), which is done by reweighting the seasonal factors so that their average values over a year are close to 1.0. In seeking a simple modification $\tilde{H}(\tau,t)$ of $H(\tau,t)$ so that the factors $e^{\hat{\alpha}\tilde{H}(\tau,t)} \doteq 1 + \hat{\alpha}\tilde{H}(\tau,t)$ have this property, we are led to consider modifications $\tilde{H}(\tau,t)$ such that

$$\frac{1}{12}\sum_{year} \tilde{H}(\tau,t) \doteq 0. \qquad (7)$$

The procedure of Bell (1984), which is implemented in X-12-ARIMA, achieves (7) by removing from $H(\tau,t)$ calendar month means of the form

$$\begin{aligned} H^*(\tau,t) &= \frac{1}{N_1-N_0}\sum_{n=N_0}^{N_1-1} H(\tau,\widetilde{t}+12(n-1)) \\ &= \frac{\frac{1}{N_1-N_0}\sum_{n=N_0}^{N_1-1}\tau_{\widetilde{t}+12(n-1)}}{\tau} \qquad (8) \\ &= \frac{\tau_t^*}{\tau}, \end{aligned}$$

with $N_1 > N_0$, where $\widetilde{t} = t - 12[t/12]$, i.e. the remainder after division of t by 12. Thus τ_t^* is the average of values τ_s from the same calendar month as t over an interval of $N_1 - N_0$ years. The modified regressor is

$$\tilde{H}(\tau,t) = H(\tau,t) - H^*(\tau,t) = \frac{\tau_t - \tau_t^*}{\tau}. \quad (9)$$

Note that if $\sum_{year} H(\tau,t) = 1$, as is the case with the U.S. holidays for any plausible τ, then $\sum_{year} H^*(\tau,t) = 1$, so $\sum_{year} \tilde{H}(\tau,t) = 0$, i.e. equality holds in (7). When this happens, the product of the adjustment factors $\exp\left(\hat{\alpha}\tilde{H}(\tau,t)\right)$ is 1 over each calendar year,

$$\prod_{year} e^{\hat{\alpha}\tilde{H}(\tau,t)} = 1. \quad (10)$$

The definition (8) yields

$$H^*(\tau,t) - H^*(\tau,t+12) = 0 \quad (11)$$

for any choice of N_0, N_1 $(N_1 > N_0)$. As a consequence, if $D \geq 1$ in (3) (or if $d \geq 1$ and seasonal indicators variables are included in X_t), then the coefficient estimates, forecasts, AICC's, etc. obtained from the use of $\tilde{H}(\tau,t)$ will be *identical* to those obtained with $H(\tau,t)$. This applies to the holiday analyses of the Retail Trade series presented above, because the models for the series all include a seasonal difference. Thus, changing to (9) would not change the model selections presented above. We note that (11) does not imply that $H^*(\tau,t)$ models a purely seasonal effect, because it has a level component, a long term mean close to $1/12$, due to $\sum_{year} H^*(\tau,t) \doteq 1$.

In every calendar month in which $H(\tau,t)$ is always 0, e.g. May and later months for Easter regressors, the $H^*(\tau,t)$ defined by (8) are 0, and consequently, $\tilde{H}(\tau,t) = 0$ and $e^{\hat{\alpha}\tilde{H}(\tau,t)} = 1$, i.e. adjustment does not modify the data, a property data users would expect. (There are no such months for Ramadan, which moves through the entire year. So for Ramadan, one could alternatively define $H^*(\tau,t) = \sum_{n=12(N_0-1)+1}^{12N_1} H(\tau,n)/12(N_1 - N_0)$, i.e., a global mean that does not depend on t, in order to have a constant adjustment factor, $e^{-\hat{\alpha}H^*(\tau,t)}$, for all months outside the holiday interval in a given year.)

4.1 Choices of N_0, N_1 and interpretation of $\exp\left(\hat{\alpha}\tilde{H}(\tau,t)\right)$

If the holiday calendar repeats itself every $12(N_1 - N_0)$ months, then it follows from the Fourier decomposition formulas of Section 4.2.3 of Anderson (1971) that $H^*(\tau,t)$ is the period 12 component of $H(\tau,t)$. Then $\tilde{H}(\tau,t)$ is the theoretical deseasonalized and mean-corrected component of $H(\tau,t)$ (equality will hold in (7)). It is theoretically attractive to choose N_0 and N_1 to achieve this. But when very long periods are involved, as in the case of Easter, it can simplify discussion with data users to choose smaller numbers that yield approximate periodicity (400 years in the case of Easter in X-12-ARIMA). Or one can simply specify N_0 and N_1 so that the sum in (8) covers the span of the available data. If both seasonal adjustment and holiday adjustment are being done, the differences in adjustment factors $e^{\hat{\alpha}\tilde{H}(\tau,t)}$ resulting from different choices of N_0 and N_1 are unimportant, because the products of the seasonal and holiday factors (the "combined factors") will be essentially unchanged for the following reason. Given different versions, say $H^*_{(1)}(\tau,t)$ and $H^*_{(2)}(\tau,t)$, of (8), if we define $\Delta H^*(\tau,t) = H^*_{(2)}(\tau,t) - H^*_{(1)}(\tau,t)$, then the corresponding adjustment factors satisfy

$$e^{\hat{\alpha}\tilde{H}_{(2)}(\tau,t)} = e^{\hat{\alpha}\Delta H^*(\tau,t)} e^{\hat{\alpha}\tilde{H}_{(1)}(\tau,t)}.$$

