# Department of Economics

## Signal Extraction, Maximum Likelihood Estimation and the Start-up Problem

Stephen Pollock

Working Paper No. 433 February 2001 ISSN 1473-0278

Queen Mary
University of London

# Signal Extraction, Maximum Likelihood Estimation and the Start-up Problem

**by D.S.G. POLLOCK**

Queen Mary, University of London

*and*

GREQAM: Groupement de Recherche en Economie Quantitative d'Aix–Marseille

*Email:* stephen_pollock@sigmapi.u-net.com

In this paper, we portray the essential features of the finite-sample signal extraction problem in both the stationary and the nonstationary cases. The computational procedures can be simplified in the light of our analysis. An important outcome of the analysis is a demonstration that the start-up problem can be handled far more easily that one might expect from a passing acquaintance with the usual practices.

*Date:* January, 2001

*Key words:* Signal extraction, Linear filtering, Trend estimation

*JEL Classification:* C22

## 1. Introduction

The treatment of signal-extraction problems in econometrics rests heavily upon the Wiener–Kolmogorov theory, which is a mainstay of communications engineering. The theory envisages that the data are generated by stationary stochastic processes and that they form a lengthy sequence. Therefore, the classical results of the Wiener–Kolmogorov theory have been developed in the context of the theoretically tractable assumption that the data form a doubly infinite sequence stretching into the indefinite past and the indefinite future. (For the original expositions, see Kolmogorov [5] and Wiener [12]. For an accessible modern treatment of the theory see Whittle [13].)

In practice, econometric time series are of a strictly limited duration and they are often highly nonstationary; and some radical adaptations of the Wiener–Kolmogorov theory are necessary if it is to be applied successfully in such circumstances. In econometric analysis, the adaptations have been accomplished, typically, by adopting the device of the Kalman filter and its associated smoothing algorithms, (see Harvey [2] and Harvey and Todd [3]). The non-stationarity of the data necessitates a careful handling of the start-up problem; and this has been accomplished via a development known as the diffuse Kalman filter, (see De Jong [1], whose algorithm has been used by Koopman *et al.* [6]).

The Kalman filter is a very flexible device which is also a complex one, as it must be if it is to achieve a high degree of generality. (For an account of the Kalman filter and of the associated smoothing algorithms, see Pollock [9]). Part of the complexity of its associated algorithms is due to fact that they are intended for online processing, which is rarely a requirement in econometric analysis. Therefore, there is scope for simplifying the treatment of the econometric problem, the essential features of which are masked by the complexity of the Kalman filter. In fact, the econometric problem is a relatively simple one which can be approached in a direct manner.

In this paper, we portray the essential features of the finite-sample signal extraction problem in both the stationary and the nonstationary cases. In the light of our analysis, the computational procedures can be simplified somewhat. An important outcome of the analysis is a demonstration that the start-up problem can be handled far more easily that one might expect from a passing acquaintance with the usual practices.

## 2. The Statistical Assumptions

Consider a vector

$$y = \xi + \eta, \tag{1}$$

of $T$ observations, which consists of a signal component $\xi$ and a noise component $\eta$. It is assumed that

$$\xi \sim N(0, \Omega_\xi) \quad \text{and} \quad \eta \sim N(0, \Omega_\eta) \tag{2}$$

are independent, normally distributed, random vectors generated by two contemporaneous stationary stochastic processes. The assumption of independence implies that

$$\begin{aligned} D(y) &= \Omega_\xi + \Omega_\eta, \\ C(\xi, y) &= \Omega_\xi \quad \text{and} \\ C(\eta, y) &= \Omega_\eta. \end{aligned} \tag{3}$$

The joint density function of $\xi$ and $\eta$ is

$$N(\xi, \eta) = (2\pi)^{-T} |\Omega_\xi|^{-1/2} |\Omega_\eta|^{-1/2} \exp\{-\frac{1}{2}(\xi'\Omega_\xi^{-1}\xi + \eta'\Omega_\eta^{-1}\eta)\}, \tag{4}$$

whereas the density function of their sum $y = \xi + \eta$ is

$$N(y) = (2\pi)^{-T/2} |\Omega_\xi + \Omega_\eta|^{-1/2} \exp\{-\frac{1}{2} y'(\Omega_\xi + \Omega_\eta)^{-1} y\}. \tag{5}$$

