Estimation of Parameters in Models for Cesium Beam Atomic Clocks

Peter V. Tryon* and Richard H. Jones**

National Bureau of Standards, Boulder, CO 80303

August 2, 1982

This paper is intended to serve as an introduction to the use of the Kalman filter in modeling atomic clocks and obtaining maximum likelihood estimates of the model paramaters from data on an ensemble of clocks. Tests for the validity of the model and confidence intervals for the parameter estimates are discussed. Techniques for dealing with unequally spaced and partially or completely missing multivariate data are described. The existence of deterministic frequency drifts in clocks is established and estimates of the drifts are obtained.

Key words: atomic clocks; Kalman filter; maximum likelihood; missing observations; random walks; state space; time series analysis; unequally spaced data.

1. Introduction

The recursive updating of least squares estimates that would later become a special case of the Kalman "filter" was apparently known to Gauss [1],[1] and published in the modern statistical literature by Plackett [2]. Kalman's contribution [3] was the generalization to dynamic systems. Kalman's filter has found immense application in diverse areas of engineering. In classical applications, recursive estimates are obtained of the "state" of the system but parameters appearing elsewhere in the state space representation for the system and the mathematical model for the system must be considered known.

In our application, Kalman's recursive equations allow us to compute the likelihood function for given values of parameters occurring anywhere in the state space representation. Nonlinear optimization techniques can then be used to find the maximum likelihood estimates. The recursive residuals, or innovations, may be examined to judge the adequacy-of-fit of the model, and generalized likelihood ratio tests may be used to test the significance of model parameters such as frequency drifts and obtain confidence intervals for the estimated parameters. Kalman filter techniques make it quite easy to deal with unequally spaced and completely or partially missing multivariate data. We describe these methods in this paper.

The next article in this issue, "Estimating Time from Atomic Clocks," by Jones and Tryon [4] describes statistical procedures for detecting clock errors and allowing for the insertion or deletion of clocks in the ensemble during the parameter estimation process. It also describes the development of a time scale algorithm based on the model developed in this paper.

2. Models For Atomic Clocks

A cesium atomic clock is a feedback control device whose frequency locks onto the fundamental resonance of the cesium atom at (ideally) 9,192,631,770 Hz, which defines the second. Frequency bias due to fundamental instrumentation limitations and environmental effects are determined by calibration against primary frequency standards. Stochastic fluctuations arise from shot noise in the cesium beam and the probabilistic nature of quantum

mechanical transition rates which cause white noise in the frequency of the clock which is integrated into a random walk in time. In addition, empirical studies have demonstrated that frequency wanders independently as a random walk, introducing the integral of a random walk into time.

Finally, there is considerable international debate over the inclusion of linear drift in frequency to account for break-in and aging. Such drifts could be strictly constant over a long time period, or allowed to change slowly as a third random walk. The methods developed in this paper provide, for the first time, valid statistical tests of these hypotheses.

The proposed stochastic model for a clock's behavior is

$$
\begin{aligned}
x(t) &= x(t-1) + \delta(t)\,y(t-1) + \tfrac{1}{2}\,\delta^2(t)w(t-1) + \varepsilon(t) \\
y(t) &= y(t-1) + \phantom{\tfrac{1}{2}\,\delta^2(t)} \delta(t)w(t-1) + \eta(t) \\
w(t) &= \phantom{x(t-1) + \delta(t)\,y(t-1) + \tfrac{1}{2}\,\delta^2(t)} w(t-1) + \alpha(t)
\end{aligned}
\tag{2.1}
$$

where

$x(t)$ is the clock's deviation from "perfect" time at sample point t in nanoseconds;

$\delta(t)$ is the time interval in days between sample points t and $t-1$;

$y(t)$ is the clock's frequency deviation from perfect frequency at sample point t in nanoseconds/day;

$w(t)$ is the drift in frequency in nanoseconds/day^2;

$\varepsilon(t)$, $\eta(t)$, and $\alpha(t)$ are mutually independent white zero mean Gaussian random variables with standard deviations $\sqrt{\delta(t)}\sigma_\varepsilon$, $\sqrt{\delta(\tau)}\sigma_\eta$, and $\sqrt{\delta(t)}\sigma_\alpha$, respectively, in nanoseconds, nanoseconds/day, and nanoseconds/day^2. The $\sqrt{\delta(t)}$ arises because the variance of a random walk is proportional to the time interval.

If $\sigma_\alpha = 0$, $w(t)$ is a constant, w, representing a deterministic linear trend. If both $\sigma_\alpha = 0$ and $w = 0$, the model is drift-free. The purpose of this study is to evaluate the validity of these models, and to obtain generalized likelihood ratio tests for the significance of the drift parameters and maximum likelihood estimates of the parameters σ_ε, σ_η, and w (or σ_α, if appropriate), for each clock in the ensemble.

3. The Kalman State Space Model for a Clock Ensemble

The combination of clocks into a Kalman state space model will be explained by considering the special case of a three-clock model. The *state transition equation* is

$$
\begin{bmatrix}
x_1(t) \\ y_1(t) \\ w_1(t) \\ x_2(t) \\ y_2(t) \\ w_2(t) \\ x_3(t) \\ y_3(t) \\ w_3(t)
\end{bmatrix}
=
\begin{bmatrix}
1 & \delta(t) & \tfrac{1}{2}\delta^2(t) & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & \delta(t) & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & \delta(t) & \tfrac{1}{2}\delta^2(t) & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & \delta(t) & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & \delta(t) & \tfrac{1}{2}\delta^2(t) \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \delta(t) \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
x_1(t-1) \\ y_1(t-1) \\ w_1(t-1) \\ x_2(t-1) \\ y_2(t-1) \\ w_2(t-1) \\ x_3(t-1) \\ y_3(t-1) \\ w_3(t-1)
\end{bmatrix}
+
\begin{bmatrix}
\varepsilon_1(t) \\ \eta_1(t) \\ \alpha_1(t) \\ \varepsilon_2(t) \\ \eta_2(t) \\ \alpha_2(t) \\ \varepsilon_3(t) \\ \eta_3(t) \\ \alpha_3(t)
\end{bmatrix}
$$

where subscripts indicate the clocks.

4

In more compact form this is

$$X(t) = \Phi(t)\, X(t\text{–}1) + U(t) \; , \; t = 1, 2, ..., \tag{3.1}$$

where $X(t)$ is the state vector and $\Phi(t)$ is the state transition matrix. The stochastic elements of the transition from one state to the next are contained in $U(t)$.

