Isolating Cyclical Patterns in Irregular Time Series Data

A. Stan Hurn
The University of Glasgow

and

A. David McDonald
CSIRO Division of Fisheries
Hobart

Abstract Time series data which are observed at irregular time intervals often arise in economics and the bio-sciences. Existing methods for modelling these data have focused on the discretisation of continuous processes. A method is proposed for fitting cyclical components to irregular time-series data based on the continuous-discrete Kalman filter which incorporates numerical integration of the differential equations describing the model. The method is applied to seawater temperature data and empirical sampling distributions for parameter estimators are enumerated. The supporting sampling distributions suggest that the method yields estimates which have satisfactory statistical properties.

1. INTRODUCTION

The problem of estimating the parameters of dynamic economic and biophysical processes when data are sampled at irregular time intervals has long been of interest to applied researchers. Irregular data arise commonly in the monitoring of biophysical processes as a result of sampling, equipment failure and random events. Irregular economic time series are becoming more common as a result of increased availability of reliable, high-frequency data. Exchange-rate markets, for example, give rise to continuously-generated data that are reported at irregular discrete intervals. Standard estimation methods become less precise when data are not spaced evenly.

In this paper we explore structural time series methods for estimating cyclical components in time series when data are spaced irregularly. We propose an estimation technique based on the likelihood function recovered from the prediction errors of a continuous-discrete filtering problem, where the differential equations comprising the filter are integrated numerically. As the method is entirely numerical it has application beyond the simple structural time series paradigm we adopt here. We demonstrate the resulting technique using monthly seawater temperate data and report empirical sampling distributions for the parameter estimates.

2. A STRUCTURAL TIME SERIES MODEL FOR ISOLATING CYCLICAL PATTERNS

Structural time series models enable time-series data to be modelled by cyclical components and/or other unobserved components as dictated by economic theory or the structural characteristics of the data. The literature on structural time series models is relatively well developed (see Harvey [1989] for the original exposition and Harvey et al., [1992] and McDonald and Hurn [1994] for applications) and the treatment offered here will be relatively terse.

In this paper attention will be restricted to a simple additive-cycle model in continuous time. The model takes the form

$$\begin{bmatrix} d\mu_t \\ d\psi_t \\ d\psi_t^* \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \ln(\alpha) & \lambda \\ 0 & -\lambda & \ln(\alpha) \end{bmatrix} \begin{bmatrix} \mu_t \\ \psi_t \\ \psi_t^* \end{bmatrix} dt + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} dz_\mu \\ dz_\psi \\ dz_\psi{}^* \end{bmatrix} \quad (1)$$

with μ_t the random walk component of the time series and ψ_t and ψ_t^* the unobservable components relating to the cycle. The error vector $\left(d\mu_t,\ dz_{\psi_t},\ dz_{\psi_t^*}\right)'$ is a vector of increments in the Wiener process with mean zero and variances

σ_μ^2, σ_ψ^2 and $\sigma_{\psi*}^2$[1]. The cyclical component has two parameters: λ, the cycle frequency, and α, the damping factor. If $0<\exp(\alpha)<1$ the cycle is stationary (Harvey, 1989 p. 40).

The cyclical component is in fact a mixture of sine and cosine waves and hence facilitates the modelling of the periodic component that may exist in the data series y_t. This is demonstrated easily using the matrix exponential definition together with the power series expansions for the sine and cosine functions. Ignoring the random walk component, without loss of generality, the discrete analogue of (1) is

$$\begin{bmatrix}\psi_\tau \\ \psi_\tau^*\end{bmatrix} = \alpha^\tau \begin{bmatrix}\cos\lambda\tau & \sin\lambda\tau \\ -\sin\lambda\tau & \cos\lambda\tau\end{bmatrix}\begin{bmatrix}\psi_{\tau-1} \\ \psi_{\tau-1}^*\end{bmatrix} + \begin{bmatrix}\varepsilon_{\psi\tau} \\ \varepsilon_{\psi*\tau}\end{bmatrix} \quad (2)$$

where the error-vector covariance matrix is

$$\text{var}\begin{bmatrix}\varepsilon_\tau \\ \varepsilon_\tau^*\end{bmatrix} = \left(\sigma_\psi^2/\ln(\alpha^{-2})\right)\left(1-\alpha^{2\tau}\right)I \quad (3)$$

