The Application of the Kalman Filter to Nonstationary Time Series through Time Deformation

Zhu Wang * Henry L. Gray † Wayne A. Woodward ‡

March 17, 2008

Abstract

An increasingly valuable tool for modeling nonstationary time series data is time deformation. However, since the time transformation is applied to the time axis, equally spaced data become unequally spaced data after time transformation. Interpolation is therefore often used to obtain regularly sampled data, which can be modeled by classical ARMA modeling techniques. However interpolation may be undesirable since it can introduce spurious frequencies. In this paper, the continuous time autoregressive model is used in order to eliminate the need

*Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, LE-400, Seattle, WA 98109

†Statistical Science Department, Southern Methodist University, Dallas, Texas 75275-0332. E-mail: hgray@mail.smu.edu

‡Statistical Science Department, Southern Methodist University, Dallas, Texas 75275-0332. E-mail: waynew@mail.smu.edu

to interpolate. To estimate the parameters, a reparameterization and model suggested by Belcher, Hampton, and Tunnicliffe-Wilson (1994)is employed. The resulting modeling improvements include, more accurate estimation of the spectrum, better forecasts, and the separation of the data into its smoothed time-varying components. The technique is applied to simulated and real data for illustration.

1 Introduction

The analysis of time series is often difficult when data are nonstationary. For instance, while the definition of the spectrum for stationary time series is well established, there appears to be no analogous way of defining a time-dependent spectrum uniquely. One approach, which is particularly useful when the frequencies are changing monotonically is to transform time in such a way that the process is stationary on the transformed index set. This then leads, in a natural way, to an appropriate definition of the spectrum at any time, t, i.e., the "instantaneous spectrum". Gray and Zhang (1988) introduced continuous multiplicative-stationary (M-stationary) processes for the case in which the time deformation is logarithmic, i.e. $u = \ln t$. Jiang, Gray, and Woodward (2006) extended the M-stationary processes by introducing $G(\lambda)$-stationary processes. A general class of the $G(\lambda)$ processes, referred to as the $G(p, \lambda)$ process, is then introduced through a Box-Cox transformation of time. The estimation approach for the discrete M-stationary (Gray, Vijverberg, and Woodward, 2005) and discrete $G(\lambda)$-stationary (Jiang et al., 2006) time series models can be summarized as follows:

(1) Beginning with equally spaced nonstationary data, a time deformation

is applied to obtain unevenly spaced data from a continuous ARMA process.

(2) Interpolation is then employed to obtain evenly spaced discrete ARMA$(p, p-1)$ data so that the conventional AR or ARMA model can be used.

(3) If forcasting is needed, 're-interpolation" is then required in order to return to the original time scale.

While it may suffice for some purposes, interpolation often introduces distortion and noise. In fact, properly interpolating time series data itself is a challenging research topic.

Recognizing both the difficulties and the inherent disadvantage of the above discrete interpolation approach, the purpose of this paper is to utilize the continuous time autoregressive modeling as an alternative. Parameter estimation is based on the maximied likelihood function, which is decomposed by the Kalman filter. Taking advantage of this approach, familiar applications to forecasting, interpolation, smoothing and latent components estimation can be applied to nonstationary time series data. The idea of fitting continuous autoregressive models using time deformation is not new. For instance, Gray and Zhang (1988) advocated taking the approach of Jones (1981), and Stock (1988) used a similar approach to analyze the postwar U.S. GNP. However, the estimation approach proposed by Jones (1981) for unequally-spaced data implicitly limits the applications to lower order models. In fact, both of the two examples illustrated in Stock (1988) are first order autoregressive. However, Belcher, et al. (1994) introduced a

reparametriction of the Jones method which provided a much more stable likelihood algorithm. In this paper we will make extensive use of the model and the numerical methodology suggested in Belcher, et al. (1994). The remainder of this paper is organized as follows. In Section 2, the basics of M-stationary and $G(\lambda)$-stationary processes are reviewed. Section 3 describes continuous time autoregressive models and the Kalman filter. Section 4 presents three applications, and Section 5 summarizes the conclusions and discusses future research directions and possible extensions.

2 M-stationary and $G(\lambda)$-stationary processes

A continuous time series process $X(t), t \in (0, \infty)$, is M-stationary if $X(t)$ has a constant mean, finite, constant variance and $E[(X(t) - \mu)(X(t\tau) - \mu)] = R_X(\tau)$. An important M-stationary process is the Euler process. $X(t)$ is a pth-order Euler process, denoted Euler(p), if $X(t)$ satisfies

$$t^p X^{(p)}(t) + \psi_1 t^{p-1} X^{(p-1)}(t) + \ldots + \psi_p (X(t) - \mu) = a(t) \qquad (1)$$

where $X^{(i)}(t)$ is the ith derivative of $X(t)$, $E[X(t)] = u$ and $\psi_1, ..., \psi_p$ are constants, and $a(t)$ is M-white noise, meaning that $a(t) = \epsilon(\ln(t))$ where $\epsilon(t)$ is the formal derivative of a Brownian process. See Gray and Zhang (1988). Without loss of generality it will be assumed that $E[X(t)] = 0$. If $X(t)$ is a zero mean Euler(p) process defined by (1), and $Y(u) = X(e^u)$, i.e., $Y(\ln(t)) = X(t)$, then