The clocks are never read directly, but, using sophisticated electronics, the time differences between clocks can be read with a precision of less than a nanosecond. "Perfect" time, which we can never observe, cancels out, leaving an observation of the clock error differences. $z_1(t) = x_1(t) - x_2(t)$ and $z_2(t) = x_1(t) - x_3(t)$. We write the *observation equation* in matrix form as

$$
\begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix}
=
\begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \end{bmatrix}
\begin{bmatrix} x_1(t) \\ y_1(t) \\ w_1(t) \\ x_2(t) \\ y_2(t) \\ w_2(t) \\ x_3(t) \\ y_3(t) \\ w_3(t) \end{bmatrix}
+
\begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix}
$$

or

$$Z(t) = H(t)\, X(t) + V(t) \; , \; t = 1, 2, ... \tag{3.2}$$

$Z(t)$ is the vector of observed clock time differences and $H(t)$ is the observation matrix. $V(t)$ is a vector of observational errors. We write $H(t)$ as a time-varying matrix because, in the multivariate case, partially missing data is represented by deleting rows from $H(t)$ thus changing the dimension of $Z(t)$, $H(t)$, and $V(t)$. In some cases it is possible to recover from a problem with the reference clock by defining a new (temporary) reference and changing the patterns of $\pm$ 1 and zeros in $H(t)$. This is described in detail in Jones and Tryon [4].

For small $\delta(t)$ the covariance matrix of the transition errors, $U(t)$, is of the diagonal form $\delta(t)Q$ where Q is

$$
Q =
\begin{bmatrix}
\sigma_{\epsilon 1}^{2} & & & & & & & & \\
& \sigma_{\eta 1}^{2} & & & & & & & \\
& & \sigma_{\alpha 1}^{2} & & & & & 0 & \\
& & & \sigma_{\epsilon 2}^{2} & & & & & \\
& & & & \sigma_{\eta 2}^{2} & & & & \\
& & & & & \sigma_{\alpha 2}^{2} & & & \\
& & & & & & \sigma_{\epsilon 3}^{2} & & \\
& 0 & & & & & & \sigma_{\eta 3}^{2} & \\
& & & & & & & & \sigma_{\alpha 3}^{2}
\end{bmatrix}
\tag{3.3}
$$

Note that the matrix Q contains all the parameters to be estimated except w.

If the time offset of each clock is measured to the nearest nanosecond and is otherwise error free, the observational error vector $V(t)$ has covariance matrix

$$
R(t) =
\begin{bmatrix} 1/12 & 0 \\ 0 & 1/12 \end{bmatrix},
\tag{3.4}
$$

since the variance of a uniformly distributed random variable from –0.5 to 0.5 is 1/12. The dimension of $R(t)$ will also vary with time if there are partially missing data.

4. The Kalman Recursion

First, we adopt some notation. By $X(t|s)$, $s \leqslant t$ we denote the best estimate of the state vector at sample time t, based on all the data up to and including sample time s. We define $P(t|s)$ to be the covariance matrix of $X(t|s)$. Similarly, $Z(t|s)$ will be the predicted observation at time t based on all the data up to time s.

The recursion proceeds through the following steps (Kalman [3]).

1. Given the state vector and its covariance matrix at any sample time, say t, we can predict the state vector one observation interval into the future using the transition eq (3.1). The prediction is

$$X(t+1|t) = \Phi(t+1) \, X(t|t) \tag{4.1}$$

since the expected value of the stochastic component of the transition is zero. The covariance matrix of the prediction is

$$P(t+1|t) = \Phi(t+1) \, P(t|t) \, \Phi'(t+1) + \delta(t+1)Q, \tag{4.2}$$

where the first term on the right is the covariance matrix of the linear operation in the state transition and the second term is the covariance of the additive stochastic component.

2. From the observation matrix and the predicted state we can predict the next observation.

$$Z(t+1|t) = H(t+1) \, X(t+1|t). \tag{4.3}$$

3. We next compare the prediction with the actual data (when it becomes available in a real-time application). The difference is the recursive residual, or innovation,

$$I(t+1) = Z(t+1) - Z(t+1|t), \tag{4.4}$$

which has covariance matrix

$$C(t+1) = H(t+1) \, P(t+1|t) \, H'(t+1) + R(t+1). \tag{4.5}$$

The residuals are critically important. Later, they will be used for checking the fit of the model and estimating the parameters. For now, we will use the information in the residual to complete the recursion.

4. Let

$$\Delta(t+1) = P(t+1|t) \, H'(t+1) \, C^{-1}(t+1). \tag{4.6}$$

This is sometimes called the Kalman gain.

Using the Kalman gain we can update the predicted state and its covariance

$$X(t+1|t+1) = X(t+1|t) + \Delta(t+1) \, I(t+1) \tag{4.7}$$

and

$$P(t+1|t+1) = P(t+1|t) - \Delta(t+1) \, H(t+1) \, P(t+1|t). \tag{4.8}$$

This completes the recursion.

5. Maximum Likelihood Modeling and Estimation

In this section we will compare the classic use of the Kalman filter with our method of application. In the classical application, unknown parameters to be estimated can appear only in the state vector. All other quantities which appear in the recursion must be known. A single pass of the recursion through the data (often in real time) provides the latest estimate of the state vector and its covariance at each point in time. This is exactly how the Kalman filter is applied to forming a time scale. Given the model and estimates of the parameters, the recursion gives the updated state vector after each new observation. The first, fourth, and seventh, etc., elements of the state vector are estimates of the clock errors, from which, along with their uncertainties from the state covariance matrix, we can estimate time.

Our interest, however, is in evaluating how well the clock model fits real data, and in estimating the parameters. To do this, we make use of two facts concerning the residuals, $I(t)$, $t = 1, 2, ..., N$.

1. If the model is correct, and the recursion is run through the data with the correct values of the unknown parameters, the residuals will form Gaussian white noise series. Standard statistical tests can then be used to determine if the residuals are in fact Guassian white noise. If not, the structure of the residuals can often give some clues as to the defects in the model.

2. From the residuals and their covariance matrix, obtained by running the recursion with *any* fixed values of the unknown parameters, the likelihood function for those parameter values can be computed. Specifically, the –2 ln likelihood function (ignoring the additive constant) is

$$L = \sum_t [\ln |C(t)| + I'(t) \, C^{-1}(t) \, I(t)]. \tag{5.1}$$

The maximum likelihood estimates of the parameters are those for which L is minimized. Starting from initial guesses, we can use iterative nonlinear function minimization methods to find the estimates. Evaluating the function L for new trial values of the parameters requires one pass of the Kalman recursion through the data. Note that the parameters to be estimated may appear anywhere in the structure of the Kalman recursion.

