Introduction to the Special Issue on Statistical Signal Extraction and Filtering

D.S.G. Pollock

Queen Mary, University of London, Mile End Road, London E1 4NS

Abstract

The papers of the *Special Issue on Statistical Signal Extraction and Filtering* are introduced briefly and the invitation to contribute to the next issue to be devoted to this topic is reiterated. There follows an account of the history and the current developments in the areas of Wiener–Kolmogorov and Kalman filtering, which is a leading topic of the present issue. Other topics will be treated in like manner in subsequent introductions.

The decision to devote an issue of *Computational Statistics and Data Analysis* to the topic of Statistical Signal Extraction gave rise to a call for papers which included the following statement:

> The need to extract signals and other components from time series is a requirement in many empirical sciences, including Medicine, Engineering, Economics and Climatology, to name but a few. Nowadays, a wide variety of methods are available, including Wiener–Kolmogorov Filtering, Kalman Filtering, Principal Components Analysis, and Wavelet Analysis. Given the variety of the available methods of Statistical Signal Extraction and Filtering, and given the diversity of the subject areas in which they are applied, there are plentiful opportunities for cross fertilisation and technology transfer.

The call for papers has resulted in numerous submissions that, in the majority of cases, have fallen squarely into one or other of the areas mentioned above, with only a handful that defy such neat categorisations.

The largest group of papers deals with the theory and practice of Wiener–Kolmogorov filtering and Kalman filtering. There are seven papers in this category. The paper of Broto and Ruiz applies the methodology of structural

Email address: d.s.g.pollock@qmw.ac.uk (D.S.G. Pollock).

times series modelling to data that are affected by GARCH-type heteroskedasticity. (GARCH is also the topic of another paper by Ruiz, Pascual and Romo.) The paper of Maravall is a classic study in the analysis of unobserved components via the method of canonical decompositions, and that of Mazzochi is an application of Kalman filter methodology. The remaining papers by McElroy and Sutcliffe, by Godolphin and Triantafyllopoulos, by Strickland, Forbes and Martin and by Pollock all propose methodological innovations and extensions of the basic methods of Wiener–Kolmogorov and Kalman filtering.

The paper by Fried, Bernholt and Gather concerns a class of non-linear filtering methods that aim to preserve abrupt shifts of level and monotonic trends while rejecting the sporadic and impulsive noise that may obscure the signal. It serves to remind us that the topic of time series filtering extends far beyond the range of the linear methods.

The next group of papers concerns principal components analysis. These papers are by Ombao and Ho, by Coakley, Fuertes and Smith and by AminGhafari, Cheze and Poggi. Although links can be forged between unobserved components and principal components, none of the papers here have done so; and this remains a topic for future issues.

The paper by AminGhafari, Cheze and Poggi forges links between principal component analysis and the third major topic which is that of wavelets analysis. This is also represented by the papers by Whitcher, by Amato, Antoniadis and De Feis and by Stoev, Taqqu, Park, Michailidis and Marron. The latter uses wavelets to analyse local dependence structures in the data in pursuit of a test of long range dependence and self similarity.

The remaining papers in this collection fall outside the main clusters but well within the ambit of the special issue on statistical signal extraction. The first to mention is the paper by Zeng and Garcia-Frias concerning the use of hidden Markov models in the analysis of gene expression. The paper of Strickland Forbes and Martin also uses Markov chain modelling. It applies the Kalman filter and the fixed-interval smoothing algorithms to a linear Gaussian approximation of a non-Gaussian state-space model; and thus it closes the circle. The final paper, which lies outside the boundary of the tight circle, is that of Lehtinen on signal extraction for simulated games. This is a fascinating application, which is doubtless a surprising one to many. The paper demonstrates the wide scope of the subject of this issue in a way that is most welcome.

More than enough material has arisen from the call for papers to suggest that this is a worthwhile enterprise; and we envisage a regular succession of special issues of the journal devoted to the topic of statistical signal extraction. An invitation to submit papers to the second issue has already been broadcast, and it is appropriate to reiterate it here.

The text that accompanied the original call for papers, which is reproduced above, serves equally for the second call; and we see no merit in limiting the scope. In successive issues, we shall attempt, via a policy of active solicitations, to shift the emphasis from one major area to the next; but this should not be a concern to any potential contributor who is attracted by the broad terms of the invitation.

In line with these intentions, it seems appropriate to devote the remainder of this introduction to a brief account of the history and the current developments in one of the areas covered by this issue, which is that of Wiener–Kolmogorov and Kalman filtering. This is the leading topic of the present issue. Other topics will be treated in like manner in subsequent introductions.