$$Y^{(p)}(u) + \alpha_1 Y^{(p-1)}(u) + \ldots + \alpha_{p-1} Y^{(1)}(u) + \alpha_p Y(u) = \epsilon(u) \qquad (2)$$

where $\alpha_1, ..., \alpha_p$ are constants determined by $\psi_1, ..., \psi_p$ and $\epsilon(u) = a(e^u)$. $Y(u)$ is referred to as the dual stationary process of $X(t)$. While the natural logarithmic transformation is considered above, it may be shown that if a process is M-stationary it can be viewed as stationary on any log scale. For $\delta > 0$ and $t > 0$, the process defined by $Z(t) = X(t - \delta)$ is referred to as a shifted M-stationary process. For more on M-stationary processes, see Gray and Zhang (1988) and Gray et al. (2005).

A more general process is the so called $G(\lambda)$-stationary processes. $X(t)$ is said to be a $G(\lambda)$-stationary process on $t \in (0, \infty)$ if $X(t)$ has a constant mean, finite constant variance, and for any $t^\lambda + \tau\lambda \in (0, \infty)$, then it follows that $E[X(t) - \mu][X(t^\lambda + \tau\lambda)^{\frac{1}{\lambda}} - \mu] = B_X(\tau, \lambda)$, for $\lambda \in (-\infty, \infty)$. The class of $G(\lambda)$ stationary processes contain the usual stationary processes (when $\lambda = 1$ and $t > 0$). In the limiting case as $\lambda \to 0$, the $G(\lambda)$-stationary process becomes the M-stationary process. The time deformation required to transform this class of nonstationary processes to usual stationarity is the Box-Cox transformation, $u = \frac{t^\lambda - 1}{\lambda}$.

The fundamental theory for discrete, evenly or unevenly sampled $G(\lambda)$-stationary processes has been developed in Jiang, et al. (2006). To be specific, consider the case $\lambda = 0$, i.e. the M-stationary case. Then $\{X(t)\}$ is formally defined through the model given by (1). If $u = \log_h t$ and $Y(u) = X(h^u) = X(t)$, $h > 1, t > 0$, it is shown that $Y(u)$ is a continuous AR(p), so that for $k = 0, \pm 1, \ldots$, then $Y_k = Y(k) = X(h^k)$ is a discrete ARMA($p, p-1$). Thus, if the data are sampled at $t_k = h^k$, the resulting data can be viewed as equally-spaced observations from a discrete ARMA($p, p - 1$). See Choi, Woodward and Gray (2006). However, in most cases the data available

are equally spaced in the original time scale so that $X(h^k)$ is not available (unless the data are heavily over sampled), but instead one obtains $Y(u)$ at $u = \frac{\ln k}{\ln h} = \log_h k$, $k = 1, ..., n$, i.e. $Y(u)$ is an unequally spaced sample from a continuous AR(p). Thus Gray, et al. (2005) interpolated the data to obtain an evenly spaced dual process Y_m, $m = 1, ..., n$ to obtain an ARMA($p, p-1$) which could then be analyzed by standard methodology. The same approach was taken by Jiang, et al. (2005) with $u = \frac{t^\lambda - 1}{\lambda}$. Then $Y(u) = X((\lambda u + 1)^{1/\lambda)} = X(t)$. It is shown there that the limit as $\lambda \to 0$ results in the M-stationary case while the case $\lambda = 1$ is an AR(p). Thus in the general case the data are required at $(\lambda k\delta + 1)^{1/\lambda}$, where δ is the sampling increment and $k = 1, 2, ..., n$. Consequently if the data are equally spaced then interpolation is required in this case as well. In the next few sections we make use of the Kalman filter and the methods of Jones (1981) and Belcher, et al. (1994) to eliminate the need for interpolating the data and thus eliminate errors introduced by interpolation. In closing this section we should mention that λ and the realization origin (see Gray, et al. (2005) and Jiang, et al. (2006)) must be estimated. However, the software for estimating these parameters as well as the model in general is available at http://faculty.smu.edu/hgray/research.htm.

3 Continuous time autoregressive models

Consider now estimating a continuous time autoregressive model (2) from the observed discrete dual process Y_{k_i} for $i = 1, ..., n$ Note that the Y_{k_i} are unequally spaced observations from a continuous AR(p) process. For such

data, Jones (1981) applied the Kalman filter to decompose the maximum likelihood function and Belcher et al. (1994) further developed this approach. We will use operator notation to express model (2) as $\alpha(D)Y(t) = \epsilon(t)$ where

$$\alpha(D) = D^p + \alpha_1 D^{p-1} + \ldots + \alpha_{p-1} D + \alpha_p, \tag{3}$$

where D is the derivative operator. The corresponding characteristic equation is then given by

$$\alpha(s) = s^p + \alpha_1 s^{p-1} + \ldots + \alpha_{p-1} s + \alpha_s = 0. \tag{4}$$

The characteristic polynomial $\alpha(s)$ can be factored as

$$\alpha(s) = \prod_{i=1}^{p} (s - r_i). \tag{5}$$

A necessary and sufficient condition for stationarity of model (2) is that all the zeros have negative real parts. This can be achieved by letting

$$s^2 + e^{\theta_{2k-1}} s + e^{\theta_{2k}} = (s - r_{2k-1})(s - r_{2k}) \tag{6}$$

where $k = 1, \ldots, \frac{p}{2}$ if p is even and $k = 1, \ldots, \frac{p-1}{2}$ if p is odd, with $r_p = -e^{\theta_p}$. Without any constraint on θ_i, the stationarity is then automatically guaranteed by this reparameterization, and the parameter space is R^p.