6. Computational Procedures

It is not necessary to invert the residual covariance matrix, $C(t)$, when computing the likelihood function or updating the predicted state vector and its covariance. If the upper triangular portion of $C(t+1)$ is augmented by $H(t+1)P(t+1|t)$, and $I(t+1)$ to form the partitioned matrix

$$[C(t+1) \mid H(t+1) \, P(t+1|t) \mid I(t+1)], \tag{6.1}$$

a single call to a Choleskey factorization routine (Graybill [5], p. 232) which factors

$$C(t+1) = T'(t+1) \, T(t+1), \tag{6.2}$$

where $T(t+1)$ is upper triangular, will replace the partitioned matrix by

$$[T(t+1) \mid [T'(t+1)]^{-1} \, H(t+1) \, P(t+1|t) \mid [T'(t+1)]^{-1} \, I(t+1)]. \tag{6.3}$$

If we call the parts of the new partitioned matrix

$$[T(t+1) \mid B(t+1) \mid D(t+1)], \tag{6.4}$$

the updated state and state covariance matrices become

$$X(t+1|t+1) = X(t+1|t) + B'(t+1) D(t+1) \tag{6.5}$$

and

$$P(t+1|t+1) = P(t+1|t) - B'(t+1) B(t+1). \tag{6.6}$$

The −2 ln likelihood function becomes

$$L = \sum_t \left(\sum_i \ln t^2_{ii}(t) + D'(t) D(t) \right) \tag{6.7}$$

where $t_{ii}(t)$ are the diagonal elements of $T(t)$.

Missing data are easily handled by the Kalman filter procedure. However, there are two distinct situations. First, all the data may be missing at one or several successive observation times, and second, occurring only in multivariate data, only part of the data may be available on a given occasion. For example, one or more, but not all, of the clock differences are missing on that occasion.

The first case is most easily dealt with. In the present model, the unequal spacing feature can account for any size gap in the data by proper choice of $\delta(t)$. This, however, is a feature of random walk based models and may not apply to other models. A more general procedure for missing data is the following: Suppose $k \geq 1$ successive observations are missing. Simply repeat step one in the recursion k additional times to obtain the predicted state, $X(t+k+1|t)$, and its covariance, $P(t+k+1|t)$, for the next observation time for which there is data. Then proceed with the remainder of the recursion. There is no residual and no contribution to the likelihood function when there is no data (Jones [6]).

In the second case, partial loss of data, i.e., missing elements of the data vector $Z(t)$, can be represented by eliminating the corresponding rows of the observation matrix, $H(t)$. Similarly, rows of the observational error vector $V(t)$ and rows and columns of the error covariance matrix $R(t)$ causes no problem with the remaining steps of the recursion, which proceed normally. The information available in the data is propagated through the recursion to correctly update the state vector prediction and its estimated covariance matrix.

The matrix operations in the recursion are, for reasons of efficiency, best coded to take advantage of the specific structure (many zeros) of each matrix. Using available general purpose subroutines to perform the required matrix operations would increase the cost of computer resources enormously. These calculations can be adapted to a time-varying $H(t)$. One approach is to use indicator variables for missing data to form the various reduced matrices directly. Another approach is to note that pre- and/or post-multiplying by $H(t)$ with rows missing is equivalent to performing the operation with the original H matrix and then eliminating the appropriate rows and/or columns of the product matrix and closing it up. For example, the residual covariance matrix

$$C(t+1) = H(t+1) P(t+1|t) H'(t+1) + R(t+1)$$

with, say, the second and fourth elements of the data vector missing can be obtained by using the original H and R but then eliminating the second and fourth rows and columns from $C(t+1)$. It is straightforward to write general purpose subroutines to eliminate rows and/or columns of a matrix and close up the spaces. Again, this can be done in response to indicator variables denoting missing observations.

To start the recursion the initial state $X(0|0)$ and its covariance $P(0|0)$ are needed. Both depend on the application. The time states could be set to zero and the clocks set to agree with International Atomic Time as defined by the *BIH*. The corresponding variances on the diagonal of $P(0|0)$ would be determined by the accuracy of the intercomparison. If a strictly internal application such as parameter estimation or a time interval measurement was desired, the time states could be set to agree with the difference readings at $t = 0$, with observational error variances set to the appropriate value for roundoff error. The frequency states might be measured by intercomparison with the frequency standards and the variance determined from the measurement uncertainty. When a measurement is not possible a best guess at the frequency offset can be used with a corresponding guess at its uncertainty used for the variance.

7. Data Analysis

The data used for this analysis were collected over a period of 333 days beginning on February 16, 1979, during which seven clocks listed below were in essentially uninterrupted normal operation. Clock differences were recorded daily to the nearest nanosecond.

Clock Number	Identification
601 (Reference)	HP model 5061A option 004 serial #601
167	HP model 5060A serial #167
137	HP model 5060A serial #137
1316	HP model 5061A option 004 serial #1316
323	HP model 5061A 004 serial #323
324	HP model 5061A 004 serial #324
8	Frequency and Time System model 400 serial #8

In addition to occasional variability in the time of data collection, two successive days were missing entirely. On three other occasions, single clocks had obvious read errors which when discarded resulted in partially missing data. Several dozen other observations which had been flagged as possible errors by the ad hoc procedure then in use were examined visually, judged to be not serious, and retained in the analysis.

Table 1 gives the results of the computations. First we note that the inclusion of deterministic trends (model II) results in a drop of 41 in the –2 ln likelihood function value over the no-trend model (model I). The reduction is distributed as $X^2(6)$ under the hypothesis of model I, so this is highly significant. There is less than one chance in a thousand that this is a spurious result. Three of the seven clocks had significant trends. The inclusion of random walk drifts (model III), however, makes no improvement. Furthermore, the estimated values for the σ_a's are all approximately zero and the estimated initial values of the drift, $w(0)$, are about the same as the deterministic drifts. The conclusion is that several frequency trends are present in this data and that they are constant over the year, rather than wandering as a random walk. It should be noted that one of the drifting clocks had just begun operation while the remaining two were nearing their end of life and failed shortly after the test period. The remaining clocks were in their "middle years." The standard errors of the parameter estimates given in the table are obtained from the estimated covariance matrix as approximated by twice the inverse of the Hessian of the –2 ln likelihood function, L.

In model II it is necessary to constrain the solution so that the sum of the drifts is zero. This means that we cannot detect a common drift from clock difference readings. Because of this constraint, a standard error for clock #8 is not available. Due to the large cost of computing the 28 x 28 Hessian for model III, which we have rejected anyway, the standard errors of the estimates have not been computed.