The transition matrix in (2) contains trigonometric terms which are clearly cyclical. It should be noted that some data or theoretical considerations may prompt the addition of several cyclical components. In modelling seawater temperatures from the North Sea, however, we have found that a model with a random walk and one cyclical component is sufficient to yield homoscedastic, uncorrelated residuals.

The model is completed by the addition of a measurement equation that relates the dependent time-series variable to the state variables

$$y_t = [1 \ \ 1 \ \ 0]\begin{bmatrix}\mu \\ \psi \\ \psi_t^*\end{bmatrix} + V_t \quad (4)$$

where V_t is an error term of mean zero and covariance matrix R.

3. ESTIMATION PROCEDURE

Let the linear dynamic structural time series model be denoted using the vector (Ito) stochastic differential equation

$$dx(t) = \mathbf{F}x(t)dt + \mathbf{G}d\beta(t) \quad (5)$$

where **x** is the vector of the states, **F** and **G** are non-random coefficient matrices and $d\beta$ is the vector of increments in the Wiener processes with $E\{d\beta\, d\beta'\}=\mathbf{Q}dt$.

Linear observations are taken at time instants t, not necessarily spaced evenly, which we now denote

$$y_t = \mathbf{M}\mathbf{x}_t + \mathbf{V}_t \quad (6)$$

where y_t is the vector of observations and **M** is a non-random coefficient matrix and $\mathbf{V}_t$ is white noise with $\mathbf{V}_t \sim N(\mathbf{O},\mathbf{R}_t)$.

Given a sequence of observations $\{y_1 \ldots y_T\}$ and the coefficient matrices **F**, **G** and **M**, the Kalman filter (Kalman [1960], [1963]; Kalman and Bucy [1961]) is a recursive procedure for this continous-discrete system which computes an estimate of the conditional mean $\hat{x}_t$ and error covariance matrix $\hat{\mathbf{P}}_t$. The theory of Kalman filtering is well developed and its implementation well documented (see, for example, Jazwinski [1970]; Gelb [1974]; Maybeck [1979]; Harvey [1989]) so a brief statement of the filter will suffice here. The optimal filter for the continuous-discrete system (5) and (6) is as follows. Between observations the conditional mean and covariance matrix satisfy the differential equations

$$\frac{d\hat{\mathbf{x}}(t)}{dt} = \mathbf{F}\mathbf{x}(t)$$

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{F}\mathbf{P}(t) + \mathbf{P}(t)\mathbf{F}' + \mathbf{G}\mathbf{Q}\mathbf{G}'$$

At observation t they satisfy the difference equations

$$\hat{\mathbf{x}}_t(+) = \hat{\mathbf{x}}_t(-) + \mathbf{K}_t\left[y_t - \mathbf{M}\hat{\mathbf{x}}_t(-)\right] \quad (7)$$

$$\mathbf{P}_t(+) = \mathbf{P}_t(-) - \mathbf{K}_t\mathbf{M}\mathbf{P}_t(-) \quad (8)$$

where $\mathbf{K}_t$ is the Kalman gain matrix given by

$$\mathbf{K}_t = \mathbf{P}_t(-)\mathbf{M}'\left[\mathbf{M}\mathbf{P}_t(-)\mathbf{M}' + \mathbf{R}_t\right]^{-1} \quad (9)$$

where (+) refers to the value in period t after including the measurement of that time. Of course in the current application the coefficient matrices **F**, **M** and **G** are unknown. The central role of the filter, however, is to enable the construction of the likelihood function of the unknown model parameters via what is known as the prediction-error decomposition. This opens the way for estimation of the parameters in the coefficient matrices when these are unknown.