The method of Jones, however, is well known to have convergence problems for higher order models. To solve this problem, Belcher et al. (1994) proceed as follows. Let

$$z = \frac{1 + s/\delta}{1 - s/\delta}, \tag{7}$$

where $\delta > 0$ is referred to as a scaling factor. Now let f denote frequency in the frequency domain of $Y(u)$ and transform to a new frequency variable,

w, by letting

$$f = \delta \tan(w/2). \tag{8}$$

It is easily shown that if $z = \exp(iw)$ then $s = if$. That is, the boundary of the unit circle maps into the imaginary axis, if, when equations (7) and (8) hold.

Now for real constants $\phi_1, \ldots, \phi_p$ define

$$\phi(z^{-1}) = 1 + \phi_1 z^{-1} + \ldots + \phi_p z^{-p}. \tag{9}$$

Using (7), then (9) gives

$$\phi(z^{-1}) = \frac{\beta(s)}{(1 + s/\delta)^p},$$

where

$$\begin{aligned} \beta(s) &= \beta_0 s^p + \beta_1 s^{p-1} + \ldots + \beta_{p-1} x + \beta_p & (10) \\ &= \sum_{i=0}^{p} \phi_i (1 - s/\delta)^i (1 + s/\delta)^{p-i}, & (11) \end{aligned}$$

with $\phi_0 = 1$. Note that the β_i's are linear combination of the ϕ_i's. If we set $\alpha(s) = \beta(s)$ with $\alpha_i = \beta_i/\beta_0$, then since

$$z^p \phi(z^{-1}) = z^p + \phi_1 z^{p-1} + \ldots + \phi^p \tag{12}$$

it follows from (7) and (8) that zeros of $\alpha(s)$ lie in the left hand plane if and only if the zeros of $\phi(r)$ lie inside the unit circle.

We therefore have the very nice result which we quote from Belcher, et al. (1994), "Thus the space of parameters ϕ_i corresponding to stationary model parameters, α_i is precisely the same as that associated with the stationary discrete time autoregressive operator $\phi(B)$, where the backward shift operator B is identified with z^{-1}."

Letting $S(f)$ denote the spectrum of a stationary process $Y(u)$, then it can be seen that this transforms, using (8), to the spectrum $F(w)$ given by

$$F(w) = S(f)[1 + (f/\delta)^2] \left(\frac{\delta}{2}\right). \tag{13}$$

Using (13), it can be seen that the spectral density, $F_m(w)$ associated with model (14) is

$$F_m(w) = \frac{\delta}{2}\sigma^2 \left| \frac{(1 + if/\delta)^{p-1}}{\alpha(if)} \right|^2 (1 + f/\delta)^2,$$

With this in mind consider the modification of the model in (2) to

$$\alpha(D)Y_t = (1 + D/\delta)\epsilon(t). \tag{14}$$

Since $|1 + iX| = 1 + X^2$, it follows that

$$\begin{aligned} F_m(w) &= \frac{\delta}{2} \frac{(1 + (f/\delta)^2)^p}{|\alpha(if)|^2} \sigma^2 \\ &= \frac{\delta}{2} \left| \frac{\beta_0}{\phi\{\exp(iw)\}} \right|^2 \sigma^2, \end{aligned}$$

where $\beta_0 = \phi(-1)\delta^2$. Thus the spectral density of $F_m(w)$ of the process $Y(t)$ given in (14) has the form of the spectral density of a discrete AR(p). For further disussion of this and the advantage of replacing the model in (2) with the model in (14), and there are many, see Belcher et al. (1994).

Thus to improve numerical stability as well as eliminating the need for interpolation we will henceforth replace model (2) with the model given by (14). We should mention that this is equivalent to replacing the $G(\lambda, p)$ model by the corresponding $G(\lambda; p, p-1)$. See Choi, et al. (2006). The model in (14) will then be fitted via the maximum likelihood function calculated through the Kalman filter. The details of this can be found in

Jones (1981), Belcher et al. (1994) and Wang (2004). The software for finding and fitting a $G(\lambda, p)$ model can be downloaded from the website http://faculty.smu.edu/hgray/research.htm.