Figures 1a-f are integrated periodograms of the six series of residuals. The parallel lines represent five percent significance test limits for white noise. That is, only five percent of actual white noise series so tested will wander outside the lines. In this data set, only one pair of clocks (601 minus 167 two) show a slight deviation from white noise. We conclude that this model fits very well. Figures 2a-f are histograms of the six residual series. Also given are three different tests for normality. All the series are well approximated by the normal (Gaussian) distribution which verifies that the assumed likelihood function was reasonable.

We note in conclusion that, having selected the constant drift model, we could reduce the dimension of the state space model by one-third by rewriting the clock model as

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} 1 & \delta(t) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x(t-1) \\ y(t-1) \end{bmatrix} + \begin{bmatrix} \delta^2(t)/2 \\ \delta(t) \end{bmatrix} w + \begin{bmatrix} \varepsilon(t) \\ \eta(t) \end{bmatrix}$$

The state prediction in the first step of the recursion then becomes

$$X(t+1|t) = \Phi(t+1) X(t|t) + A(t+1)$$

9

where

$$A(t) = \begin{bmatrix} \dfrac{\delta^2(t)\ w_1}{2} \\[1em] \delta(t)w_1 \\[1em] \dfrac{\delta^2(t)\ w_2}{2} \\[1em] \delta(t)w_2 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix},$$

and $X(t)$ and $\Phi(t)$ are redefined from the clock model in the obvious way. This model would be suitable for estimation purposes but not for a time scale algorithm. With the drift in the state vector we can enter the uncertainty of its estimation in the initial state covariance matrix. It is then propagated through to the uncertainty in time. The simplified model above assumes w is known without error.

TABLE 1. Results of data analysis.

Clock Number	Model I No Trend $L = 10609.1$			Model II Deterministic Trend $L = 10567.8$			Model III Random Walk Drift $L = 10567.7$	
	Param. Est.*		S.E	Param. Est.*		S.E.	Param. Est.*	
601	$\hat{\sigma}_\epsilon$ =	7.42	.32	$\hat{\sigma}_\epsilon$ =	7.46	.32	$\hat{\sigma}_\epsilon$ =	7.47
	$\hat{\sigma}_\eta$ =	.86	.24	$\hat{\sigma}_\eta$ =	.44	.26	$\hat{\sigma}_\eta$ =	.48
				$\hat{w}$ =	.152	.038	$\hat{w}(0)$ =	.152
							$\hat{\alpha}_a$ =	4.1×10^{-5}
167	$\hat{\sigma}_\epsilon$ =	13.45	.56	$\hat{\sigma}_\epsilon$ =	13.45	.56	$\hat{\sigma}_\epsilon$ =	13.46
	$\hat{\sigma}_\eta$ =	1.15	.39	$\hat{\sigma}_\eta$ =	1.11	.36	$\hat{\sigma}_\eta$ =	1.07
				$\hat{w}$ =	.052	.061	$\hat{w}(0)$ =	.054
							$\hat{\sigma}_a$ =	3.7×10^{-5}
137	$\hat{\sigma}_\epsilon$ =	10.03	.45	$\hat{\sigma}_\epsilon$ =	10.04	.45	$\hat{\sigma}_\epsilon$ =	10.06
	$\hat{\sigma}_\eta$ =	1.71	.36	$\hat{\sigma}_\eta$ =	1.60	.36	$\hat{\sigma}_\eta$ =	1.57
				$\hat{w}$ =	.179	.081	$\hat{w}(0)$ =	.178
							$\hat{\sigma}_a$ =	2.2×10^{-4}
1316	$\hat{\sigma}_\epsilon$ =	3.61	.24	$\hat{\sigma}_\epsilon$ =	3.62	.25	$\hat{\sigma}_\epsilon$ =	3.59
	$\hat{\sigma}_\eta$ =	1.29	.24	$\hat{\sigma}_\eta$ =	1.36	.24	$\hat{\sigma}_\eta$ =	1.38
				$\hat{w}$ =	−.017	.070	$\hat{w}(0)$ =	−.015
							$\hat{\sigma}_a$ =	1.3×10^{-4}
323	$\hat{\sigma}_\epsilon$ =	3.27	.24	$\hat{\sigma}_\epsilon$ =	3.53	.22	$\hat{\sigma}_\epsilon$ =	3.52
	$\hat{\sigma}_\eta$ =	1.54	.21	$\hat{\sigma}_\eta$ =	.73	.20	$\hat{\sigma}_\eta$ =	.70
				$\hat{w}$ =	−.313	.046	$\hat{w}$ =	−.314
							$\hat{\sigma}_\epsilon$ =	4.9×10^{-5}
324	$\hat{\sigma}_\epsilon$ =	3.30	.25	$\hat{\sigma}_\epsilon$ =	3.30	.25	$\hat{\sigma}_\epsilon$ =	3.30
	$\hat{\sigma}_\eta$ =	1.42	.23	$\hat{\sigma}_\eta$ =	1.40	.22	$\hat{\sigma}_\eta$ =	1.41
				$\hat{w}$ =	.035	.072	$\hat{w}(0)$ =	.035
							$\hat{\sigma}_\epsilon$ =	7.9×10^{-5}
8	$\hat{\sigma}_\epsilon$ =	9.08	.43	$\hat{\sigma}_\epsilon$ =	9.09	.43	$\hat{\sigma}_\epsilon$ =	9.09
	$\hat{\sigma}_\eta$ =	2.68	.39	$\hat{\sigma}_\eta$ =	2.65	.39	$\hat{\sigma}_\eta$ =	2.66
				$\hat{w}$ =	−.088	−	$\hat{w}(0)$ =	−.090
							$\hat{\sigma}_r$ =	3.9×10^{-5}

*Units are in nanoseconds.

8. Conclusion

The Kalman filter is a powerful tool for maximum likelihood model fitting and parameter estimation in time series analysis. We have established the validity of the proposed models for clock behavior and obtained precise estimates of clock parameters. The existence of deterministic frequency drifts over a period of a year has been demonstrated in clocks at the beginning and end of their life span.

The primary disadvantage of these methods is the cost of function evaluation in the nonlinear optimization process. The analysis of a year of data on seven clocks takes 10 minutes on a CDC Cyber 170/750 computer and 10 hours on the DEC 11/70. Use of state-of-the-art optimization codes is essential. We use an optimization package especially designed for maximum likelihood applications. It is adapted from the package written by Weiss [7], based on algorithms by Dennis and Schnabel [8]. The code used on the CDC Cyber 170/750 is part of a preliminary version of the National Bureau of Standard's STARPAC library [9].

