A NOTE ON THE EXTRACTION OF COMPONENTS FROM TIME SERIES

Adrian Pagan

Econometric Research Program
Research Memorandum No. 148
May 1973

The research described in this paper was supported by NSF Grant GS 32003X.

Econometric Research Program
PRINCETON UNIVERSITY
207 Dickinson Hall
Princeton, New Jersey

A NOTE ON THE EXTRACTION OF COMPONENTS FROM TIME SERIES*

Adrian Pagan

1. Introduction

In recent years some attention has been directed to the concept of a time series as the summation of a number of distinct components to which economic agents may react in differing ways. Applications of such a schema have included the modelling of lags in economic behaviour [12] [13], the determination of a structural model to give adaptive forecasting "optimal" properties [14] [17], the decomposition of income into its permanent and transitory parts [5] and, of course, the older objective of distinguishing between trend, seasonal and irregular characteristics [7]. Much of the work has been theoretical in nature -- although a notable exception is [11] -- and this may have been a consequence of the method employed for extracting components, i.e., the Wiener-Kolmogorov filtering theory, as set out in Whittle [18] and employed by Grether and Nerlove [4] [6] [7].

To briefly summarize this approach, consider a series $y(t)$ with an imbedded component $z(t)$ which is to be estimated;

* I would like to thank Gregory C. Chow and Alan J. Preston for helpful comments on an earlier draft.

then the theory provides a backward filter of the form

$\hat{z}(t) = \sum_{j=0}^{\infty} \gamma(j)\, y(t-j)$ to yield the minimum mean square estimate $\hat{z}(t)$ of $z(t)$. The parameters $\gamma(j)$ are, in the most general case, derived from the spectrum of $y(t)$ but, under suitable assumptions, they may also be regarded as functions of the roots of certain high-order polynomials (see [7] for details) and, as the exact relationship is complex, must be solved by numerical techniques.

Three difficulties stand out in attempts to apply the Wiener-Kolmogorov theory:

(i) Numerical procedures for locating the zeroes of polynomials are notoriously unstable and therefore should be avoided. In this instance one should be doubly careful as the polynomials are frequently ill-conditioned.

(ii) Although the infinite backward filter summarized in the parameters $\{\gamma(j), j=0,\dots\}$ can be approximated by a rational function, thereby converting the filter to a recursive one, there is no set procedure for doing this.

(iii) The theory is uni-dimensional, i.e., confined to single series.

It is the purpose of this note to indicate how these difficulties may be surmounted, thereby integrating the Grether-Nerlove approach to components into a framework that is both easily comprehended and amenable to solution by standard matrix operations upon a computer. To achieve the latter objective we will employ Kalman-Bucy filtering theory [10], so that section 2 is devoted to a statement of its major propositions while section 3 formulates the component extraction problem in the Kalman-Bucy framework. Section 4 describes an estimation technique for the parameters of the model while the fifth section gives an application to an economic time series. Throughout, a trend /seasonal/ irregular model is assumed for expository purposes.

2. Kalman Filtering Theory

With only slight modifications the sources of the following statements on Kalman-Bucy filtering theory are [16] -- this being chosen as a convenient reference for economists. Given a $(q \times 1)$ unobservable vector $z(t)$, later referred to as the state vector, and a $(p \times 1)$ observable vector $d(t)$ connected by the system

$$d(t) = H\, z(t) + v(t) \tag{2.1}$$

$$z(t) = M\, z(t-1) + w(t) \tag{2.2}$$

where $v(t)$ and $w(t)$ are uncorrelated Gaussian processes with covariance matrices R and Q respectively, the minimum mean square error forecast of $z(t)$ given $d(t)$ is $\hat{z}(t|t)$ where

$$\hat{z}(t|t) = M\,\hat{z}(t-1|t-1) + P(t)\,H'R^{-1}[d(t) - HM\,\hat{z}(t-1|t-1)] \quad (2.3)$$

$$P(t) = [\bar{P}^{-1}(t) + H'R^{-1}H]^{-1} \quad (2.4)$$

$$\bar{P}(t) = M\,P(t-1)\,M' + Q\,, \quad (2.5)$$

and it is assumed that an _a priori_ normal estimate with mean $\hat{z}(0)$ and covariance matrix $P(0)$ of the initial state vector $z(0|0)$ is available to start the recurrence. Although $\hat{z}(0)$ and $P(0)$ are seldom available the difficulty may be overcome by use of the following theorem ([16, p. 99]).[1]