4 Applications

There are several issues needed to be addressed before we apply the approach of Belcher et al. (1994) to M-stationary and $G(\lambda)$-stationary processes. First, corresponding to the change from model (2) to (14), $X(t)$ is modeled using a special case of a continuous Euler(p, q) process (Choi, et al. 2006) or a $G(\lambda; p, q)$ processes (Jiang, et al. 2006). Therefore, henceforth when we speak of continuous autoregressive, Euler and $G(\lambda)$ models they should be understood as these restricted models related to (14). Second, to avoid more complication of the maximum likelihood function, time deformation is separated from the estimation of continuous time autoregressive models. We take the approach of Gray et al. (2005) and Jiang, et al. (2006) to estimate time deformation for M-stationary and $G(\lambda)$-stationary processes, respectively. Thirdly, following Belcher et al. (1994), the choice of the scaling factor κ will be a trade-off between capturing the data structure and overfitting. As illustrated in their example, the scale parameter may be chosen approximately as the reciprocal of the mean time between observations. For the current applications, it is worth noting that using the logarithmic or Box-Cox transformation discussed in Section 2 produced dual data for which the mean time between observations is equal to one. In our implementation we have found the scaling favor of 1.5 to be a good

choice. Experience gained in applications suggests this choice can caputre the spectrum well and not overfit. Lastly, the model selection is based on AIC and the t-statistic proposed in Belcher et al. (1994). The remainder of this section illustrates the procedure.

Example 1. One of the advantages of the continuous time model fitting is that there is no extra distortion and noised introduced, unlike in the interpolation-based approach. To demonstrate this point, we consider the example given in Gray et al. (2005). The data are generated from the shifted discrete Euler(2) process with sample size 400,

$$X(h^{j+k}) - 1.732X(h^{j+k-1}) + 0.98X(h^{j+k-2}) = a(h^{j+k}) \qquad (15)$$

with the $h^j = 10, h = 1.005$ and $\sigma^2 = 1$. A realization is displayed in Figure 1. All 400 data values are used to fit a continuous Euler model. For purpose of illustration, the original $h^j = 10, h = 1.005$ are used, for both the discrete Euler model and the continuous Euler model. Continuous Euler models up to order 18 were fit to the data. The resulting AIC and the t-statistics are shown in Table 1. Based on Table 1, a continuous Euler(4) is chosen. The estimated M-spectrum based on this continuous model is plotted along with the M-spectrum based on the discrete Euler model (estimated using interpolated values) for comparison. It can be seen from Figure 2, that the left panel has the second highest peak at a low frequency while the right panel has only one peak, which is consistent with the model. While it seems that the spurious low frequency peak may be exaggerated by the AIC selected high order (19) model, it is found that any fitted model with order greater than 4 has a spurious low frequency peak for a discrete

Figure 1: Realization of Euler(2)

Figure 2: Spectral estimates for Euler(2)

Euler model. However, this is not the case for a continuous Euler model even for a higher order model. This effect is caused not by chance, in fact, Press et al. (1992, p.576) described this phenomenon

> .... However, the experience of practitioners of such interpolation techniques *is not reassuring*. Generally speaking, such techniques perform poorly. Long gaps in the data, for example, often produce a spurious bulge at low frequencies (wavelengths comparable to gaps)

This is exactly the situation here. Since the time deformation is monotonic, there exist long gaps at the beginning part of data. The continuous Euler model, however, does not suffer from this distortion.

One of the objectives of the Kalman filter is to filter out the random noise from the signal. For a process with contaminating noise, it is possible to fit a continuous time model with a measurement error term; then a Kalman smoother may be applied to extract the signal from the noise. To demonstrate, random normal noise with mean zero and variance equal to

Figure 3: Forecasting comparison for Euler(2)

Order	t-statistic	AIC
1	−181.2441582	−32847.44
2	140.4169008	−52562.35
3	−9.8492519	−52657.36
4	5.1125129	−52681.50
5	−1.4639074	−52681.64
6	0.3369399	−52679.75
7	1.2906832	−52679.42
8	−1.4747653	−52679.59
9	0.7339677	−52678.13
10	0.5591542	−52676.45
11	−1.8504884	−52677.87
12	2.0998675	−52680.28
13	−1.5554970	−52680.70
14	1.5266172	−52681.03
15	−1.8150720	−52682.32
16	−0.5244112	−52680.60
17	−1.0917761	−52679.79
18	−0.4196429	−52677.97

Table 1: Order selection statistics for Euler(2)

lag	AR	Discrete Euler	Continuous Euler
10	2581.231	41.788	39.03
20	1684.479	335.917	92.311
30	2454.773	898.771	283.324
40	1665.196	1512.341	1801.851
50	1390.837	1119.811	1648.748
60	1270.888	658.738	894.293

Table 2: MSE forecast comparison for Euler(2)

Figure 4: Noise contaminated data

the sample variance of the signal was added to the data shown in Figure 1. The composite signal plus noise is shown in Figure 4. For illustration purposes, the original $h^j = 10, h = 1.005$ are used and the dual process is obtained on the transformed time. A continuous Euler model is then fitted with the measurement error term, and the model is selected by AIC. The AIC valuesContinuous $G(\lambda)$ do not change dramatically, and a 4th order model is chosen. After the parameter estimation, the smoothed series is estimated and shown in Figure 5. It can be seen that the smoothed data looks similar to the original signal in Figure 1.