9. References

[1] Gauss C. F. Theoria combinationis observationum erroribus minimis obnoxiae. Werke, 4. Gottingen. 1821. (Collected works 1873)
[2] Plackett, R. L. Some theorems in least squares. Biometrika, 19: 149-157; 1950.
[3] Kalman, R. E. A new approach to linear filtering and prediction problems. ASME Transactions, Part D, Journal of Basic Engineering, 82: 35-45; 1960.
[4] Jones, R. H.; Tryon, P. V. Estimating time from atomic clocks. Second Symposium on Atomic Time Scale Algorithms, June 23-25, 1982, Boulder, CO.
[5] Graybill, F. A. *Theory and application of the linear model. North Scituate:* Duxbury Press; 1976.
[6] Jones, R. H. Maximum likelihood fitting of ARMA models to time series with missing observations. Technometics, 22: 389-395; 1980.
[7] Weiss, B. E. A modular software package for solving unconstrained nonlinear optimization problems. Master's thesis, University of Colorado at Boulder; 1980.
[8] Dennis, J. E.; Schnabel, R. B. *Numerical methods for unconstrained optimization and nonlinear equations.* New Jersey: Prentice Hall; 1982.
[9] Donaldson, J. R.; Tryon, P. V. STARPAC: The standards time series and regression package. Unpublished—Documentation available from Janet P. Donaldson, NBS-714, 325 Broadway, Boulder, CO 80303.

Figures 1a-f and 2a-f follow.

FIGURE 1a. Residuals from clock 601 minus clock 167.

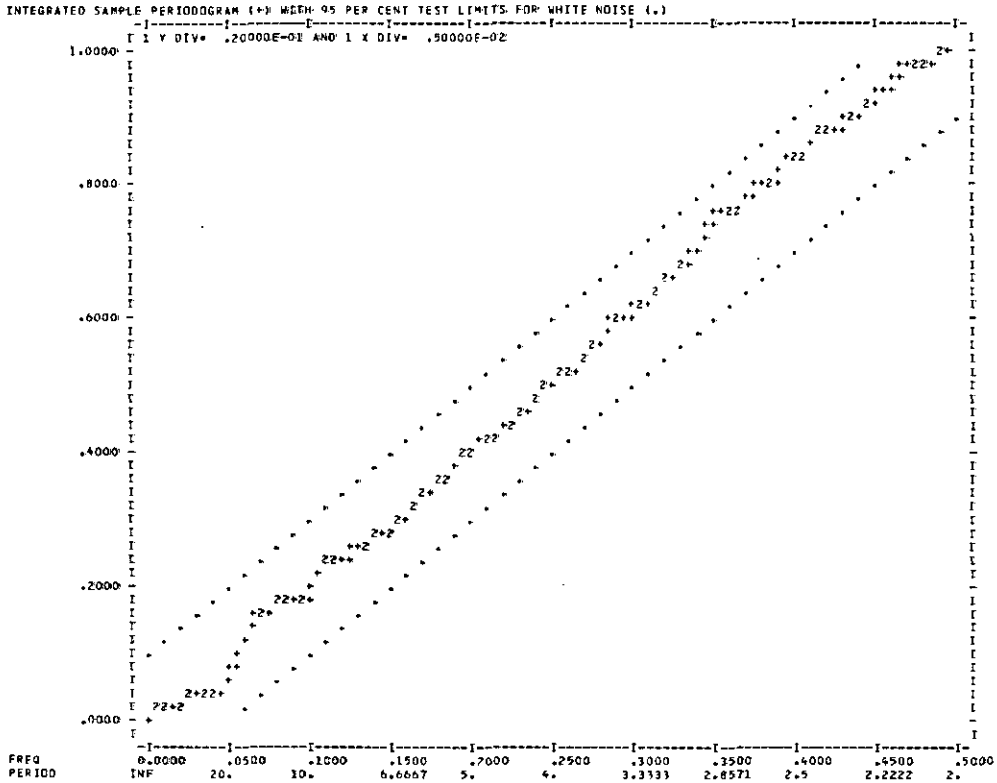FIGURE 1b. Residuals from clock 601 minus clock 137.

12

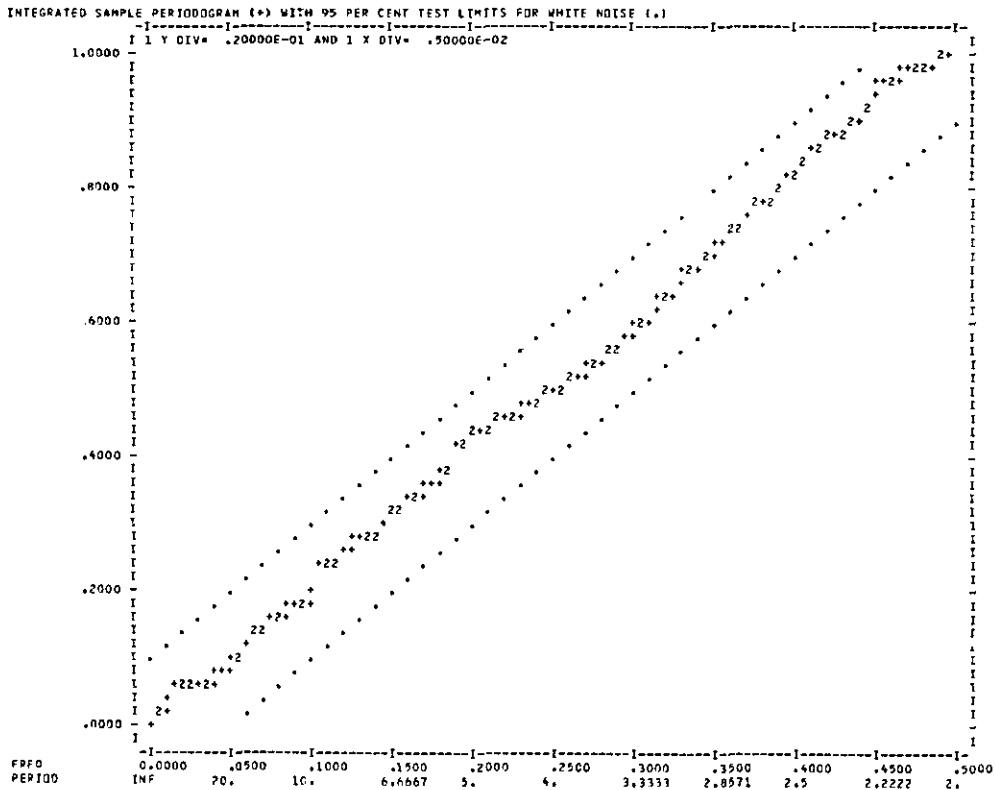FIGURE 1c. Residuals from clock 601 minus clock 1316.