Wiener–Kolmogorov Filtering and Kalman Filtering

The modern theory of statistical signal extraction has its origins in the Second World War. It was formulated independently by Norbert Wiener (1941) and Andrei Nikolaevich Kolmogorov (1941). They were both considering the problem of how to target radar-assisted anti-aircraft guns on incoming enemy aircraft. Notwithstanding their successful solution of the problem, it is doubtful whether the theory achieved any practical applications at this stage. For brief biographical accounts of Kolmogorov and Wiener see Arnold (2004) and Jerison and Stroock (1995), respectively.

The two proponents of the theory established it in different ways. Whereas Kolmogorov took a time-domain approach to the problem, Wiener worked primarily in the frequency domain. However, the unification of the two approaches was soon achieved, and a modern account of the basic Wiener–Kolmogorov theory, which encompasses both approaches, has been provided by Whittle (1983).

Another important element in the modern theory of statistical signal extraction was provided by the papers of Kalman (1960a) and Kalman and Bucy (1961), which dealt with the filtering and forecasting, in discrete and continuous time, of time-varying linear stochastic systems. This approach proposes a state-space model of the system in question; and a distinguishing feature of the formulation is the manner in which the observations are related to the underlying state variables via a linear transformation, affected by additive stochastic disturbances, in which the instruments of control can be embedded.

Almost immediately, the Kalman filter found some important applications. It was used, for example, in the late 1960's in the navigation systems that enabled the Apollo spacecraft to reach the moon. An important adjunct of

the filter in this connection was provided by the work of Kalman (1960b) and Kalman, Ho and Narendra (1961) on the subject of the observability and controllability of linear state-space systems. According to an assertion of Kalman himself, which many have heard, his papers on the subjects of filtering and control are the most cited scientific publications of all time.

The formulation of Kalman and Bucy was sufficiently general to subsume the theory of Wiener and Kolmogorov. This had been clearly apparent from the beginning. However, it has taken a while for these two approaches to statistical signal extraction to be unified, and each continues to have its separate adherents. Nevertheless, it is now widely recognised that they stand on a common ground. There have been numerous offerings that have demonstrated the connection. Econometricians commonly cite the paper of Burridge and Wallis (1988). The latter have shown how, in the case of a stationary ARMA process, the Kalman filter converges asymptotically to the backward-looking Wiener–Kolmogorov filter applied to a semi-infinite data sequence.

One has to be careful here in the use of terminology. Some econometricians insist in describing the theory of Kalman filtering, in so far as it applies to the extraction of signals from data generated by time-invariant statistical processes, as Wiener–Kolmogorov theory. On this understanding, the hallmark of the theory is minimum-mean-square-error estimation, and there is no restriction on the nature of the sample. It may be finite, semi-infinite or doubly-infinite.

Others propose that Wiener–Kolmogorov theory comprises only the original formulations, which were in respect of semi-infinite and doubly-infinite samples. However, it is worth noting that the generalisation of the theory to the case of finite samples occurred early in its history, as is evident from the paper of Zadeh and Ragazzini (1950). That there are still some major differences in outlook is evident from a comparison of some of the papers in this issue.

In econometrics, the business of statistical signal extraction had been pursued in two quite distinct ways. On the one hand, there are the heuristic filtering methods that have been developed in the central statistical agencies of several of countries and which continue to be used predominantly in the preparation of official statistics.

These methods, which are aimed, primarily, at removing the seasonal fluctuations from economic series and for estimating their trends, do not make much reference to the Wiener–Kolmogorov theory. Instead, they rely on time-honoured filtering recipes, such as the Henderson (1916) filters. (See Kenny and Durbin (1982) for an analysis of the Henderson filters.) These filters have been investigated and justified by determining their frequency response functions. They are barely represented in this issue.

The model-based methods that are prevalent in this issue are based on Wiener–

Kolmogorov and Kalman filter theory; but, within this framework, some further distinctions arise. In the first place, there is a question of whether the models of the processes that generate the data are held to be realistic representations or whether they are regarded merely as heuristic devices that are used to assist in the design of the filters.

Some econometricians take the view that, for the methods to be successful, the models must be realistic and that, for this purpose, they need to be estimated from the data. Others, who dissent from this opinion, believe that such estimated models are liable, sometimes, to be inadequate. Therefore, they prefer, on occasion if not always, to adopt an approach based on pre-specified heuristic models that are chosen to ensure that the resulting filters have the desired characteristics.