Figure 5: Signal extraction

Figure 6: Realization of Euler(4)

Example 2. Cohlmia et al. (2004) considered a realization generated from an Euler(4) model

$$(1 - 1.97B + 0.98B^2)(1 - 1.7B + 0.99B^2)X(h^k) = a(h^k), \qquad (16)$$

where B is the autoregressive operator, and the offset $h^j = 15, h = 1.005$ and $\sigma^2 = 1$. The data are shown in Figure 6. The fitted discrete model is an Euler(11) with $\hat{h} = 1.0082, \hat{h}^j = 15$. The two dual system frequencies are $f = 0.026$ and $f = 0.142$. Here We use the same realization to fit a continuous Euler model. For model selection using this time deformation, first a continuous Euler(16) is fit, and the model order selection statistics are shown in Table **??**. Both of these statistics choose an order 15, thus an order 15 continuous Euler model is finally fit, and the roots r_i of the characteristic equation based on the fitted coefficients are shown in Table 3.

The smaller the value of $|Re(r_i)|$, the higher the power associated with the corresponding frequency. As can be seen in Table 3, the power is contained mostly in the M-frequencies 17.31 and 2.89 which are associated with complex roots -0.011 ± 0.888i and -0.030 ± 0.148i, respectively. From (16), the M-frequencies associated with the factor $1 - 1.97B + 0.98B^2$ and $1 - 1.7B + 0.99B^2$ are 17.44 and 3.19, respectively. Using a discrete Euler(11) model, as described in Cohlmia et al. (2004), the fitted M-frequencies are 17.42 and 3.22 respectively. The estimated continuous M-spectrum is shown in Figure 7, along with the M-spectrum from the discrete Euler(11) model.

	Roots	M-frequency	Dual frequency
1	$-0.011 \pm 0.888i$	17.31	0.141
2	$-0.030 \pm 0.148i$	2.89	0.024
3	$-0.095 \pm 0.224i$	4.36	0.036
4	$-0.121 \pm 1.057i$	20.62	0.168
5	$-0.135 \pm 0.522i$	10.18	0.083
6	$-0.217 \pm 1.783i$	34.77	0.284
7	$-0.461 \pm 9.750i$	190.18	1.552
8	-6.082	0.00	0.000

Table 3: Factors of fitted model for Example 4.4

Figure 7: Spectral estimates for Example 4.4

The highest two peaks are approximately located at the dual frequencies 0.141 and 0.024, from Table 3.

To compare forecasting performance, an AR(18) model chosen by AIC was also fitted as described in Cohlmia et al. (2004), and the forecasts for the last 30 steps were found for the three different models. As can be seen in Figure 8, the continuous model forecasts are more accurate.

It can be seen from Table 4, that the continuous Euler(15) model overwhelming outperforms the discrete Euler(11) model. While the discrete Euler(11) does consider the time-varying frequency of the time series, the modeling procedure first interpolates the original data to obtain the evenly spaced dual data, then models the dual data and obtain forecasts using AR

lag	AR	Discrete Euler	Continuous Euler	Continuous Euler (Wrong)
10	15975.83	84.74	77.17	0.19
20	21960.41	338.30	1928.38	0.18
30	13835.11	2658.27	1500.88	0.24
40	17827.06	62110.1	18847.41	0.65
50	18642.87	26735.55	27028.65	0.87
60	13786.61	25600.46	26802.83	0.91

Table 4: MSE forecasting comparison for Example 4.4

Figure 8: Forecasting comparison for Example 4.4

model techniques, and then finally involves reinterpolation on the original time scale. It is seen that these two steps of interpolation cause the forecasts to be less accurate, even though the discrete Euler(11) clearly does a better job than a direct application of an AR model. To check the assumptions of the fitted continuous model, the standard residual diagnostics are plotted in Figure 9. As can be seen, the diagnostics do not find anything suspicious. In addition, the standardized residuals passed the Box-Ljung white noise test for both 24 and 48 lags.

This data set is considered again in a filter setting. The components associated with the two dominant M-frequencies 17.31 and 2.89, or the dual

Figure 9: Model diagnostics for Example 4.4

Figure 10: Filtered data for Example 4.4

frequencies 0.141 and 0.024, can be estimated from the dual process. Using the Kalman smoothing algorithm, the estimated two dominant frequency components after mean corrected are shown in Figure 10. In Figure 10 we also show the data after M-filtering out the M-frequency 17.31, by subtracting the estimated component corresponding to M-frequency 17.31, from the original data. The M-filtered data look similar to that given in Cohlmia et al. (2004), who described a Butterworth filter approach to filter the interpolated dual data, and then reinterpolated back to the original time scale. Again, the method discussed in this example avoided any interpolation for the purpose of filtering.