FIGURE 1d. Residuals from clock 601 minus clock 323.
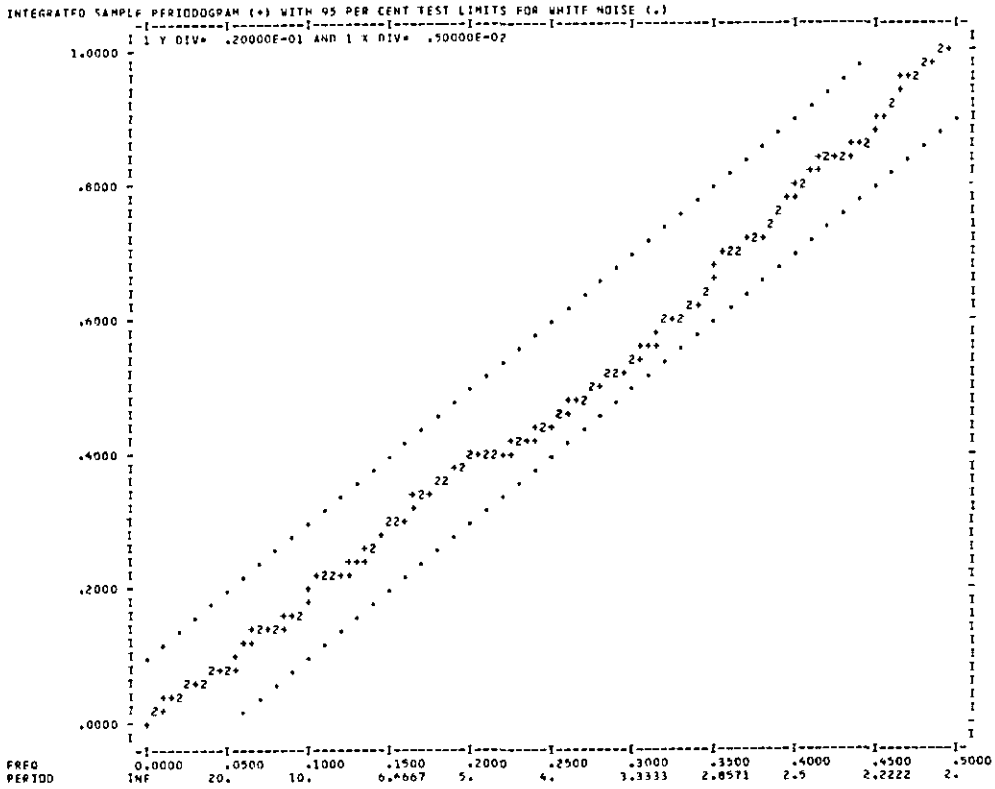
13

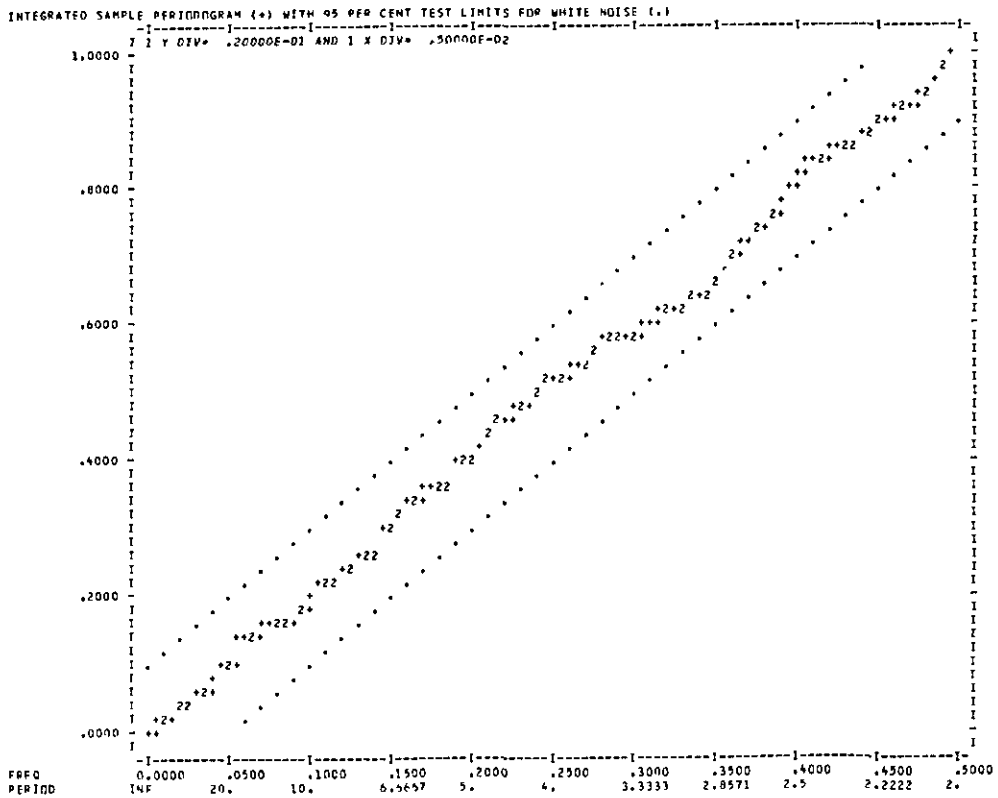FIGURE 1e. Residuals from clock 601 minus clock 324.

FIGURE 1f. Residuals from clock 601 minus clock 8.

14

```
                NUMBER OF OBSERVATIONS  =           332
                MINIMUM OBSERVATION     = -.43333971BE+02
                MAXIMUM OBSERVATION     =  .428583010E+02

                HISTOGRAM LOWER BOUND   = -.450000000E+02
                HISTOGRAM UPPER BOUND   =  .450000000E+02

                OBSERVATIONS USED       =           332        25 PCT TRIMMED MEAN = -.333716985E+00
                MIN. OBSERVATION USED   = -.433339718E+02       STANDARD DEVIATION  =  .160738390E+02
                MAX. OBSERVATION USED   =  .428583010E+02       MEAN DEV./STD. DEV. =  .805496712E+00
                MEAN VALUE              = -.153504202E+00       B1**0.5             =  .145368914E-02
                MEDIAN VALUE            = -.691321082E+00       B2                  =  .279156726E+01
```