An alternative assumption regarding $\xi$ is that it represents a trend component generated by a nonstationary ARIMA process incorporating $d$ roots of unity. Then $\xi$ will be reduced to stationarity by a compound matrix difference operator $Q'$ of order $(T-d)\times T$, with the result that

$$Q'\xi = \zeta \sim N(0, \Omega_\zeta). \tag{6}$$

In that case, we must consider, in place of (4), the density function

$$N(\zeta,\eta) = (2\pi)^{-(2T-d)/2}|\Omega_\zeta|^{-1/2}|\Omega_\eta|^{-1/2}\exp\{-\frac{1}{2}(\xi'Q\Omega_\zeta^{-1}Q'\xi+\eta'\Omega_\eta^{-1}\eta)\}, \tag{7}$$

whereas (5) must be replaced by

$$N(g) = (2\pi)^{-(T-d)/2}|\Omega_\zeta + Q'\Omega_\eta Q|^{-1/2}\exp\{-\frac{1}{2}g'(\Omega_\zeta + Q'\Omega_\eta Q)^{-1}g\}, \tag{8}$$

where $g = Q'y$ is the differenced version of the data vector.

In statistical signal extraction, we face two related problems. The first problem is to estimate the parameters of the distributions of $\xi$ and $\eta$. The second problem is to find estimates of $\xi$ and $\eta$ given the data vector $y$ and given the values of the statistical parameters, which amounts to knowing $\Omega_\xi$ or $\Omega_\zeta$ and knowing $\Omega_\eta$. We shall concentrate on the second of these two problems, and we shall pass a few comments on the first problem at the end of the paper.

## 3. Extracting the Latent Components

The problem of estimating $\xi$ and $\eta$, in the case where $\xi$ is stationary, can be construed as a matter of maximising the likelihood function $N(\xi,\eta)$ of (4) subject to the condition that $\xi+\eta = y$. This entails minimising a chi-square criterion function:

$$S(\xi) = (y-\xi)'\Omega_\eta^{-1}(y-\xi) + \xi'\Omega_\xi^{-1}\xi. \tag{9}$$

It makes no odds whether we consider this criterion function or an analogous function with $\eta = y-\xi$ as its argument. The eventual results will be the same.

Differentiating $S$ with respect to $\xi$ and setting the result to zero gives the first-order condition from which the minimising value can be obtained:

$$x = (\Omega_\xi^{-1} + \Omega_\eta^{-1})^{-1}\Omega_\eta^{-1}y. \tag{10}$$

However, a familiar matrix identity concerning the inverse of a sum of matrices (see Pollock [9, p. 228], for example) enables us to write

$$\begin{aligned}(\Omega_\xi^{-1} + \Omega_\eta^{-1})^{-1} &= \Omega_\xi - \Omega_\xi(\Omega_\xi+\Omega_\eta)^{-1}\Omega_\xi\\ &= \Omega_\eta - \Omega_\eta(\Omega_\xi+\Omega_\eta)^{-1}\Omega_\eta.\end{aligned} \tag{11}$$

Using the second of these identities, we get

$$
(12)\qquad \begin{aligned} x &= \{I - \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}\}y \\ &= y - h, \end{aligned}
$$

where $h = \Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1}y$ is the estimate of $\eta$. Since $\Omega_\eta(\Omega_\xi + \Omega_\eta)^{-1} + \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1} = I$, we can also write $x = \Omega_\xi(\Omega_\xi + \Omega_\eta)^{-1}y$.

It is easy to see, in reference to (3), that

$$
(13)\qquad \begin{aligned} x &= E(\xi|y) = E(\xi) + C(\xi, y)D^{-1}(y)\{y - E(y)\} \quad \text{and} \\ h &= E(\eta|y) = E(\eta) + C(\eta, y)D^{-1}(y)\{y - E(y)\} \end{aligned}
$$

are equally the conditional expectations and the minimum-mean-square-error estimates of $\xi$ and $\eta$. Here, we have $E(\xi) = E(\eta) = 0$ and, therefore, $E(y) = 0$.

