Ann Inst Stat Math (2013) 65:439–456
DOI 10.1007/s10463-012-0373-x

Forecasting continuous-time processes with applications to signal extraction

Tucker S. McElroy

Received: 23 March 2011 / Revised: 20 June 2012 / Published online: 23 September 2012

Abstract The paper derives forecasting and signal extraction estimates for continuous time processes. We present explicit formulas for filters and filter kernels that yield minimum mean square error estimates of future values of the process or an unobserved component, based on a continuum of values in the semi-infinite past. The class of processes considered are cumulations of moving average processes, which includes the CARIMA class. Explicit examples are calculated, and some discussion of applications to signal extraction is provided. We also provide an explicit algorithm for spectral factorization of continuous-time moving averages.

Keywords CARIMA · Signal extraction · Stochastic process

1 Introduction

Data that arise from a continuous stream in time are prevalent in many engineering and industrial applications, such as the regulation of a thermostat or the production of chemicals. There may be some underlying physical process that is observed at any set of sampling times, which can be made arbitrarily frequent. Analysts refer to such data as being "continuously observed," though of course this does not mean that an actual continuum of information is stored—such an uncountable quantity would be

This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the author and not necessarily those of the U.S. Census Bureau.

T. S. McElroy (✉)
Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100, USA
e-mail: tucker.s.mcelroy@census.gov

impossible to retain. Rather, "continuous observation" means that the actual sampling rate is quite high relative to the interesting features of the process, and sampling can be effected whenever desired. Examples include EKG readings (heart and brain monitoring), process control used in industrial applications, and financial records of price movements. See Karatzas and Shreve (1998), Barndorff-Nielsen and Shepherd (2001), Kilian (2005), Andrei (2006), and Astrom and Murray (2008) for background.

In these situations it is very natural to model the physical process via a continuous-time stochastic process, such as a continuous-time autoregressive moving average (CARMA) process; see Brockwell (2000, 2001, 2004) and Brockwell et al. (2012) for a recent treatment, although this important class of processes has been in use for many decades. Cumulations (or integrations) of such processes may also be considered, in order to represent nonstationary effects. Such models can be fitted using a finite number of observations by writing down the corresponding Gaussian likelihood; see Jones (1981) or Bergstrom (1988, 1990) for further details.

Two of the chief statistical problems of interest for any observed stochastic process are forecasting and signal extraction. This paper presents forecasting and concurrent signal extraction filters constructed in continuous-time, such that the resulting mean squared error (MSE) is minimal among all linear estimators. For the forecasting results, it is important to assume the process is causal, and the optimal forecasting filter is typically represented as the convolution of a tempered distribution with the data process. When combined with a smooth signal extraction filter, these forecast tempered distributions can result in tractable concurrent filters for estimating trends and other movements of interest. In practice, one might proceed to discretize the signal extraction estimates in accordance with the observed data, which is typically sampled regularly at a fixed frequency from the data process.

We begin by setting out the forecasting results for stationary causal processes (including, but not limited to, the CARMA processes) in Sect. 2, also introducing some basic concepts about continuous time concurrent filters. Section 3 extends these results to integrations of causal moving averages, so that nonstationarity can be addressed. Then Sect. 4 treats concurrent signal extraction by combining known results on symmetric signal extraction (McElroy and Trimbur 2006) with the preceding forecasting results. In some applications involving unobserved components, it is necessary to perform spectral factorization of spectra. This concept is explained in Sect. 4, and an algorithm to accomplish spectral factorization is described in the Appendix, along with the proofs of all theorems. An illustration of concurrent signal extraction is also developed in Sect. 4, with some discussion of how the method can be used on observed data.

2 Forecasting causal filtered noise processes

Define a causal filtered noise (CFN) process Y via $Y = g * \epsilon$ with g supported on $\mathbb{R}^+ = [0, \infty)$ and ϵ continuous time white noise (Priestley 1981, pp. 156–158), which may or may not be Gaussian. Here $*$ denotes convolution, so that at each time t we have $Y(t) = \int_0^\infty g(x)\epsilon(t-x)\, \mathrm{d}x$. This process consists of noise $\{\epsilon(t)\}$ filtered by g; since $Y(t)$ only depends upon present and past values of the noise process, it is called causal. These processes form the core class for this paper, and are fairly broad. They are mean zero and covariance stationary, with autocovariance function

$$R(h) = \mathbb{E}[Y(t)Y(t+h)] = \sigma^2 \, (g * \overline{g})(h) \tag{1}$$

for any $h \in \mathbb{R}$, where σ is the scale of the continuous time white noise (WN (σ^2) hereafter), and $\overline{g}(x) = g(-x)$. A simple example of a CFN process is provided by $g(x) = 1_{[0,c]}(x)$ for some $c > 0$, which produces a c-dependent covariance structure on the resulting process Y.