Example 3. Figure 11 shows brown bat echolocation data. The data set has been analyzed in the framework of time frequency analysis, e.g., Forbes and Schaik (2000). In the context of time deformation, it was discussed in Gray et al. (2005) and Cohlmia et al. (2004). The realization analyzed here is of length 381. The fitted discrete Euler(11) model has $\hat{h} = 1.00278, \hat{h}^j = 202$ with strong dual system frequencies at .048, 0.294, and 0.435 or M-frequencies (a scaled frequency defined as $f^* = f/\ln h$, where f is the dual frequency) at 0, 53.2, 105.8, and 156.7, respectively. Using time deformation, a CAR(18) model is fitted. The model order selection statistics are shown in Table 5. Both t-statistic and AIC statistics choose an order 18.

Thus, a continuous model of order 18 was fitted to the data, and the roots r_i of characteristic equation from fitted coefficients are shown in Table 7.

Figure 11: Brown bat echolocation data

Order	t-statistic	AIC
1	20.60	−422.35
2	234.06	−55205.13
3	−10.73	−55318.25
4	146.06	−76648.91
5	64.37	−80790.60
6	56.00	−83924.57
7	−45.30	−85974.37
8	−98.09	−95594.57
9	−27.84	−96367.40
10	−8.35	−96435.15
11	−17.01	−96722.43
12	−21.54	−97184.53
13	−17.06	−97473.66
14	20.64	−97897.57
15	−2.11	−97900.03
16	13.49	−98080.08
Continuous $G(\lambda)$ 17	−5.12	−98104.34
18	3.16	−98112.34

Table 5: Order selection statistics for brown bat data

Figure 12: Spectral estimates of brown bat data

lag	AR	Discrete Euler	Continuous Euler
10	2.77e-05	1.12e-06	4.77e-06
20	3.24e-05	8.51e-06	2.33e-05
30	6.71e-05	4.24e-05	6.13e-05
40	2.18e-04	6.21e-05	4.22e-05
50	2.38e-04	1.09e-04	1.68e-04
60	1.94e-04	1.29e-04	2.91e-04

Table 6: MSE forecast comparison for the brown bat data

Figure 13: Forecasting comparison for the brown bat data

As can be seen, the power is concentrated mostly in the dual frequencies 0, 0.294, 0.146, and 0.435, or M-frequencies 0, 105.8, 52.5 and 156.7, respectively, which are associated with roots 0, -0.006 ± 1.846i and -0.01 ± 0.916i, respectively. In this case, the spectra associated with the discrete and continuous Euler models are similar, as can be seen in Figure 12.

	Roots	M-frequency	Dual Frequency
1	-0.003	0.00	0.000
2	-0.006±1.846i	105.81	0.294
3	-0.01±0.916i	52.53	0.146
4	-0.016±0.95i	54.48	0.151
5	-0.018±2.733i	156.69	0.435
6	-0.055±3.623i	207.71	0.577
7	-0.11	0.00	0.000
8	-0.26±1.789i	102.54	0.285
9	-0.271±12.949i	742.37	2.061
10	-0.399±0.456i	26.15	0.073

Table 7: Factors of fitted model for brown bat data

It is of interest to decompose the signal into its "basic components" which are a by product of the Kalman filter. Note that after the model parameters are estimated, the components associated with the dominant M-frequencies can be estimated. From Table 7, it can be seen that there

Figure 14: Components of brown bat data

Figure 15: Hunting bat echolocation data

are four dominant frequencies. The first dominant dual low frequency is 0. The dual frequency of 0.294 and 0.285 are close, 0.146 and 0.151 are close together, thus those close frequency components will be combined together. The remaining dominant dual frequency is 0.435. The results obtained here are similar to those given in Cohlmia et al. (2004). Using the Kalman smoother, the four dominant frequency components may be estimated and are shown, after mean correction, in Figure 14.

Example 4. This data set contains 280 observations from a Nyctalus noctula hunting bat echolocation at 4×10^{-5} second intervals, which was analyzed in Gray et al. (2005) using a discrete Euler model. The data are shown in Figure 15. The fitted discrete Euler(12) model has $\hat{h} = 1.00326, \hat{h}^j = 188$. Using this time deformation, a continuous Euler(15) model is fitted. The model order selection statistics are shown in Table **??**. Both t-statistic and AIC statistics choose an order 14.

Thus, a continuous Euler(14) model is fit to the data and the roots r_i of characteristic equation from the estimated coefficients are shown in Table 8.

Figure 16: Spectral estimates of hunting bat data

Figure 17: Forecasting comparison for hunting bat data

As can be seen in Figure 16, the power is contained mostly in the dual frequencies 0.121, 0.242, 0 and 0.366.