[1] To satisfy identification conditions σ_ψ^2 and σ_ψ^2* will be constrained to be equal (Harvey, [1989]).

The prediction errors from the filter[2] are defined as
$\varepsilon_t = y_t - M\hat{x}_t(-)$

with covariance matrix

$$\mathrm{var}(\varepsilon_t) = \Omega_t = MP_t(-)M' + R_t.$$

The likelihood function for T observations is then given by

$$\log L = -\frac{T}{2}\log 2\pi - \frac{1}{2}\sum_{K=1}^{T}\log|\Omega_t| - \frac{1}{2}\sum_{K=1}^{T}\varepsilon_t'\Omega_t^{-1}\varepsilon_t \quad (10)$$

which is to be maximised by some numerical optimisation procedure.

Existing estimation methods have focussed on the use of the matrix exponential function as a convenient representation of the solution of systems of this type. The procedure advocated by Harvey [1989] when dealing with real time applications is to discretise (5) to yield

$$x_t = \exp(F\Delta t)x_{t-1} + \eta_t \quad (11)$$

where Δt represents the time period between the observation at t and that at t-1, with η_t a white noise disturbance term of mean zero and covariance matrix

$$Q_t = \int_0^{\Delta t} \exp(F\Delta t)GQG' \exp(F\Delta t)' \, ds$$

The matrix exponential, however, often does not lend itself to numerical computation (Moler and Van Loan [1978]). As many models which may be of general interest do not have convenient closed form solutions, we propose numerical integration of the initial-value problem (5) and (6) for which reliable numerical methods are available (Cheney and Kincaid [1985]; Press et al., [1986]). The solution of the equations for the conditional mean and covariance matrix of the states is obtained at time t for the current parameter estimates by numerical integration using the Runge-Kutta-Felberg (RKF) algorithm. This is an order-5 scheme which uses 6 function evaluations to advance the solution by one time step and simultaneously compute the local truncation error in performing this step. This error is then compared with the user-specified accuracy and the step accepted or re-taken if it is too large. An error tolerance of 5×10^{-11} was used for the present paper.

2 Harvey [1989], who advocated the use of the Kalman filter in time series problems gives an in depth discussion of the derivation of the likelihood function.

4. APPLICATION TO SEAWATER TEMPERATURE DATA

To demonstrate the proposed estimation procedure we report the results of fitting a random walk and one cyclical component to monthly mean seawater surface temperature from the island of Sylt, Germany. The data set spans the period January 1958–September 1991. Estimates based on the entire regularly-spaced data set of 405 observations are reported in Table 1. These results indicate that the complete data set is consistent with a stationary cycle of 12 periods. A plot of the fitted and actual data is provided in Figure 1.

Table 1. Maximum Likelihood Estimates with Regularly-spaced Data

Parameter	Estimate	Standard Error
σ_μ^2	0.560×10^{-3}	0.050×10^{-3}
σ_ψ^2	0.070×10^{-3}	0.010×10^{-3}
σ_R^2	0.0156	0.140×10^{-3}
α	0.9994	0.030×10^{-3}
λ	0.5240	0.060×10^{-3}
$\psi(o)$	-0.6273	0.006
$\psi^*(o)$	-0.6708	0.006

$\psi(o)$ and $\psi^*(o)$ are estimated starting values for the unobservable cyclical component.

Having established a benchmark with the complete data set, random samples of 90% of the data set were taken to create irregularly-spaced data with which maximum likelihood estimates were recalculated. Individual data points were discarded at random to leave a sample of 364 monthly observations spanning the 405 months of the complete series. Each sample, therefore, consists of 364 seawater temperature records spaced between one and three months apart. In total 1620 samples were generated, the estimation results for which appear in Table 2 and Figures 2–4.