A very useful class of examples is afforded by the CARMA processes. Consider a CARMA process $\{Y(t)\}$ as described in Brockwell (2004), repeated here for easy reference. Let $\{\epsilon(t)\}$ be WN (σ^2), and let D be the mean-square differentiation operator defined in (Priestley 1981, p. 164)—also see McElroy and Trimbur (2006). Consider relatively prime polynomials $a(z) = z^p + a_1 z^{p-1} + \cdots + a_p$ and $b(z) = b_0 + b_1 z + \cdots + b_q z^q$ such that

$$a(D)Y(t) = b(D)\epsilon(t) \tag{2}$$

where the distinct roots $\lambda_1, \ldots, \lambda_P$ of $a(z)$ have negative real part and multiplicities $k_1, \ldots, k_P$, so that $a(z) = \Pi_{r=1}^{P} (z - \lambda_r)^{k_r}$. The case of simple poles was expounded in Brockwell (2004) and extended to repeated roots in Tsai and Chan (2009); also see Brockwell and Lindner (2009).

In the definition of the CARMA, it is vital that $q < p$. For short the process is referred to as a CARMA(p,q). The condition that the AR roots have negative real part ensures a stationary process (Brockwell 2004). Then there exists a causal kernel g such that $Y = g * \epsilon$, given by

$$g(x) = \sum_{r=1}^{P} \frac{1}{(k_r - 1)!} \cdot \frac{\partial^{k_r - 1}}{\partial y^{k_r - 1}} \left\{ \frac{b(y) \, \exp\{yx\}}{\Pi_{\ell \neq r} (y - \lambda_\ell)^{k_\ell}} \right\} |_{y=\lambda_r} \; 1_{[0,\infty)}(x). \tag{3}$$

(By convention $0! = 1$.) Hence the autocovariance function consists of linear combinations of exponentials and sinusoids multiplied by polynomials.

Returning to the general CFN process, the h-step ahead forecasting problem is to determine the minimum mean squared error (MMSE) forecast of $Y(t+h)$ given data $\mathbf{Y} = \{Y(s)\}_{s \leq t}$. Here $h > 0$, and the forecast is denoted $\widehat{Y(t+h)}$. A limitation here is that a semi-infinite past is assumed; the forecasting problem utilizing a finite-sample is not treated here. (When the forecasting kernel decays to zero rapidly enough, the distant past has no impact, and we can truncate the filter to a finite past without any change to the estimates.) The MMSE forecast is a linear function of the data if $\{Y(t)\}$ is a Gaussian process. We focus on this case, since linear functions are easy to work with in practice. Relaxing Gaussianity, one may instead formulate the problem as seeking the filter with smallest MSE among all linear estimators in $\mathbf{Y}$, i.e., among all causal filters. Only under the Gaussian assumption do we know that our MMSE estimator is optimal among *all* estimators, linear and non-linear. Therefore we have

$$\widehat{Y(t+h)} = (\psi * Y)(t),$$

where ψ is supported on $\mathbb{R}^+$ (so that we only utilize present and past data). A solution to this problem is given in Theorem 1 below. Although ψ depends on h, this will be suppressed in the notation.

First, we introduce some helpful notation. The lag operator L is defined formally via $L = e^{-D}$, and acts on a process by shifting it back one time unit. Thus $L^s Y(t) = Y(t-s)$ for all $s, t \in \mathbb{R}$. Filtering of continuous-time processes is equivalent to convolution of the input data with a so-called kernel function. Formally this is described as the action of a filter operator defined as follows (see McElroy and Trimbur (2006) for more discussion). If θ is a kernel function and X is our input process, then our output process at time t has value

$$(\theta * X)(t) = \int \theta(u) X(t-u)\,\mathrm{d}u = \int \theta(u) L^u X(t)\,\mathrm{d}u = \int \theta(u) L^u\,\mathrm{d}u\; X(t).$$

Hence we define the filter operator (or filter for short) associated with the kernel θ via $\Theta(L) = \int \theta(u) L^u\,\mathrm{d}u$, which acts on processes via convolution of the associated kernel. In general we use lower case for a kernel, and upper case for the corresponding filter; in particular, $G(L) = \int g(u) L^u\,\mathrm{d}u$. The parallels to the discrete-time case are obvious and deliberate. Note that the filter's frequency response function (frf) is $\Theta(e^{-i\lambda}) = \mathcal{F}[\theta](\lambda)$, the fourier transform (FT) of θ. This also establishes our normalization conventions about the FT. The spectral density of a stationary stochastic process is the FT of the autocovariance function, and therefore is proportional to the squared magnitude of the FT of g. Note that in the CARMA case, $G(e^{-i\lambda}) = b(i\lambda)/a(i\lambda)$, and g in (3) equals $\mathcal{F}^{-1}[b(i\cdot)/a(i\cdot)]$.