	Roots	M-frequency	Dual Frequency
1	-0.003±0.761i	37.21	0.121
2	-0.017±1.523i	74.46	0.242
3	-0.056	0.00	0.000
4	-0.089±2.302i	112.59	0.366
5	-0.107±0.277i	13.53	0.044
6	-0.164±0.89i	43.51	0.142
7	-0.177±5.888i	287.93	0.937
8	-32.16	0.00	0.000

Table 8: Factors of fitted model for hunting bat data

To compare forecasting performance, the forecasts for the last 60 data values are obtained for the three different models and the results are shown in Figure 17. To compare forecast performance, different forecast origins are examined and the results are listed in Table 9As can be seen, the continuous Euler model forcasts marginally outperform the discrete Euler model. To check the assumptions of the fitted continuous model, the standard residual diagnostics are plotted in Figure 18.

lag	AR	Discrete Euler	Continuous Euler	Continuous Euler (Wrong)
10	0.0765	0.0209	0.0183	0.0092
20	0.1478	0.0100	0.0102	0.0006
30	0.1939	0.0080	0.0100	0.0057
40	0.2741	0.0236	0.0205	0.0065
50	0.2021	0.0305	0.0233	0.0073
60	0.2027	0.0252	0.0189	0.0067

Table 9: MSE forecast comparison for hunting bat data

Figure 18: Model diagnostics for hunting bat data

Figure 19: Components of hunting bat data

To estimate unobserved dominant components, the components associated with the dominant M-frequencies can be estimated. Corresponding to the frequencies in Table 8, the four dominant components are estimated using the Kalman filter smoothing algorithm, the results after mean correction are shown in Figure 19.

Example 5. Cohlmia et al. (2004) investigated a simulated quadratic chirp in the class of $G(\lambda)$-stationary process given by

$$X(t) = \cos[2\pi(\frac{t}{250} + .25)^3] + 2\cos[7\pi(\frac{t}{250} + .25)^3] + .1\epsilon(t) \qquad (17)$$

where $\epsilon(t)$ are standard normal variates.

A realization from (17) of length 400 is shown in Figure 20 and the two deterministic components in (17) are plotted in Figure 21. These components have time varying frequencies. A discrete $G(\lambda)$ model of order 20 is fitted to the data with $\hat{\lambda} = 3, \hat{h} = 81127, \hat{\Lambda} = 60$. In order to conduct continuous $G(3)$ modeling, time deformation using the above parameters is performed to obtain unevenly spaced stationary dual data. These dual data are to be fitted using continuous AR models. An 11th order model is fitted to the data and is chosen to be the final model according to the AIC, see Table **??**. After model fitting, the $G(3)$ spectrum is shown, along with the spectrum generated from the fitted discrete $G(3)$ model with order 20. As shown in Figure 22, the two estimates are similar to each other, and both have the two close peaks. The two dominant dual frequencies as shown in Table 11, are 0.056 and 0.016, and these are the same as those given in the discrete model. To compare forecast performance, the last 60 points are forecast and shown in Figure 23. For a variety of forecast origins, the results

Figure 20: Realization of quadratic chirp

Figure 21: Deterministic Components of quadratic chirp

are listed in Table 10. The model diagnostics are shown in Figure 24.

For the two components associated with the dominant $G(3)$ frequencies, or dual frequencies at 0.056 and 0.016, which are the corresponding original components after reverting the time deformation, we apply the Kalman smoother. The estimated two dominant frequency components after mean correction are shown in Figure 25. Comparing Figure 21 and 25, it can be seen that the original components seem to be reproduced accurately. Similar results were achieved in Cohlmia et al. (2004) by applying a Butterworth filter to the dual process, which involves twice interpolations. However, the method discussed here, by a structural time domain approach, essentially avoids interpolations.

Example 6. A Doppler-type signal was considered in Cohlmia et al. (2004), in which the data are simulated from

$$X(t) = \sin(\frac{840\pi}{t+50}) + .5\sin(\frac{2415\pi}{t+50}) + .05\epsilon(t) \tag{18}$$

where $\epsilon(t)$ are standard normal variates.

Figure 22: Spectral estimates of quadratic chirp

Figure 23: Forecasting comparison for quadratic chirp

Figure 24: Model diagnostics for quadratic chirp

lag	AR	Discrete $G(\lambda)$	Continuous $G(\lambda)$	Continuous $G(\lambda)$ (wrong)
10	0.613	0.105	0.178	0.054
20	0.948	0.052	0.177	0.040
30	0.931	0.081	0.304	0.040
40	2.070	0.062	0.294	0.038
50	2.150	0.110	0.325	0.041
60	2.625	0.153	0.409	0.040

Table 10: MSE forecast comparison for quadratic chirp

	Roots	G-frequency	Dual Frequency
1	-0.001±0.349i	$7.0e-07$	0.056
2	-0.037±0.101i	$2.0e-07$	0.016
3	-0.161±1.238i	$2.4e-06$	0.197
4	-0.343±2.372i	$4.7e-06$	0.377
5	-1.987±4.606i	$9.0e-06$	0.733
6	-440.295	0.000	0.000