FOR A NORMAL DESTRIBUTION, THE VALUES (MEAN DEVIATION/STANDARD DEVIATION), B1**0.5 AND B2 ARE APPROXIMATELY
0.8, 0.0 AND 3.0, RESPECTIVELY. TO TEST THE NULL HYPOTHESIS OF NORMALITY, SEE TABLES OF CRITICAL VALUES
PP. 207-208, BIOMETRIKA TABLES FOR STATISTICIANS, VOL. 1. SEE PP. 67-68 FOR A DISCUSSION OF THESE TESTS.

```
  INTERVAL   CUM.   1-CUM.  CELL   NO.                          CELL FRACTION
  MID POINT  FRACT. FRACT.  FRACT. OBS.  0.00      .05      .10      .15      .20      .25
  -------------------------------------+---------+---------+---------+---------+---------+
  -40.0000   .015   1.000   .015    5    +++
  -30.0000   .042    .985   .027    9    +++++
  -20.0000   .181    .958   .139   46    +++++++++++++++++++++++++++
  -10.0000   .395    .819   .214   71    +++++++++++++++++++++++++++++++++++++++++++
   0.        .636    .605   .241   80    ++++++++++++++++++++++++++++++++++++++++++++++++
   10.0000   .816    .364   .181   60    ++++++++++++++++++++++++++++++++++++
   20.0000   .931    .184   .114   38    +++++++++++++++++++++++
   30.0000   .991    .069   .060   20    ++++++++++++
   40.0000  1.000    .009   .009    3    ++
```

FIGURE 2a. Histogram—residuals from clock 601 minus clock 167.

```
                NUMBER OF OBSERVATIONS  =           332
                MINIMUM OBSERVATION     = -.43333971BE+02
                MAXIMUM OBSERVATION     =  .428583010E+02

                HISTOGRAM LOWER BOUND   = -.450000000E+02
                HISTOGRAM UPPER BOUND   =  .450000000E+02

                OBSERVATIONS USED       =           332        25 PCT TRIMMED MEAN = -.333716985E+00
                MIN. OBSERVATION USED   = -.433339718E+02       STANDARD DEVIATION  =  .160738390E+02
                MAX. OBSERVATION USED   =  .428583010E+02       MEAN DEV./STD. DEV. =  .805496712E+00
                MEAN VALUE              = -.153504202E+00       B1**0.5             =  .145368914E-02
                MEDIAN VALUE            = -.691321082E+00       B2                  =  .279156726E+01
```

FOR A NORMAL DESTRIBUTION, THE VALUES (MEAN DEVIATION/STANDARD DEVIATION), B1**0.5 AND B2 ARE APPROXIMATELY
0.8, 0.0 AND 3.0, RESPECTIVELY. TO TEST THE NULL HYPOTHESIS OF NORMALITY, SEE TABLES OF CRITICAL VALUES
PP. 207-208, BIOMETRIKA TABLES FOR STATISTICIANS, VOL. 1. SEE PP. 67-68 FOR A DISCUSSION OF THESE TESTS.

```
  INTERVAL   CUM.   1-CUM.  CELL   NO.                          CELL FRACTION
  MID POINT  FRACT. FRACT.  FRACT. OBS.  0.00      .05      .10      .15      .20      .25
  -------------------------------------+---------+---------+---------+---------+---------+
  -40.0000   .015   1.000   .015    5    +++
  -30.0000   .042    .985   .027    9    +++++
  -20.0000   .181    .958   .139   46    +++++++++++++++++++++++++++
  -10.0000   .395    .819   .214   71    +++++++++++++++++++++++++++++++++++++++++++
   0.        .636    .605   .241   80    ++++++++++++++++++++++++++++++++++++++++++++++++
   10.0000   .816    .364   .181   60    ++++++++++++++++++++++++++++++++++++
   20.0000   .931    .184   .114   38    +++++++++++++++++++++++
   30.0000   .991    .069   .060   20    ++++++++++++
   40.0000  1.000    .009   .009    3    ++
```

FIGURE 2b. Histogram—residuals from clock 601 minus clock 137.

```
                NUMBER OF OBSERVATIONS  =           332
                MINIMUM OBSERVATION     = -.433339718E+02
                MAXIMUM OBSERVATION     =  .428583010E+02

                HISTOGRAM LOWER BOUND   = -.450000000E+02
                HISTOGRAM UPPER BOUND   =  .450000000E+02

                OBSERVATIONS USED       =           332        25 PCT TRIMMED MEAN = -.333716985E+00
                MIN. OBSERVATION USED   = -.433339718E+02       STANDARD DEVIATION  =  .160738390E+02
                MAX. OBSERVATION USED   =  .428583010E+02       MEAN DEV./STD. DEV. =  .805496712E+00
                MEAN VALUE              = -.153504202E+00       B1**0.5             =  .145368914E-02
                MEDIAN VALUE            = -.691321082E+00       B2                  =  .279156726E+01
```

FOR A NORMAL DESTRIBUTION, THE VALUES (MEAN DEVIATION/STANDARD DEVIATION), B1**0.5 AND B2 ARE APPROXIMATELY
0.8, 0.0 AND 3.0, RESPECTIVELY. TO TEST THE NULL HYPOTHESIS OF NORMALITY, SEE TABLES OF CRITICAL VALUES
PP. 207-208, BIOMETRIKA TABLES FOR STATISTICIANS, VOL. 1. SEE PP. 67-68 FOR A DISCUSSION OF THESE TESTS.

```
  INTERVAL   CUM.   1-CUM.  CELL   NO.                          CELL FRACTION
  MID POINT  FRACT. FRACT.  FRACT. OBS.  0.00      .05      .10      .15      .20      .25
  -------------------------------------+---------+---------+---------+---------+---------+
  -40.0000   .015   1.000   .015    5    +++
  -30.0000   .042    .985   .027    9    +++++
  -20.0000   .181    .958   .139   46    +++++++++++++++++++++++++++
  -10.0000   .395    .819   .214   71    +++++++++++++++++++++++++++++++++++++++++++
   0.        .636    .605   .241   80    ++++++++++++++++++++++++++++++++++++++++++++++++
   10.0000   .816    .364   .181   60    ++++++++++++++++++++++++++++++++++++
   20.0000   .931    .184   .114   38    +++++++++++++++++++++++
   30.0000   .991    .069   .060   20    ++++++++++++
   40.0000  1.000    .009   .009    3    ++
```

FIGURE 2c. Histogram—residuals from clock 601 minus clock 1316.

```
        NUMBER OF OBSERVATIONS    =        332
        MINIMUM OBSERVATION       = -.433339718E+02
        MAXIMUM OBSERVATION       =  .428583010E+02

        HISTOGRAM LOWER BOUND     = -.450000000E+02
        HISTOGRAM UPPER BOUND     =  .450000000E+02

        OBSERVATIONS USED         =        332        25 PCT TRIMMED MEAN = -.333716985E+00
        MIN. OBSERVATION USED     = -.433339718E+02    STANDARD DEVIATION   =  .160738390E+02
        MAX. OBSERVATION USED     =  .428583010E+02    MEAN DEV./STD. DEV.  =  .805496712E+00
        MEAN VALUE                = -.153504202E+00    B1**0.5              =  .145368914E-02
        MEDIAN VALUE              = -.691321082E+00    B2                   =  .279156726E+01
```