This discussion can be generalized somewhat when the kernel θ is not well-defined as a function, even when $\Theta(L)$ is well-defined. For example, we may have $\Theta(L) = L^0$ (the identity filter), which has kernel θ given by the Dirac delta function at zero (Folland 1995); this is actually a tempered distribution, not an actual function. It is a reality of continuous-time filtering that many filters of interest only have kernels that exist in the sense of tempered distributions, not being proper functions; in this case it is not useful to write down the kernels.

So the forecasting problem is therefore to determine $\Psi(L)$ (this depends on h, which is fixed throughout the discussion) such that $\Psi(L)Y(t)$ is a MMSE linear estimator of $Y(t+h)$, for any $t \in \mathbb{R}$. We present a general expression for the frf for CFN processes, and a more particular expression for the CARMA case.

Theorem 1 *The MMSE forecast filter $\Psi(L)$ for a Gaussian CFN process $Y = g * \epsilon$ has frf*

$$\Psi(e^{-i\lambda}) = \frac{\int_0^\infty g(x+h) e^{-i\lambda x}\,\mathrm{d}x}{\int_0^\infty g(x) e^{-i\lambda x}\,\mathrm{d}x}.$$

Furthermore, suppose the CFN is a CARMA process (2), *and let*

$$c^{(r)}(z) = \frac{\partial^{k_r - 1}}{\partial y^{k_r - 1}} \left\{ \frac{b(y)\ \exp\{yh\}}{(z-y)\Pi_{\ell \neq r}(y - \lambda_\ell)^{k_\ell}} \right\} |_{y=\lambda_r} \tag{4}$$

be defined for any complex $z \neq \lambda_r$, for $r = 1, \ldots, P$. Then the filter frf is given by

$$\Psi(e^{-i\lambda}) = \sum_{r=1}^{P} \frac{c^{(r)}(i\lambda)}{(k_r - 1)!} \cdot \frac{a(i\lambda)}{b(i\lambda)}.$$

The spectral density of the error process is $|\int_0^h g(u)e^{-i\lambda u}\,\mathrm{d}u|^2 \sigma^2$, *with minimal MSE of* $(2\pi)^{-1} \int_0^h g^2(x)\,\mathrm{d}x\,\sigma^2$.

Remark 1 If the domains of integration in the first formula for $\Psi(e^{-i\lambda})$ were over all $\mathbb{R}$ instead of $\mathbb{R}^+$, the frf would trivially reduce to $e^{i\lambda h}$ corresponding to L^{-h}. This "perfect" forecast function corresponds to having all data available.

Remark 2 When the AR and MA roots have multiplicity one, then the frf can be written as

$$\Psi(e^{-i\lambda}) = \sum_{r=1}^{p} \frac{b(\lambda_r)\exp\{\lambda_r h\}}{(i\lambda - \lambda_r)\,\Pi_{\ell \neq r}(\lambda_r - \lambda_\ell)}\ \frac{a(i\lambda)}{b(i\lambda)}.$$

Because $\dot{a}(\lambda_r) = \Pi_{\ell \neq r}(\lambda_r - \lambda_\ell)$, we can formally express the filter as

$$\Psi(L) = \sum_{r=1}^{p} \frac{b(\lambda_r)\exp\{\lambda_r h\}}{\dot{a}(\lambda_r)}\ \frac{\Pi_{\ell \neq r}(D - \lambda_\ell)}{b(D)}.$$

This typically has a tempered distribution for its kernel, since the numerator has order $p-1$ in D, and the denominator has order q. The forecasted process $\widehat{Y(t+h)}$ (as a function of t) is a CARMA(p,p−1). In particular,

$$a(D)\widehat{Y(t+h)} = \Psi(L)b(D)\epsilon(t) = \left(\sum_{r=1}^{p} \frac{b(\lambda_r)\exp\{\lambda_r h\}}{(D - \lambda_r)\,\dot{a}(\lambda_r)} \right) \epsilon(t).$$

Real-data applications are discussed in Sect. 4. Typically, the forecasting filters have non-integrable frfs because they involve the "whitening" transformation $b(D)/a(D)$, which has a kernel only in the sense of tempered distributions. When the forecasting filters are combined with suitably smooth filters, such as trend filters, then practical applications are possible. Below we demonstrate a few theoretical examples of forecast filters.

Example 1 **CAR(1)** Let $a(z) = z + a_1$ with $\lambda_1 = -a_1$ (so $a_1 > 0$). Then $\Psi(e^{-i\lambda}) = e^{-a_1 h}$, i.e., a constant multiplication of the current observation at time present. In this case ψ is equal to $e^{-a_1 h}$ times the Dirac delta function at zero. Compare this result with the forecast function of the discrete AR(1) with polynomial $1-\phi B$, namely ϕ^h. The CAR(1) forecast MSE is $(1-e^{-2a_1 h})/(2a_1)$ times the innovation variance, which clearly increases in h (to a limiting value of $1/(2a_1)$) and tends to zero as $h \to 0$, which is intuitive.