Table 11: Factors of fitted model for quadratic signal

Figure 25: Filtered data of quadratic chirp

Figure 26: Realization of Doppler signal

A realization with sample size 200 is shown in Figure 26 and the two deterministic components in (18) are plotted in Figure 27. These components have time varying frequencies, thus are nonstationary. To separate them, Cohlmia et al. (2004) described a discrete $G(\lambda)$ analysis by fitting a discrete $G(\lambda)$ model. The estimated model is order 12 with $\hat{\lambda} = -1.7, \hat{h} = 1.1888 \times 10^{-6}, \hat{\Lambda} = 90$. To apply a continuous $G(\lambda)$ analysis, the time deformation estimated from the discrete model is used to obtain unevenly spaced stationary dual data. These dual data are thus to be fitted using continuous AR models. The model selection statistics are shown in Table **??**, from which, an order 14 model is chosen. After model fitting, the $G(-1.7)$ spectrum is shown, along with the spectrum generated from the fitted discrete $G(-1.7)$ model with order 20. There seem to be three dominant dual frequencies as shown in Table 13, as 0.033, 0.100 and 0.091, but the last two are close together and jointly contribute to the second component. The last 20 points are forecast and shown in Figure 29. It can be seen that the continuous $G(\lambda)$ model outperforms the discrete $G(\lambda)$ model. The continuous model captures the cyclical structure of the data more accurately. The model diagnostics are shown in Figure 30, and the standardized residuals pass the Box-Ljung white noise tests for lags 24 and 48.

Figure 27: Deterministic components of Doppler signal

Figure 28: Spectral estimates of Doppler signal

To estimate unobserved components associated with the dominant $G(-1.7)$ frequencies, the Kalman smoothing algorithm was applied and the estimated components after mean correction were shown in Figure 31. Comparing Figure 27 and 31, it can be seen that the two components are almost perfectly recovered. Similar results were achieved in Cohlmia et al. (2004).

lag	Discrete $G(\lambda)$	Continuous $G(\lambda)$	Continuous $G(\lambda)$ (wrong)	
10	0.7901	0.0032	0.0021	0.0028
20	0.4737	0.0581	0.0033	0.003
30	0.3741	0.0828	0.0052	0.003
40	0.3909	0.0721	0.0035	0.003
50	0.388	0.0506	0.0415	0.003
60	0.5276	0.0386	0.0988	0.004

Table 12: Forecast comparison for Doppler signal

Figure 29: Forecasting comparison for Doppler signal

Figure 30: Model diagnostics for Doppler signal

	Roots	G-frequency	Dual Frequency
1	-0.001±0.205i	27429	0.033
2	-0.008±0.626i	83858	0.100
3	-0.011±0.572i	76585	0.091
4	-0.255±1.478i	197836	0.235
5	-0.418±2.292i	306792	0.365
6	-0.838±4.838i	647657	0.770
7	-1.652±11.201i	1499557	1.783

Table 13: Factors of fitted model for Doppler signal

Figure 31: Filtered data of Doppler signal

5 Discussions

We believe the approach taken in this paper may be applicable to other time series applications subject to time deformation transforms. Recently, time deformation has received broad interest. For instance, Flandrin et al. (2003) considered the Lamperti transformation and self-similar processes. The work of Gray and Zhang (1988) was treated there to be a weakened form of Lamperti's theorem and M-stationary processes were specific classes of self-similar processes.

6 Acknowledgments

We are grateful to Dr. G. Tunnicliffe Wilson for providing the Fortran code to fit continuous time autoregressive models and for helpful communications with him. We wish to thank Curtis Condon, Ken White, and Al Feng of the Beckman Institute of the University of Illinois for the bat data and for permission to use it in this paper.

References

Belcher, J., Hampton, J. S., and Tunnicliffe Wilson, G. (1994). Parameterization of continuous time autoregressive models for irregularly sampled time series data. *Journal of the Royal Statistical Society, Series B, Methodological*, 56:141–155.

Choi, E.-H. (2003). *EARMA Processes with Spectral Analysis of Nonstationary Time Series*. PhD thesis, Southern Methodist University.

Cohlmia, K. B., Gray, H. L., Woodward, W. A., and Jiang, H. (2004). On filtering time series with monotonic varying frequencies. *Submitted*.

Flandrin, P., Borgnat, P., and Amblard, P.-O. (2003). From stationarity to self-similarity, and back: Variations on the Lamperti transformation. In Raganjaran, G. and Ding, M., editors, *Processes with Long-Range Correlations*, pages 88–117. Springer.

Forbes, E. and Schaik, A. (2000). Fourier transform and wavelet transform for the time-frequency analysis of bat echolocation signals. Technical report, The University of Sydney. http://www.eelab.usyd.edu.au/andre/avs_techrep.html.

Gray, H. L., Vijverberg, C. P. C., and Woodward, W. A. (2005). Nonstationary data analysis by time deformation. *Communications in Statistics, Series A*. To appear.

Gray, H. L. and Zhang, N. F. (1988). On a class of nonstationary processes. *J. Time Ser. Anal.*, 9(2):133–154.

Jiang, H., Gray, H. L., and Woodward, W. A. (2004). Time-frequency analysis-G(λ) process. Technical report, Southern Methodist University.

Jones, R. H. (1981). Fitting a continuous time autoregression to discrete data. In *Applied Time Series Analysis II*, pages 651–682.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. (1992). *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, second edition edition.

Stock, J. H. (1988). Estimating continuous-time processes subject to time deformation: An application to postwar U.S. GNP. *Journal of the American Statistical Association*, 83:77–85.

Wang, Z. (2004). *The Application of the Kalman Filter to Nonstationary Time Series through Time Deformation*. PhD thesis, Southern Methodist University.