In the empirically determined model-based procedures, the unobserved components of a time series are typically modelled by ARMA and ARIMA processes of relatively low orders. The assumption that the components are driven by mutually uncorrelated white-noise processes implies that the sum of their autocovariance generating functions must equal the autocovariance generating function of the aggregate process.

In the structural time series modelling approach, which is implemented in the STAMP program of Koopman *et al.* (1995) that reflects the methodology of Harvey (1989), the component processes are modelled individually and they are combined to produce a model of the aggregate process. The CAPTAIN MatLab Toolbox of Young *et al.* (2004), which can be used in estimating unobserved components, also adopts the structural approach.

In the alternative canonical-decompositions approach, which is implemented in the SEATS–TRAMO program of Caporello and Maravall (2004), the aggregate process is modelled at the outset before being decomposed into its component parts. The autocovariance generating functions of the component processes are obtained from that of the aggregate processes via a partial fraction decomposition.

If the spectral density function of a linear stochastic process is nowhere zero in the available frequency range, then the process can be decomposed into two parts, one of which is white noise. According to the canonical principle, the substantive components of a model should be purged of their white-noise elements, which should be attributed to a separate and independent irregular component.

In the decompositions approach of SEATS–TRAMO, it is necessary to adopt this principle in order to identify the individual components. The canonical principle might also be adopted in the structural times series modelling approach, but it is unnecessary for the purpose of achieving identification and

it is commonly disregarded. (An exposition of the canonical-decompositions approach is given in the article of Maravall in this issue.)

There is a tendency for the trends that are estimated via an empirical model-based approach to contain high-frequency elements of the data. The consequence is that their profiles are rough and jagged when we expect them to be smooth. This tendency will be exacerbated by the failure to represent the high-frequency components adequately within the model. An example is provided by the model that underlies the filter of Hodrick and Prescott (1997). This uses an integrated random walk as the model of the trend component and a white-noise disturbance term to represent the remaining data components.

The only free parameter of the model is the ratio that is formed from the variance of the white noise that powers the trend component and the variance of the white-noise disturbance. This ratio is known as the smoothing parameter. When the model is fitted by maximum likelihood, it is commonly found that the value of the estimated smoothing parameter is far too low to deliver a smooth trend. For this reason, the smoothing parameter is commonly determined by a rule of thumb.

Apart from ensuring that the components of the data are properly represented in the model, there are other recourses that may be adopted. In the first place, a resort to the canonical principle will ensure that any white noise that resides within the estimated trend is removed and assigned to the irregular component. The problem can also be alleviated by modifying the criterion of estimation.

One recourse, which is employed by the CAPTAIN Toolbox, is to adopt the criterion of minimising the sum of squares of the h-steps-ahead forecast errors for fitting the model in place of the usual maximum-likelihood criterion, which is based on the one-step-ahead forecast errors. Proietti (2004), has recently shown that this is tantamount to pre-processing the data via a lowpass filter before proceeding with a maximum-likelihood or a least-squares estimation. The effect is to concentrate attention on the low-frequency end of the spectrum.

Another recourse, which is also provided by the Toolbox, is to adopt a frequency-domain criterion that is designed to ensure that the parametric spectrum of the model adheres as closely as possible to the empirical periodogram or to the parametric spectrum of a high-order AR model that has been fitted to the data.

Part of the problem that can affect the empirical model-based procedures lies in the nature of the low-order ARIMA processes that are used to represent the data components. These have widely dispersed spectra that extend across the entire frequency range. The filters that are derived from such models are

liable to transmit a wide range of frequencies. In order to represent processes with localised spectra, ARIMA models of much higher orders may be required, which have impulse responses that are widely dispersed in the time domain.

However, there are difficulties in estimating high-order processes from limited data sequences. Therefore, when we wish to extract components with localised spectra, we are liable to resort to pre-specified filters of high orders that embody only one or two free parameters that govern their design. Typically, in highpass and lowpass filters, these parameters determine the nominal cut-off frequency of the pass band and the rate of transition between the pass band and the stop band.

Amongst econometricians, the digital Butterworth filter, which is familiar to electrical engineers, has been making some headway recently—see Pollock (2000). The filter is derivable from an heuristic model on Wiener–Kolmogorov principles. It has a nominal cut-off point, marking the mid point in the transition from the pass band to the stop band, that can be specified at will, and the rate of the transition can be increased by increasing the order of the filter.

In many cases, it will found, from an inspection of the periodogram of the detrended data, that the spectral structures of the various components are separated by virtual dead spaces where the periodogram ordinates show no significant power. In that case, it may be appropriate to use a method that can be applied directly to the ordinates of the Fourier transform of the detrended data. Such a method is presented in the paper of Pollock (this issue), where it is related to the Wiener–Kolmogorov methods that operate in the time domain.