Now consider the case where $\xi$ is generated by a nonstationary ARIMA process. Then the chi-square criterion function of (9), by which the likelihood is maximised, is replaced by

$$
(14)\qquad S(\xi) = (y - \xi)'\Omega_\eta^{-1}(y - \xi) + \xi' Q\Omega_\zeta^{-1}Q'\xi,
$$

of which the minimising value is

$$
(15)\qquad x = (Q\Omega_\zeta^{-1}Q' + \Omega_\eta^{-1})^{-1}\Omega_\eta^{-1}y.
$$

The matrix inversion lemma indicates that

$$
(16)\qquad (Q\Omega_\zeta^{-1}Q' + \Omega_\eta^{-1})^{-1} = \Omega_\eta - \Omega_\eta Q(Q'\Omega_\eta Q + \Omega_\zeta)^{-1}Q'\Omega_\eta,
$$

and it follows that

$$
(17)\qquad \begin{aligned} x &= \{I - \Omega_\eta Q(Q'\Omega_\eta Q + \Omega_\zeta)^{-1}Q'\}y \\ &= y - h, \end{aligned}
$$

Equation (17) seems markedly more complicated that equation (12). However, the elaboration is due entirely to the matrix difference operator $Q'$ which will have zeros everywhere except on $d + 1$ diagonal bands. It is unlikely that $d$ will exceed 2.

The task of computing $h$ can be accomplished by a handful of direct multiplications and recursions. The matter may be illustrated by the case where $\Omega_\zeta$ and $\Omega_\eta$ are variance–covariance matrices of moving-average processes. Therefore consider

$$
(18)\qquad \begin{aligned} h &= \Omega_\eta Q(Q'\Omega_\eta Q + \Omega_\zeta)^{-1}g \\ &= \Omega_\eta Qb. \end{aligned}
$$

The first task is to calculate $b$ by solving the equation

$$(Q'\Omega_\eta Q + \Omega_\zeta)b = g. \tag{19}$$

The solution is found via a Cholesky decomposition which sets $Q'\Omega_\eta Q + \Omega_\zeta = GG'$, where $G$ is a lower-triangular matrix. The system $GG'b = g$ can be cast in the form of $Gp = q$ and solved for $p$. Then $G'b = p$ can be solved for $b$. These are the recursive operations. Finding $h = \Omega_\eta Qb$ thereafter entails only direct multiplications.

If $\Omega_\xi$ and $\Omega_\eta$ are variance–covariance matrices of autoregressive processes, then, knowing that their inverses are matrices with a limited number of diagonal bands, we might choose to calculate the signal estimate $x$, in the stationary case, via the formula of (10).

In general, the matrix on the LHS of (19) has the structure of the variance–covariance matrix of an ARMA process. At its most complicated, it takes the form of $Q'\Omega_\eta Q + \Omega_\zeta = A^{-1}VA'^{-1}$, where $A$ is a banded lower-triangular Toeplitz matrix, due to the autoregressive part, and $V$ is a symmetric matrix with a limited number of nonzero bands around the diagonal, (see, for example, Pollock [9, Ch. 22]). The system $A^{-1}VA'^{-1}b = g$ can be cast in the form of $Vp = q$, where $p = A'^{-1}b$ and $q = Ag$. The vector $q$ can be calculated by a direct multiplication and, when this is available, the solution for $p$ can be obtained via the Cholesky decomposition of $V$. Then $b = A'p$ can be calculated.

## 4. Handling the Start-Up Problem

Problems often arise from not knowing how to supply the initial conditions with which to start the recursive filtering procedures that are entailed in the extraction of latent signals and other components of the data. By choosing inappropriate starting values, one can generate so-called transient effects which are liable, in fact, to affect all of the processed values.

The start-up problem can sometimes be overlooked without detriment when the data components are generated by zero-mean stationary stochastic processes. In that case, the zero-valued expectations of the process can serve as the initial conditions for the recursive filtering. However, what may seem remarkable is the absence of an explicit treatment of the problem in our foregoing analysis, which has been devoted to the extraction of nonstationary trends; and an explanation is called for.