Table 2. Maximum Likelihood Estimates with Irregularly-Spaced Sample Data

Parameter	Mean of Estimates	Standard Deviation
σ_μ^2	0.539×10^{-3}	0.283×10^{-3}
σ_ψ^2	0.065×10^{-3}	0.025×10^{-3}
σ_R^2	0.0162	0.0012
α	0.9994	0.092×10^{-3}
λ	0.5241	0.112×10^{-3}
$\psi(o)$	-0.6239	0.009
$\psi^*(o)$	-0.6605	0.015

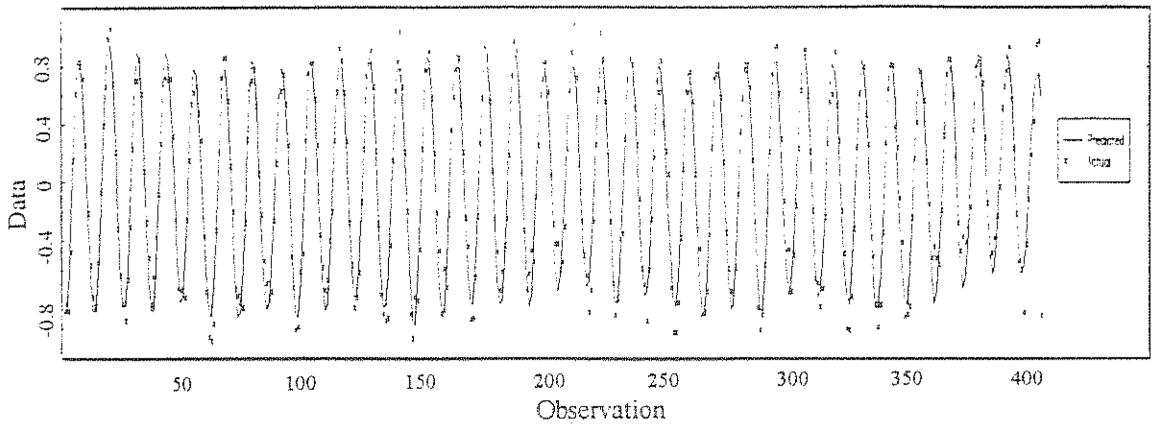

Figure 1: Fitted and Actual Data (Complete Data Set)