<u>THEOREM:</u> If the equations (2.1) and (2.2) are observable and controllable then (i) the matrix equations (2.4) and (2.5) tend uniformly to a constant positive definite matrix P^* if $P(0)$ is non-negative definite (ii) the equation (2.3) written in the form $\hat{z}(t|t) = [M - P^*\,H'R^{-1}HM]\,\hat{z}(t-1|t-1) + P^*\,H'R^{-1}d(t)$ is uniformly asymptotically stable, i.e., all eigenvalues of the matrix $[M - P^*\,H'R^{-1}H\,M]$ lie within the unit circle.[2]

---

[1] One case in which $\hat{z}(0)$ and $P(0)$ are available is when parameter estimation has been by Bayesian techniques.

[2] Observability and controllability are defined in any text on control theory, e.g. Astrom [1].

From the above theorem it is apparent that the establishment of observability and controllability is sufficient to guarantee that the initial conditions cease to be important as the sample size grows, allowing their choice to be somewhat arbitrary. Of course this statement must be tempered by the knowledge that the rapidity of convergence to the equilibrium value P^*, and the influence of the initial condition $z(0)$, will be a function of the values chosen so that one should select wisely. As will be seen in section 4, the Grether-Nerlove type of component model reduces to a mixed autoregressive moving-average (ARMA) equation and, in effect, the above theorem is satisfied only if the conditions for the identification of the parameters of such equations are fulfilled; the seminal analysis of these is Hannan [8].

3. A Components Model

For illustrative purposes it is useful to select one of the models in [7] as, with one exception to be dealt with at the close of this section, generalization of the approach to other models is trivial. Specifically, the series to be decomposed $(y(t))$ is regarded as the sum of mutually uncorrelated trend $(T(t))$, seasonal $(S(t))$ and irregular $(I(t))$ terms. In the spirit of [4] the parametric form is

$$y(t) = T(t) + S(t) + I(t) \tag{3.1a}$$

$$(1-\beta_1 L - \beta_2 L^2)\, T(t) = e_1(t) \tag{3.1b}$$

$$(1 - \beta_3 L^4)\, S(t) = e_2(t) \tag{3.1c}$$

$$I(t) = e_3(t) \tag{3.1d}$$

where L is the lag operator with the property $L^k\, x(t) \to x(t-k)$ and (3.1) will be referred to as the structural form of the model while, after all substitutions have been made, the remaining single equation will be designated the reduced form.[3]

(3.1b) - (3.1d) may be written as

$$\begin{bmatrix} T(t) \\ T(t-1) \\ S(t) \\ S(t-1) \\ S(t-2) \\ S(t-3) \end{bmatrix} = \begin{bmatrix} \beta_1 & \beta_2 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta_3 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} T(t-1) \\ T(t-2) \\ S(t-1) \\ S(t-2) \\ S(t-3) \\ S(t-4) \end{bmatrix} + \begin{bmatrix} e_1(t) \\ e_2(t) \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{3.2}$$,

---

[3]The justification for the chosen forms is given in detail in [7]; suffice it to say that the spectral properties of the modelled components accord with _a priori_ notions. Note that, whereas [7] was concerned with monthly series, it is convenient for us to deal with models for quarterly series here; because of this the model for the seasonal factor undergoes an obvious modification.

while (3.1a) is

$$y(t) = [1\ 0\ 1\ 0\ 0\ 0]\begin{bmatrix} T(t) \\ T(t-1) \\ S(t) \\ S(t-1) \\ S(t-2) \\ S(t-3) \end{bmatrix} + e_3(t) \qquad (3.3).$$