For a simple explanation of how the problem has been handled, we can concentrate on equation (17) which provides the estimate $x$ of the tend. This estimate is obtained by subtracting the estimate $h$ of a stationary noise or residue component from the data vector $y$. Since the noise process has a zero-valued expectation, it transpires that the recursive procedures that generate the vector $h$ do not require explicit start-up values. However, there are some more intriguing explanations of the treatment of the start-up problem.

Let us consider reducing the data vector to stationarity using the matrix $Q'$. The result is

$$
\begin{aligned}
Q'y = g &= \zeta + \kappa \\
&= z + k,
\end{aligned} \tag{20}
$$

where $\zeta = Q'\xi$ and $\kappa = Q'\eta$ are the differenced versions of the trend and the residue, respectively, and where $z = Q'x$ and $k = Q'h$ are their estimates. It can be show, via the foregoing analysis of the signal-extraction problem in the stationary case, that

$$
\begin{aligned}
z &= \Omega_\zeta (Q'\Omega_\eta Q + \Omega_\zeta)^{-1} g \quad \text{and} \\
k &= Q'\Omega_\eta Q (Q'\Omega_\eta Q + \Omega_\zeta)^{-1} g.
\end{aligned} \tag{21}
$$

Now it might seem enough to remark that the estimate of $h$ of $\eta$, given by equation (18), is obtained from the estimate $k$ of $\kappa = Q'\eta$, given above, simply by stripping away the matrix factor $Q'$. However, since $Q'$ is a noninvertible matrix, such an argument is insufficient.

There are two legitimate ways of recovering $x$ and $h$ from $z$ and $k$ which are mathematically equivalent. The first way is to solve the following problem:

$$
\text{Minimise} \quad (y - x)'\Omega_\eta^{-1}(y - x) \quad \text{subject to} \quad Q'x = z. \tag{22}
$$

This is a matter of ensuring that the trend estimate $x$ follows the data sequence $y$ as closely as possible, subject to the condition that the differenced trend estimate $Q'x$ equals the estimate $z$ of (21). This prescription leads to the value of $x$ given in equation (17).

The second way of approaching the problem is to seek to generate the value of $h$ by accumulating the values of its differenced version $k$. For this purpose, it is necessary to estimate some starting values with which to begin the process of cumulation. Let $M = [S_*, S]$ be the inverse of the (invertible) matrix difference operator $\nabla = (I - L_T)^d$, where $L_T = [e_1, \ldots, e_{T-1}, 0]$ is the matrix lag operator which is obtained from the identity matrix $I_T = [e_0, e_1, \ldots, e_{T-1}]$ of order $T$ by deleting the leading column and by appending a vector of zeros to the last column. Then the estimate of $h$ will be given by the equation

$$
h = S_* k_* + Sk, \tag{23}
$$

where $k_*$ is a vector of $d$ initial conditions or start-up values. The start-up values are obtained by evaluating the criterion

$$
\text{Minimise} \quad (S_* k_* + Sk)'\Omega_\eta^{-1}(S_* k_* + Sk) \quad \text{with respect to} \quad k_*. \tag{24}
$$

As one might expect, this leads to the value of $h$ given in equation (18). A formal demonstration of this result had been provided by Pollock [11].

## 5. Determining the Statistical Parameters

The estimation of the statistical parameters is a common accompaniment of the business of extracting the latent components, but, as we shall suggest, there are circumstances in which the requisite parameters can be derived by other means. If the parameters are to be estimated, then it is appropriate to pursue the method of maximum likelihood.

If the components of the data are generated by stationary ARMA processes, then their sum will also be an ARMA process. In the special case where both the components are generated by moving-average processes, their sum will also be a moving-average process.

The likelihood function, in the case of stationarity, is provided by (5). In the case of a nonstationary data process, the data vector $y$ is generated by an ARIMA process, and it is the likelihood function pertaining to the differenced data $g = Q'y$ which must be considered. This is provided by equation (8).

The likelihood function can be written either in terms of the parameters of the component processes or in terms of the parameters of a composite univariate process which results from adding the components. Harvey and Todd [3] have described the former as the structural likelihood function and the latter as the reduced-form function. The parameters of the reduced form may be subject to certain restrictions arising out of the specification of the structural form.