By (11), the ratio of the adjustment factors, $e^{\hat{\alpha}\Delta H^*(\tau,t)}$, satisfies $e^{\hat{\alpha}\Delta H^*(\tau,t+12)} = e^{\hat{\alpha}\Delta H^*(\tau,t)}$, and since $\sum_{year} \Delta H^*(\tau,t) \doteq 0$, its average value is a constant close to 1.0. So the ratio is essentially a multiplicative seasonal effect and, in practice, it seems to get absorbed almost completely into the seasonal factors of $Y_t/e^{\hat{\alpha}\tilde{H}_{(2)}(\tau,t)}$. Consequently, the combined factors will be almost the same.

Adjustment by $\exp\left(\hat{\alpha}\tilde{H}(\tau,t)\right)$ is an adjustment for the percentage difference of $\exp(\hat{\alpha}\tau_t)$ from $\exp(\hat{\alpha}\tau_t^*)$, the latter being the holiday effect were the calendar month of t to contain exactly its average of holiday interval days. Thus, we can interpret this adjustment as presupposing that if the holiday did not exist, we would have observed the value $\exp(\hat{\alpha}\tau_t^*)(Y_t/\exp(\hat{\alpha}\tau_t))$ instead of Y_t, and we are using adjustment to obtain this value. It follows from (10), or the more approximate relation implied by (7), that application of these holiday adjustment factors provides a percentage redistribution of these holiday effects throughout the calendar year.

The largest factors $\exp\left(\hat{\alpha}\tilde{H}(\tau,t)\right)$ obtained from the retail sales series for each U.S. holiday were: for Easter 1.06 (for Shoe Stores), for Thanksgiving 1.012 (for Department Stores and for Children's and Miscellaneous Apparel), and for Labor Day 1.015 (for Shoe Stores and for Discount Department Stores).

4.2 Indirect estimation with $\tilde{H}(\tau,t)$

If, instead of using a regARIMA model to directly estimate holiday effects from Y_t, they are estimated indirectly via a regression model for the irregular component of a seasonal adjustment of Y_t, the case for using a deseasonalized and level adjusted regressor like (9) becomes even stronger, because the irregular component is a deseasonalized and detrended version of Y_t. OSFE analysis can be used to compare indirect and direct modeling of holiday effects.

5. Concluding Remarks

Our conclusions from our study are the following. AICC comparisons can be used to identify series that are good candidates for holiday effect modeling as described in this paper. However, it is worthwhile to perform OSFE analyses of the series that are identified by AICC as having a holiday effect, as a confirmatory diagnostic. The OSFE results will usually be consistent with AICC results, but there are occasional exceptions due, for example, to a change in the strength or nature of the holiday's effect on the series in the latest years of the data.

All of the holiday effect estimates, AICC values, and out-of-sample forecast error sums of squares required for this study were obtained from the Census Bureau's X-12-ARIMA program, which is available free from

http://www.census.gov/srd/www/x12a/ (12)

It includes built-in regressors for Easter, Labor Day and Thanksgiving. An accompanying graphical diagnostics program, X-12-Graph, which can produce plots like (6) from X-12-ARIMA output, can also be downloaded from (12). A program *genhol* is available from *brian.c.monsell@census.gov* for generating regressors (1) and (9) for X-12-ARIMA from an input file of holiday dates. Regressors can be obtained for intervals before, surrounding, and after one or more holidays.

We thank Bill Bell and Donald Martin for comments that improved this paper.

REFERENCES

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, Wiley: New York

Bell, W. R. (1984) "Seasonal Decomposition of Deterministic Effects," Census Bureau, Stat. Res. Div. Research Report RR84-1, *http://www.census.gov/srd/papers/pdf/rr84-1.pdf*

Bell, W. R., and Hillmer, S. C. (1983) "Modeling Time Series with Calendar Variation," *Journal of the American Statistical Association,* **78**, 526–534.

Findley, D. F. (1999) "Akaike's Information Criterion II," in *Encyclopedia of Statistical Science Update Volume 3,* eds. S. Kotz, C. B. Read, and D. R. Banks. New York: Wiley, 2-6.

Findley D. F., Monsell, B. C., Bell, W. R., Otto, M. C., and Chen, B.-C. (1998), "New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program," *Journal of Business and Economic Statistics*, 16, 127-177.

Hurvich, C. M. and Tsay, C. L. (1989), "Regression and Time Series Modeling in Small Samples," *Biometrika*,76, 297-307.

Morris, N. D. and Pfeffermann, D. (1984), "A Kalman filter approach to the forecasting of monthly time series affected by moving festivals," *Journal of Time Series Analysis 5, 255-268.*

Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986) "Akaike Information Criterion Statistics," D. Reidel: Dordrecht.