Now (3.2) and (3.3) have the respective forms

$$z(t) = M\, z(t-1) + w(t)$$

$$d(t) = H\, z(t) + v(t) \quad ,$$

which is the system format required to extract optimal estimates of $z(t)$ by the Kalman filter. Given the assumption that $E(e_j(t)\, e_k(t-s)) = \{ {\sigma_{jj} \atop 0} \quad {j=k,\ s=0 \atop s \neq 0} \}$ the covariance matrix of $w(t)$ and $v(t)$ is easily derived, so that, for any $\beta_1, \beta_2, \beta_3$, Q and R, it is possible to obtain estimates of $T(t)$ and $S(t)$. To illustrate this take $\beta_1 = -1.7$, $\beta_2 = 0.7125$, $\beta_3 = 0.9$, $\sigma_{jj} = 1$ (j=1,2.3) (the β_j are taken from [7] and σ_{jj} are arbitrary). Then application of (2.3) - (2.5) yields the recursive equations (where the steady-state version of the matrix equations (2.4) and (2.5) is used):

$$\hat{T}(t|t) = 1.7\ \hat{T}(t-1|t-1) - 0.7125\ \hat{T}(t-2|t-2) + 0.596\ \Phi(t)$$

$$\hat{T}(t-1|t) = \hat{T}(t-1|t-1) + 0.299\ \Phi(t)$$

$$\hat{S}(t|t) = 0.9\ \hat{S}(t-4|t-4) + 0.253\ \Phi(t)$$

$$\hat{S}(t-1|t) = \hat{S}(t-1|t-1) - 0.177\ \Phi(t)$$

$$\hat{S}(t-2|t) = \hat{S}(t-2|t-2) - 0.055\ \Phi(t)$$

$$\hat{S}(t-3|t) = \hat{S}(t-3|t-3) - 0.012\ \Phi(t)$$

and

$$\Phi(t) = y(t) - 1.7\ \hat{T}(t-1|t-1) + 0.7125\ \hat{T}(t-2|t-2) - 0.9\ \hat{S}(t-4|t-4) \quad .$$

Some confusion may be caused by the presence of equations for $\hat{T}(t-1|t)$, $\hat{S}(t-1|t)$, $\hat{S}(t-2|t)$ and $\hat{S}(t-3|t)$, but this is resolved by noting that the Kalman filter yields the optimal extraction based on information up to time t, while, in the above model, because of the dependence in the series, an improved estimate of $S(t-1)$ etc. may be made as later observations become available; to incorporate this feature it is necessary that there be extra updating equations.[4] Once estimates of the

[4] The Kalman filter quoted in section 2 is concerned with the best estimate of $z(t)$ given $y(0),\ldots,y(t)$. A more general formulation might be to ask for the best estimate of $z(t)$ given

trend and seasonal, $\hat{S}(t)$ and $\hat{T}(t)$, are obtained, it is a simple matter to find $\hat{I}(t) = y(t) - \hat{S}(t) - \hat{T}(t)$ and to decompose any new observation into its constituent parts. Moreover the recursive nature of the computations represents a great simplification over the complex computations required to factorize the high order polynomials associated with the optimal filtering formulae given by the Kolmogorov-Wiener theory.[5]

There is one extension needed to the above presentation viz. when one of the components (say $S(t)$) is not represented by a pure autoregression (AR) as above but by an ARMA form. In fact, such a model was propounded in [7] on the basis that a moving average (MA) term was needed to generate spectral peaks in the seasonal that decline over the range $0 - \pi$. To handle ARMA processes it is only necessary to augment the state vector e.g., if the seasonal is taken to be

---

$y(0),\ldots,y(t+\tau)$ i.e., a combined backward and forward filter. Essentially the equations for $\hat{S}(t-1|t)$, $\hat{S}(t-2|t)$, $\hat{S}(t-3|t)$ correspond to the optimal filter with $\tau=1,2,3$. A complete generalization of the Kalman filter is possible but it would increase the length of this note.