If the data sequence were infinitely lengthy, then the Wiener–Kolmogorov filter for extracting a particular component would be formed from the ratio of the component's autocovariance generating function and that of the aggregate process. Such filters have infinite impulse response functions, and, unless they are implemented via bidirectional rational feedback processes, they are liable to generate doubly-infinite sequences of filter coefficients. This seems, at first, to make it difficult to apply them to data sequences of limited duration, such as are typical in econometric applications.

The problem of finite data sequences is properly overcome by adopting a finite-sample version of the Wiener–Kolmogorov filter. As the paper of Pollock (this issue) shows, this is readily accessible. Moreover, it is also provided by the Kalman filter in combination with a fixed-interval smoothing algorithm.

These two finite-sample approaches generate the same centrosymmetric filter matrix. Its effect is that, at the start of the process, the filter looks forwards in time to the end of the sample. In the middle of the process, it looks symmetrically towards both ends of the sample; and, at the end of the process,

it looks backward in time to the start of the sample. The sequence of filter coefficients that is used at the end of the sample is the mirror image of the sequence that is used at the start.

There are various ways of adapting the finite sample Wiener–Kolmogorov filters to deal with nonstationary data sequences. One way of doing so is to rely upon the Kalman filter, of which the transitions equation can readily accommodate nonstationary dynamics. However, it has taken a while to settle the matter of how best to provide the requisite start-up values in cases where the data are modelled by nonstationary processes. A definitive solution appears to have been reached by Ansley and Kohn (1985), but it has required some further interpretation and exposition by others before being widely adopted. See, for example, Koopman (1997) and Pollock (2003).

An alternative method of dealing with trended data is to reduce it to stationarity by differencing, or by some other means, before applying the signal extraction procedures. The components that have been extracted from the differenced data can be re-inflated, via the summation operator, which is the inverse of the difference operator, to produce the estimates of the original components. The process of re-inflation also requires some initial conditions; and Pollock (this issue) shows how these may be obtained.

It is notable that much of the exposition of Ansley and Kohn (1985) is devoted to a demonstration that the method which they propose for adapting the Kalman filter to nonstationary data entails a transformations approach in which the differencing operator is used to eliminate the dependence of the likelihood function upon unknown initial conditions with a diffuse or improper statistical distribution.

It has been claimed that the differencing approach destroys much of the low-frequency information. In practice, however, the signal extraction processing is applied to the high-frequency components; and the low-frequency trend component, of which the contents would be destroyed by differencing, is found by subtracting from the data the sum of the other components.

It has never been easy to understand how the Kalman filter fulfils the remit of a finite-sample Wiener–Kolmogorov theory. That it does do so is evident only from the fact that both approaches satisfy the same criteria. Indeed, it has been remarked by one authority that "One could have ignored the Wiener–Kolmogorov filtering literature and still one could have obtained the Kalman filter/fixed-interval smoothing solution to the signal extraction problem." Two of the papers of this issue, namely those of Pollock and of McElroy and Sutcliffe, should help in reaching an understanding of the relationship between the two approaches.

8

References

[1] Ansley, C.F., and Kohn, R., 1985. Estimation, filtering and smoothing in state space models with incompletely specified initial conditions. The Annals of Statistics, 13, 1286–1316.

[2] Arnold, V.V., 2004. A.N. Kolmogorov and natural science. Russian Math. Surveys, 59, 27–46.

[3] Burridge, P., and Wallis, K.F., 1988. Prediction theory for autoregressive–moving average processes. Econometric Reviews, 7, 65–95.

[4] Caporello, G., and A. Maravall, 2004. Program TSW: Revised Reference Manual, Working Paper 0408, Servicio de Estudios, Banco de España.

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

[6] Henderson, R., 1916. Note on graduation by adjusted averages. Transactions of the Actuarial Society of America, 17, 43–48.

[7] Hodrick R.J., and Prescott, E.C., 1997. Postwar U.S. business bycles: an empirical investigation. Journal of Money, Credit and Banking, 29, 1–16.

[8] Jerison, D., and Stroock, D., 1995. Norbert Wiener. Notices of the AMA, 42, 430–438.

[9] Kalman, R.E., 1960a. A new approach to linear filtering and prediction problems. Transactions of the ASME Journal of Basic Engineering, Series D, 82, 35–45.

[10] Kalman, R.E., 1960b. On the general theory of control systems. Proceedings of the First International Congress of Automatic Control, Moscow, USSR.