FOR A NORMAL DESTRIBUTION, THE VALUES (MEAN DEVIATION/STANDARD DEVIATION), B1**0.5 AND B2 ARE APPROXIMATELY
0.8, 0.0 AND 3.0, RESPECTIVELY. TO TEST THE NULL HYPOTHESIS OF NORMALITY, SEE TABLES OF CRITICAL VALUES
PP. 207-208, BIOMETRIKA TABLES FOR STATISTICIANS, VOL. 1. SEE PP. 67-68 FOR A DISCUSSION OF THESE TESTS.

```
  INTERVAL    CUM.   1-CUM.  CELL   NO.                         CELL FRACTION
  MID POINT  FRACT.  FRACT.  FRACT. OBS.  0.00     .05     .10     .15     .20     .25
 ------------------------------------------+---------+---------+---------+---------+---------+
  -40.0000    .015   1.000   .015    5    +++
  -30.0000    .042    .985   .027    9    +++++
  -20.0000    .181    .958   .139   46    +++++++++++++++++++++++++++++++
  -10.0000    .395    .819   .214   71    ++++++++++++++++++++++++++++++++++++++++++++++
   0.         .636    .605   .241   80    ++++++++++++++++++++++++++++++++++++++++++++++++++++
   10.0000    .816    .364   .181   60    ++++++++++++++++++++++++++++++++++++++++
   20.0000    .931    .184   .114   38    +++++++++++++++++++++++++
   30.0000    .991    .069   .060   20    +++++++++++++
   40.0000   1.000    .009   .009    3    ++
```

FIGURE 2d. Histogram—residuals from clock 601 minus clock 323.

```
        NUMBER OF OBSERVATIONS    =        332
        MINIMUM OBSERVATION       = -.433339718E+02
        MAXIMUM OBSERVATION       =  .428583010E+02

        HISTOGRAM LOWER BOUND     = -.450000000E+02
        HISTOGRAM UPPER BOUND     =  .450000000E+02

        OBSERVATIONS USED         =        332        25 PCT TRIMMED MEAN = -.333716985E+00
        MIN. OBSERVATION USED     = -.433339718E+02    STANDARD DEVIATION   =  .160738390E+02
        MAX. OBSERVATION USED     =  .428583010E+02    MEAN DEV./STD. DEV.  =  .805496712E+00
        MEAN VALUE                = -.153504202E+00    B1**0.5              =  .145368914E-02
        MEDIAN VALUE              = -.691321082E+00    B2                   =  .279156726E+01
```

FOR A NORMAL DESTRIBUTION, THE VALUES (MEAN DEVIATION/STANDARD DEVIATION), B1**0.5 AND B2 ARE APPROXIMATELY
0.8, 0.0 AND 3.0, RESPECTIVELY. TO TEST THE NULL HYPOTHESIS OF NORMALITY, SEE TABLES OF CRITICAL VALUES
PP. 207-208, BIOMETRIKA TABLES FOR STATISTICIANS, VOL. 1. SEE PP. 67-68 FOR A DISCUSSION OF THESE TESTS.

```
  INTERVAL    CUM.   1-CUM.  CELL   NO.                         CELL FRACTION
  MID POINT  FRACT.  FRACT.  FRACT. OBS.  0.00     .05     .10     .15     .20     .25
 ------------------------------------------+---------+---------+---------+---------+---------+
  -40.0000    .015   1.000   .015    5    +++
  -30.0000    .042    .985   .027    9    +++++
  -20.0000    .181    .958   .139   46    +++++++++++++++++++++++++++++++
  -10.0000    .395    .819   .214   71    ++++++++++++++++++++++++++++++++++++++++++++++
   0.         .636    .605   .241   80    ++++++++++++++++++++++++++++++++++++++++++++++++++++
   10.0000    .816    .364   .181   60    ++++++++++++++++++++++++++++++++++++++++
   20.0000    .931    .184   .114   38    +++++++++++++++++++++++++
   30.0000    .991    .069   .060   20    +++++++++++++
   40.0000   1.000    .009   .009    3    ++
```

FIGURE 2e. Histogram—residuals from clock 601 minus clock 324.

```
        NUMBER OF OBSERVATIONS    =        332
        MINIMUM OBSERVATION       = -.433339718E+02
        MAXIMUM OBSERVATION       =  .428583010E+02

        HISTOGRAM LOWER BOUND     = -.450000000E+02
        HISTOGRAM UPPER BOUND     =  .450000000E+02

        OBSERVATIONS USED         =        332        25 PCT TRIMMED MEAN = -.333716985E+00
        MIN. OBSERVATION USED     = -.433339718E+02    STANDARD DEVIATION   =  .160738390E+02
        MAX. OBSERVATION USED     =  .428583010E+02    MEAN DEV./STD. DEV.  =  .805496712E+00
        MEAN VALUE                = -.153504202E+00    B1**0.5              =  .145368914E-02
        MEDIAN VALUE              = -.691321082E+00    B2                   =  .279156726E+01
```

FOR A NORMAL DESTRIBUTION, THE VALUES (MEAN DEVIATION/STANDARD DEVIATION), B1**0.5 AND B2 ARE APPROXIMATELY
0.8, 0.0 AND 3.0, RESPECTIVELY. TO TEST THE NULL HYPOTHESIS OF NORMALITY, SEE TABLES OF CRITICAL VALUES
PP. 207-208, BIOMETRIKA TABLES FOR STATISTICIANS, VOL. 1. SEE PP. 67-68 FOR A DISCUSSION OF THESE TESTS.

```
  INTERVAL    CUM.   1-CUM.  CELL   NO.                         CELL FRACTION
  MID POINT  FRACT.  FRACT.  FRACT. OBS.  0.00     .05     .10     .15     .20     .25
 ------------------------------------------+---------+---------+---------+---------+---------+
  -40.0000    .015   1.000   .015    5    +++
  -30.0000    .042    .985   .027    9    +++++
  -20.0000    .181    .958   .139   46    +++++++++++++++++++++++++++++++
  -10.0000    .395    .819   .214   71    ++++++++++++++++++++++++++++++++++++++++++++++
   0.         .636    .605   .241   80    ++++++++++++++++++++++++++++++++++++++++++++++++++++
   10.0000    .816    .364   .181   60    ++++++++++++++++++++++++++++++++++++++++
   20.0000    .931    .184   .114   38    +++++++++++++++++++++++++
   30.0000    .991    .069   .060   20    +++++++++++++
   40.0000   1.000    .009   .009    3    ++
```

FIGURE 2f. Histogram—residuals from clock 601 minus clock 8.

16