[5]As Taylor says (p. 98), "Kalman's results are an ingenious formulation of classical Wiener filtering theory in a computationally practical form."

$$(1-\beta_3 L^4)\, S(t) = (1+\alpha_1 L)\, e_2(t) ,$$

the state dynamics are now

$$\begin{bmatrix} T(t) \\ T(t-1) \\ S(t) \\ S(t-1) \\ S(t-2) \\ S(t-3) \\ e_2(t) \end{bmatrix} = \begin{bmatrix} \beta_1 & \beta_2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta_3 & \alpha_1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} T(t-1) \\ T(t-2) \\ S(t-1) \\ S(t-2) \\ S(t-3) \\ S(t-4) \\ e_2(t-1) \end{bmatrix} + \begin{bmatrix} e_1(t) \\ e_2(t) \\ 0 \\ 0 \\ 0 \\ 0 \\ e_2(t) \end{bmatrix} ,$$

and the analysis continues as before.

4. Estimation

Although the estimation technique will not be described in detail (this has been done in [15] and an alternative method may be found in [11]) it is of interest to consider the simple example of (4.1) -- utilized in the following section -- for an exposition of the approach.

$$
\begin{aligned}
y(t) &= T(t) + S(t) + I(t) \\
(1-L)T(t) &= e_1(t) \\
(1-\beta L^4)S(t) &= e_2(t) \\
I(t) &= e_3(t)
\end{aligned}
\qquad (4.1)
$$

In (4.1) it is assumed that $E(e_i(t)\ e_j(t-s)) = \sigma_{ij}$ when $s=0$, $i=j$ and zero otherwise. The reduced form of (4.1) is

$$
(1-L)(1-\beta L^4)\ y(t) = u(t) = (1-\beta L^4)\ e_1(t) + (1-L)\ e_2(t) + + (1-L)(1-\beta L^4)\ e_3(t)
$$

Setting $\lambda_1 = \sigma_1^2/\sigma_3^2$ and $\lambda_2 = \sigma_2^2/\sigma_3^2$ the autocovariances of $u(t)$ are (when divided by the constant σ_3^2)

$$
\begin{aligned}
\gamma_u(0) &= (1+\beta^2)\ \lambda_1 + 2\lambda_2 + 2(1+\beta^2) \\
\gamma_u(1) &= -\lambda_2 - 1 \\
\gamma_u(2) &= 0 \\
\gamma_u(3) &= 0 \\
\gamma_u(4) &= -\lambda_1\beta - \beta \\
\gamma_u(5) &= \beta
\end{aligned}
$$

and $\gamma_u(j)$ $(j=6,\ldots) = 0$. As this is the covariance function of a fifth order moving average (MA) process an alternate representation for $u(t)$ is

$$u(t) = (1+\alpha_1 L + \alpha_2 L^2 + \alpha_3 L^3 + \alpha_4 L^4 + \alpha_5 L^5)\ \epsilon(t) . \qquad (4.2)$$

Denoting the covariances from the MA approximation to $u(t)$ by $\tilde{\gamma}_u(j)$, i.e., $\tilde{\gamma}_u(j)$ are formed from $\alpha_1,\ldots,\alpha_5$ and $\varphi = (\sigma_\epsilon^2/\sigma_3^2)$, the task is to find estimates of $\alpha_1,\ldots,\alpha_5$ and φ that minimize the deviation $\gamma_u(j) - \tilde{\gamma}_u(j)$.[6] Wilson [19] has provided an algorithm for this and generally the resulting discrepancies are of the order of 10^{-14} (if double precision arithmetic is used). Therefore, given estimates for $\theta = (\lambda_1, \lambda_2, \beta)$, it is possible to obtain estimates for $\alpha_1,\ldots,\alpha_5$ and φ . Furthermore, by using the estimates of $\alpha_1,\ldots,\alpha_5$ so obtained, the sum of squares $S = \sum_{j=0}^{N=1} \hat{\epsilon}^2(j)$ may be computed.[7] By this procedure it has been possible to evaluate the sum of squares corresponding to any given values of λ_1 , λ_2 and β and this information is sufficient for the

---

[6] As $\gamma_u(j)$ were normalized with σ_3^2 so must $\hat{\gamma}_u(j)$ and this results in the factor $\sigma_\epsilon^2/\sigma_3^2$.