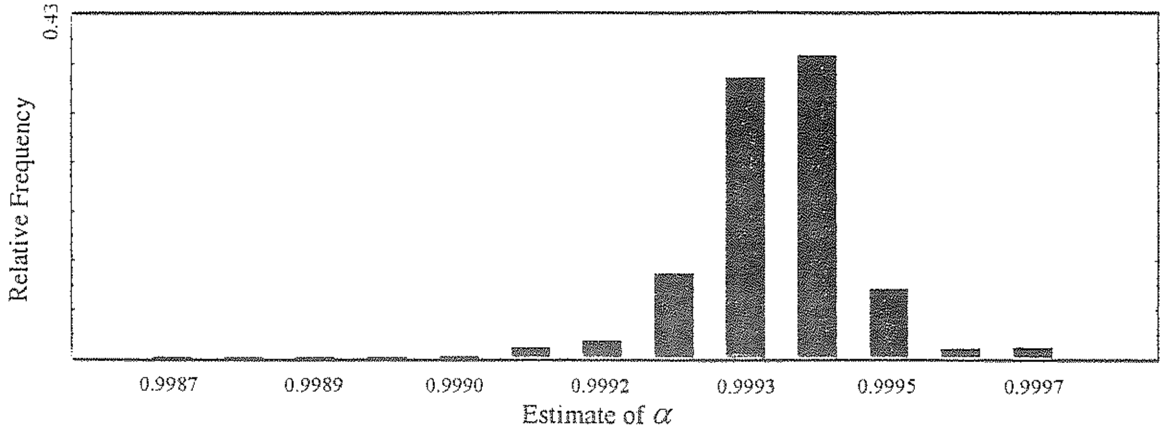

Figure 2: Histogram of Sample Estimates of α

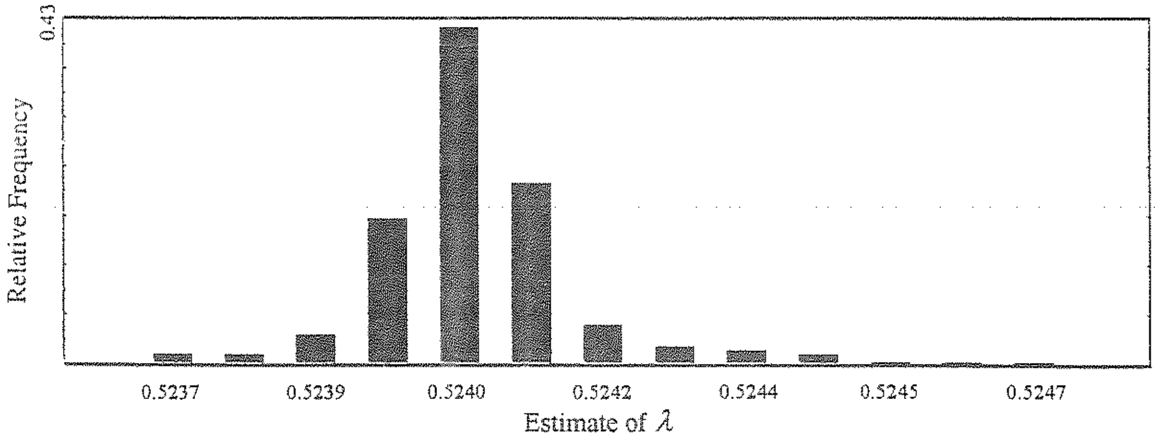

Figure 3: Histogram of Sample Estimates of λ

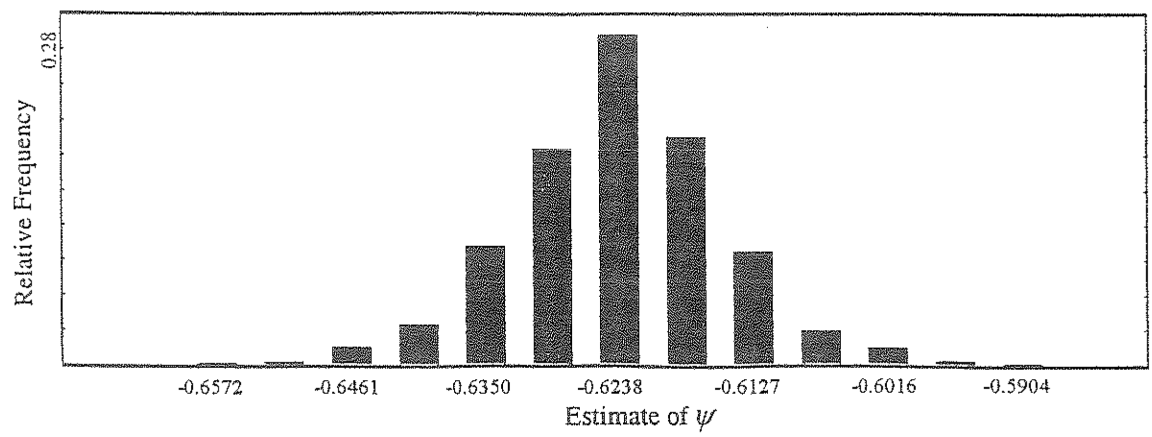

Figure 4: Histogram of Sample Estimates of $\psi(0)$

The results recorded in Tables 1 and 2 illustrate the high precision possible for parameter estimates when time-series data are reported irregularly. When 10% of the data are discarded randomly, subsequent estimates of α, λ and $\psi(o)$ are centred tightly around those obtained using the complete data set. In order to ascertain how well the method performs when smaller samples of (more) irregular data are used Table 3 contains means and standard deviations of parameter estimates from 100 simulations for 90, 80, ..., 50 percent of the full data set. These results indicate that precision is maintained quite well in the face of increased irregularity. For data of the type described above, therefore, the method proposed for isolating cyclical patterns in irregular time-series data is very promising.