If the reduced form is subject to restrictions, then it is easiest to estimate the structural parameters via the structural form of the likelihood function. This is the approach that has been pursued by Harvey [2].

If there are no restrictions on the reduced form, then the option exists of of inferring the estimates of the structural parameters from estimates of the parameters of the unrestricted reduced form. This approach has been taken by Hillmer and Tiao [4], by Pierce [8] and by Maravall and Pierce [7].

In an alternative approach to econometric signal extraction, the statistical parameters are determined by manipulation with a view to ensuring that the resulting signal-extraction filter has certain preconceived properties. A common objective is to derive a lowpass filter with a designated cut-off frequency for which the transition from the pass band to the stop band is a rapid as possible, given the constraints of the filter order and the need to maintain numerical stability.

The leading example in econometric analysis of a filter whose parameters are determined by manipulation is the so-called Hodrick–Prescott filter. This has only a single adjustable parameter. Another example is provided by the digital version of the Butterworth filter, the use of which has been advocated by the present author in a recent paper, (see Pollock [10]). The lowpass Butterworth filter, which is used in extracting the trend, can be specified by choosing a cut-off frequency, which marks the midpoint of the transition from the passband to the stopband, and by choosing the filter order, which governs the rate of transition—with the rate increasing as the filter order is increased.

The analogue version of the Butterworth filter is a time-honoured device in the field of electrical engineering; and, in the experience of the author, it has proved difficult to devise a discrete-time Wiener–Kolmogorov filter with a performance superior to that of the digital Butterworth filter.

## References

[1] De Jong, P., (1991), The Diffuse Kalman Filter, *The Annals of Statistics,* **19,** 1073–1083.

[2] Harvey, A.C., (1989), *Forecasting, Structural Time Series Models and the Kalman Filter,* Cambridge University Press, Cambridge.

[3] Harvey, A.C., and P.H. Todd, (1983), Forecasting Economic Time Series with Structural and Box–Jenkins Models: A Case Study, *Journal of Business and Economic Forecasting,* **1,** 299–307.

[4] Hillmer, S.C., and G.C. Tiao, (1982), An ARIMA-Model-Based Approach to Seasonal Adjustment, *Journal of the American Statistical Association,* **77,** 63–70.

[5] Kolmogorov, A.N., (1941), Interpolation and Extrapolation, *Bulletin de l'academie des sciences de U.S.S.R., Ser. Math.,* **5,** 3–14.

[6] Koopman, S.J., A.C. Harvey, J.A. Doornick and N. Shephard, (1995), *STAMP 5.0: Structural Time Series Analyser Modeller and Predictor—The Manual,* Chapman and Hall, London.

[7] Maravall, A., and D.A. Pierce, (1987), A Prototypical Seasonal Adjustment Model, *Journal of Time Series Analysis,* **8,** 177–193.

[8] Pierce, D.A., (1979), Signal Extraction in Nonstationary Time Series, *The Annals of Statistics,* **6,** 1303–1320.

[9] Pollock, D.S.G., (1999), *A Handbook of Time-Series Analysis, Signal Processing and Dynamics,* The Academic Press, London.

[10] Pollock, D.S.G., (2000), Trend Estimation and Detrending via Rational Square Wave Filters, *Journal of Econometrics,* **99**, 317–334.

[11] Pollock, D.S.G., (2000), Filters for Short Nonstationary Sequences, Discussion Paper 00A030 of GREQAM, University of Aix–Marseille.

[12] Wiener, N., (1950), *Extrapolation, Interpolation and Smoothing of Stationary Time Series,* MIT Technology Press and John Wiley and Sons, New York.

[13] Whittle, P., (1983), *Prediction and Regulation by Linear Least-Square Methods, Second Revised Edition,* Basil Blackwell, Oxford.

Queen Mary
University of London

This working paper has been produced by
the Department of Economics at
Queen Mary, University of London



Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096 or Fax: +44 (0)20 8983 3580
Email: j.conner@qmw.ac.uk
Website: www.econ.qmw.ac.uk/papers/wp.htm