[7] Note that $\epsilon(-1),\ldots,\epsilon(-q)$ are required to compute the $\hat{\epsilon}(t)$ if this is done from (4.2), but, for convenience, these may be set to zero. Box and Jenkins [2] have derived an estimator of these "nuisance" parameters which could be applied.

application of non-linear estimators designed to find the $\hat{\theta}$ minimizing S. The particular algorithm chosen was a modification of the Gauss-Newton algorithm [9]. Finally, when the minimum value $\hat{S}$ has been achieved, an estimate of σ_ϵ^2 may be formed from $\hat{S}/N$ and $\hat{\sigma}_3^2 = \hat{\varphi}\,\sigma_\epsilon^2$.[8]

5. Decomposing a Consumption Series

As an illustration of the foregoing methodology the series, Quarterly Personal Consumption Expenditure on Food ($ m) -- one of the thirteen constituents of consumption expenditure in the Australian National Accounts -- was modelled by (4.1) but with $\beta = 0.95$.[9] Observations were available for the period September 1950 to June 1970 inclusive, yielding eighty data points. Proceeding as in section 4 (under the assumption that $\epsilon(-1),\ldots,\epsilon(-5)$ equaled zero) yielded the following parameter estimates:

---

[8] The steps needed to apply the approach to models other than (4.1) are (i) Solve for the reduced form (ii) Determine the covariance matrix of $u(t)$, (iii) apply Wilson's algorithm to find the MA process that matches this covariance matrix.

[9] An attempt was made to estimate β but it was close to unity. When $\beta=1$ the identification conditions are broken but, even for β close to unity, it was found that rounding error lead to a lack of convergence in the Kalman filter. Given the length of the series, a fine discrimination between a non-stationary ($\beta=1$) and stationary ($\beta < 1$) signal seems impossible. It is of some interest to note that the residuals $\hat{\epsilon}(t)$ passed the tests given by Box and Pierce [3] for whiteness.

$$\lambda_1 = 1.18 \qquad \lambda_2 = 4.14 \qquad \sigma_\epsilon^2 = 10.11$$

with the implied MA parameters

$$\alpha_1 = -0.11, \quad \alpha_2 = -0.02, \quad \alpha_3 = -0.03, \quad \alpha_4 = -0.19, \quad \alpha_5 = 0.03$$

and $\sigma_3^2 = 3.47$. Using the equilibrium value of (2.4) and (2.5) the filtering equations were (where $\Phi(t) = \hat{T}(t-1|t-1) + 0.95 \hat{S}(t-4|t-4)$)

$$\hat{T}(t|t) = \hat{T}(t-1|t-1) + 0.36(y(t) - \Phi(t))$$

$$\hat{S}(t|t) = 0.95\ \hat{S}(t-4|t-4) + 0.53\ (y(t) - \Phi(t))$$

$$\hat{S}(t-1|t) = \hat{S}(t-1|t-1) - 0.20\ (y(t) - \Phi(t))$$

$$\hat{S}(t-2|t) = \hat{S}(t-2|t-2) - 0.12\ (y(t)) - \Phi(t))$$

$$\hat{S}(t-3|t) = \hat{S}(t-3|t-3) - 0.06\ (y(t) - \Phi(t)) \quad .$$

Utilizing these equations with starting values of $\hat{T}(0) = 200$ and $\hat{S}(0) = 50$ yields the estimates of trend and seasonal contained in Table 1.