5. CONCLUSION

It is yet to be established precisely how robust the method is when only small data sets are available, when the cyclical signal is weak and when data-recording intervals are very irregular. Work in progress by the present authors, however, is very encouraging in this regard. Early results indicate that data recorded at highly irregular intervals do not present difficulties for precision of the estimators or for tracking the data with forecasts. Accordingly, we believe that the proposed method is worthy of application in a broader context than that of the present paper.

Table 3. Mean and Standard Deviation of Estimates from 100 Simulations

Parameter	Mean (Standard Deviation)				
	90% Sample	80% Sample	70% Sample	60% Sample	50% Sample
σ_μ^2	0.598×10^{-3} (0.403×10^{-3})	0.573×10^{-3} (0.406×10^{-3})	0.628×10^{-3} (0.612×10^{-3})	0.645×10^{-3} (0.874×10^{-3})	0.640×10^{-3} (0.0011)
σ_ψ^2	0.067×10^{-3} (0.026×10^{-3})	0.065×10^{-3} (0.042×10^{-3})	0.068×10^{-3} (0.054×10^{-3})	0.083×10^{-3} (0.094×10^{-3})	0.077×10^{-3} (0.087×10^{-3})
σ_R^2	0.0159 (0.0015)	0.0167 (0.0019)	0.0174 (0.0031)	0.0188 (0.0045)	0.0208 (0.0057)
α	0.9994 (0.119×10^{-3})	0.9994 (0.144×10^{-3})	0.9994 (0.190×10^{-3})	0.9993 (0.278×10^{-3})	0.9993 (0.0253×10^{-3})
λ	0.5241 (0.125×10^{-3})	0.05240 (0.189×10^{-3})	0.5240 (0.233×10^{-3})	0.5241 (0.298×10^{-3})	0.5241 (0.277×10^{-3})
$\psi(o)$	-0.6249 (0.0086)	-0.6268 (0.0118)	-0.6279 (0.0133)	-0.6274 (0.0151)	-0.6347 (0.0254)
$\psi^*(o)$	-0.6614 (0.0174)	-0.6605 (0.0193)	-0.6600 (0.0241)	-0.6673 (0.0276)	-0.6623 (0.0344)

ACKNOWLEDGEMENTS

This research was conducted while A. S. Hurn was visiting the CSIRO Division of Fisheries, Hobart, whose financial support is acknowledged gratefully. The authors also wish to thank, Dr K. A. Lindsay of the Department of Mathematics, The University of Glasgow, from whose original Fortran code the numerical integration procedure was adapted and translated for use with the GAUSS programming language. Thanks are also extended to Wetterdienst Schleswig and Christian-Albrechts University of Kiel for making North Sea water temperature data available and to Adam Davidson for excellent research assistance.

REFERENCES

Cheney, W. and Kincaid, D., Numerical Mathematics and Computing, Wadsworth, Belmont California, 1985.

Gelb, A., Applied Optimal Estimation, The M.I.T. Press, 374 pp., Cambridge Massachusetts, 1974.

Harvey, A.C., Forecasting, Structural Time Series Models and the Kalman Filter, *Cambridge University Press*, 554 pp., Cambridge, 1989.

Harvey, A.C., Ruiz, E. and Sentana, E., Unobserved component time series models with ARCH disturbances", *Journal of Econometrics*, 52(1), 129–158, 1992.

Jazwinski, A.H., Stochastic Processes and Filtering Theory, *Academic Press*, 376 pp., New York, 1970.

Kalman, R.E., A new approach to linear filtering and prediction problems, *J. Basic Eng*, 82, 35–45, 1960.

Kalman, R.E., and Bucy, R.S., New results in linear filtering and prediction theory, *J. Basic Eng.*, 83, 95–108, 1961.

McDonald, A.D. and Hurn, A.S., Unobservable Cyclical Components in Term Premia of Fixed-Term Financial Instruments. *Mathematics and Computers in Simulation*, 35, 1994.

Moler, C.B. and Van Loan, C.F., Nineteen dubious ways to compute the exponential of a matrix, *SIAM Review*, 20, 801–836, 1978.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T., Numerical Recipes, *Cambridge University Press*, New York, 1986.