[11] Kalman, R.E., and Bucy, R.S., 1961. New results in linear filtering and prediction theory. Transactions of the ASME Journal of Basic Engineering, Series D, 83, 95–107.

[12] Kalman, R.E., Ho, Y.C., and Narendra, K.S., 1961. Controllability of linear dynamical systems. Contributions to Differential Equations. John Wiley, New York.

[13] Kenny, P.B., and Durbin, J., 1982. Local trend estimation and seasonal adjustment of economic and social time series. Journal of the Royal Statistical Society, Series A (General), 145, 1–41.

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

[15] Koopman, S.J., 1997. Exact initial Kalman filtering and smoothing for nonstationary time series models. Journal of the American Statistical Association, 92, 1630–1638.

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

[17] Pollock, D.S.G., 2000. Trend estimation and de-trending via rational square wave filters. Journal of Econometrics, 99, 317–334.

[18] Pollock, D.S.G., 2003. Recursive estimation in econometrics. Computational Statistics and Data Analysis, 44, 37–75.

[19] Proietti, T., 2004. Forecasting and signal extraction with misspecified models. Forthcoming in Journal of Forecasting.

[20] Wiener, N., 1941. Extrapolation, Interpolation and Smoothing of Stationary Time Series. Report on the Services Reasearch Project DIC-6037. Published in book form in 1949 by MIT Technology Press and John Wiley and Sons, New York.

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

[22] Young, P.C., Taylor , C.J., Tych, W., Pedregal, D.J., and McKenna, C.J., 2004. The Captain Toolbox. Centre for Research on Environmental Systems and Statistics, Lancaster University. (www.es.lancs.ac.uk/cres/captain).

[23] Zadeh, A., and J.R. Ragazzini, 1950, An Extension of Wiener's theory of prediction. Journal of Applied Physics, 21, 645–655.

Referees

The editors would like to thank the following persons who have acted as referees for the special issue on *Statistical Signal Extraction and Filtering.*

Umberto Amato
Kit Baum
Jermie Bigot
Fabio Busetti
Roberto Casarin
Nathalie Cheze
Daniela De Canditis
Tim Downie
Gabriele Fiorentini
Roland Fried
Ursula Garczarek
Ed Godolphin
Gregory Gonye
Andrew Harvey
Lasse Holmstrom
Alessandra Iacobucci
Regina Kaiser
Vladimir Katkovnik
Chang-Jin Kim
Joseph Lardies
W.K. Li
Jiebo Luo
Domenico Marinucci
Mario Mazzocchi
Nikolaos Mitianoudis
Vittorio Murino
Guy Nason
Francisco Ocaña
Jeffrey Perloff
Kris Popat
James Ramsay
Jörg Renner
Elmar Rothert
Esther Ruiz
Steve Satchell
Philipp Sibbertsen
Thomas Trimbur
Dick van Dijk
Mike West
Peter Young

Bhavik Bakshi
Luc Bauwens
Bill Bolstad
Jean-François Cardoso
Jean-Paul Chavas
Jean-François CoeurJolly
Ivo Dinov
Rebecca Driver
Catherine Forbes
Jean Jacques Fuchs
Richard Gerlach
Amos Golan
Raphael Gottardo
John Haywood
Soosung Hwang
Emma Inglesias
Madhu Kalimipalli
Hiroyuki Kawakatsu
Siem Jan Koopman
Matthieu Lemoine
Jia Li
Danilo Mandic
Adam Marsh
Bill McCausland
Bill Moran
Fionn Murtagh
Michael Neumann
Prakash Patil
Benoit Perron
Don Poskitt
Morten Ravn
Thomas Ribarits
Alex Rubinov
Gerhard Rünstler
Christoph Schleicher
David Stern
Klaus Troitzsch
Marina Vannucci
Chisun Won

Kelvin Balcombe
Jan Beran
Alain Boudou
Hervé Cardot
Zhao-Guo Chen
Rosaria Conte
Howard Doran
Guoliang Fan
Mario Forni
Antonio Garcia Ferrer
Liudas Giraitis
Victor Gomez
Robert Gregson
Mark Henry
Rob Hyndman
Andy Irving
Juha Karvanen
Mitsuru Kawamoto
Hans-Joachim Kuempel
Steve Leybourne
Asger Lunde
Jonathan Marchini
Kent Matthews
Tucker McElroy
David Mount
Federico Nardari
Jukka Nyblom
Joe Pearlman
Christophe Planas
Tommaso Proietti
Lucrezia Reichlin
Yves Romain
Havard Rue
Lucio Sarno
Sergei Shiskin
Kostas Triantafyllopoulos
Dimitri Van De Ville
David Veredas
David Woodruff