As the model is almost certainly not the best one for the series the results should not be taken very seriously, but of interest is the strong seasonal component in the December quarter and the large trend in the series. Each of these features was visible in a graph of the series.

TABLE 1

Estimates of Trend and Seasonal in the Food Series
Sept. 1966 to June 1970

Year	Seasonal	Trend
1966/3	49.83	675.5
1966/4	95.16	688.3
1967/1	35.90	702.5
1967/2	39.58	711.5
1967/3	51.63	722.0
1967/4	95.35	732.3
1968/1	41.54	743.3
1968/2	45.43	753.4
1968/3	45.56	757.6
1968/4	96.97	768.0
1969/1	38.08	775.0
1969/2	41.15	783.6
1969/3	43.56	795.5
1969/4	103.27	812.6
1970/1	44.46	825.6
1970/2	68.36	845.6

6. Conclusion

Of the three objectives of the paper outlined at the beginning two have been accomplished and illustrated with numerical examples, vis. standard matrix operations were employed to construct the filters and recursive forms were given. The final objective of simultaneously extracting the components from a number of time series was not explicitly dealt with but, as the filtering theory is multi-dimensional and the estimation technique clearly extends to joint estimation for a number of series, generalization is straightforward.

REFERENCES

[1] Astrom, K.J., Introduction to Stochastic Control Theory (Academic Press, New York, 1970).

[2] Box, G.E.P. and G.M. Jenkins, Time Series Analysis: Forecasting and Control (Holden-Day, San Francisco, 1970).

[3] Box, G.E.P. and D.A. Pierce, "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," Journal of the American Statistical Association, 64 (1970), 1509-1526.

[4] Couts, D., D. Grether and M. Nerlove, "Forecasting Non-Stationary Economic Time Series," Management Science, 13 (1966), 1-22.

[5] Friedman, M., A Theory of the Consumption Function (Princeton University Press, Princeton, 1957).

[6] Grether, D.M., "Studies in the Analysis of Economic Time Series," (unpublished doctoral dissertation, Stanford University, 1969).

[7] ______________, and M. Nerlove, "Some Properties of 'Optimal' Seasonal Adjustment," Econometrica, 38 (1970), 682-704.

[8] Hannan, E.J., "The Identification of Vector Mixed Auto-regressive Moving Average Systems," Biometrika, 56 (1969), 223-225.

[9] Hartley, W., "The Modified Gauss-Newton Method for the Fitting of Non-Linear Regression Functions by Least Squares," Technometrics, 3 (1961), 269-280.

[10] Kalman, R.E. and R.S. Bucy, "New Results in Linear Filtering and Prediction Theory," Journal of Basic Engineering, Transactions of ASME, 83D (1961), 95-108.

[11] Nerlove, M., "Analysis of Economic Time Series by Box-Jenkins and Related Techniques," Report No. 7156 of the Center for Mathematical Studies in Business and Economics, University of Chicago.

[12] ___________, "Distributed Lags and Unobserved Components," in W. Feller (ed.) Ten Economic Studies in the Tradition of Irving Fisher (John Wiley, New York, 1967), 127-169.

[13] ___________, "Lags in Economic Behavior," Econometrica, 40 (1972), 221-251.

[14] Nerlove, M. and S. Wage, "On the Optimality of Adaptive Forecasting," Management Science, 10 (1964), 207-224.

[15] Pagan, A.R., "Efficient Estimation of Models with Composite Disturbance Terms," Journal of Econometrics (forthcoming).

[16] Taylor, L., "The Existence of Optimal Distributed Lags," The Review of Economic Studies, 37 (1970), 95-106.

[17] Theil, H., and S. Wage, "Some Observations on Adaptive Forecasting," Management Science, 10 (1964), 198-206.

[18] Whittle, P., Prediction and Regulation (The English Universities Press, London, 1963).

[19] Wilson, G.T., "Factorization of the Covariance Generating Function of a Pure Moving Average Process," S.I.A.M. Journal of Numerical Analysis, 6 (1969), 1-7.