Example 2 **CAR(2)** Let $a(z) = z^2 + a_1 z + a_2$ with roots λ_1, λ_2. So $a_1 = -(\lambda_1 + \lambda_2)$ and $a_2 = \lambda_1 \lambda_2$. Then the filter is

$$\Psi(L) = \frac{e^{\lambda_1 h}(D-\lambda_2)}{\lambda_1 - \lambda_2} + \frac{e^{\lambda_2 h}(D-\lambda_1)}{\lambda_2 - \lambda_1}$$

when $\lambda \neq \lambda_2$. Then the causal filter of the process is $g(u) = e^{\lambda_1 u}/(2\lambda_1 + a_1) + e^{\lambda_2 u}/(2\lambda_2 + a_1)$, which is supported on $\mathbb{R}^+$, so that the MSE (with $\sigma^2 = 1$) is

$$\frac{e^{2\lambda_1 h} - 1}{2\lambda_1(2\lambda_1 + a_1)^2} + \frac{2(e^{(\lambda_1+\lambda_2)h} - 1)}{(\lambda_1+\lambda_2)(2\lambda_1 + a_1)(2\lambda_2 + a_1)} + \frac{e^{2\lambda_2 h} - 1}{2\lambda_2(2\lambda_2 + a_1)^2}.$$

When $\lambda_1 = \lambda_2$, then $\Psi(L) = \exp\{\lambda_1 h\} \cdot \{1 + h(D - \lambda_1)\}$.

3 Forecasting integrated processes

We now consider a more general process: the integrated causal filtered noise (ICFN). This can be written as

$$D^d Y(t) = G(L)\epsilon(t) \tag{5}$$

where $G(L)$ is a causal filter, and $\{\epsilon(t)\}$ is WN (σ^2). Note that the presence of D^d on the left hand side indicates a non-stationary process Y. When it exists, the kernel g of $G(L) = \int g(x) L^x \mathrm{d}x$ is assumed to be an integrable function supported on $\mathbb{R}^+$—as in Brockwell (2004) and the CFN case above. But we also allow for $G(L)$ to be a constant (without loss of generality, equal to unity) when $d \geq 1$, in which case g does not exist except as a tempered distribution, being equal to the Dirac delta function. When $G(L)$ is a rational function in D we obtain a CARMA process for $\{D^d Y(t)\}$. Now as discussed in McElroy and Trimbur (2011), the process $\{Y(t)\}$ can be represented in terms of the differentiated process $W(t) = D^d Y(t)$ and certain initial conditions. We have

$$Y(t) = \sum_{j=0}^{d-1} D^j Y(0) \frac{t^j}{j!} + [I^d W](t), \tag{6}$$

where the operator I^d is defined recursively via $[I^d W](t) = \int_0^t [I^{d-1} W](s)\, \mathrm{d}s$ with $[I^0 W] = W$ (the identity operator), and is explicitly given as $\int_0^t W(s)(t-s)^{d-1} \mathrm{d}s/(d-1)! = W * \chi_d(t)$ for $\chi_d(t) = t^{d-1} 1_{\mathbb{R}^+}(t)/(d-1)!$. This representation holds for $t > 0$, but can be extended to $t < 0$ by flipping the bounds of integration.

A principal example of an ICFN process is a CARIMA. Here the dth derivative of $Y(t)$ is equal to the CARMA process of the previous section. That is,

$$a(D)D^d Y(t) = b(D)\epsilon(t)$$

using the same notation as Sect. 2 (so $p > q$). For short this is referred to as a CARIMA (p,d,q) process.

In forecasting problems it is common practice to assume that initial conditions are uncorrelated with disturbances; see Bell (1984) for discussion of the discrete-time case and related issues of signal extraction. In our context, this amounts to the following:

Assumption A $\{D^j Y(0)\}_{j=0}^{d-1}$ are uncorrelated with $\{W(s)\}_{s\in\mathbb{R}}$.

We again consider the forecasting problem, seeking a forecast filter $\Psi(L)$ as in Sect. 2. As earlier, we assume the Gaussian distribution to get MMSE forecasts; relaxing Gaussianity, the filters can be interpreted as having MMSE among all estimators linear in $\mathbf{Y}$. We first treat the case of a ICFN, where $\{W(t)\}$ may be WN (i.e., where $G(L) = 1$ and g is Dirac) or is a CFN of the type discussed in Sect. 2. We also provide some more explicit expressions when the process is a CARIMA. Because of the integration of the processes, the forecast filters will typically have non-integrable frfs, and their kernels can only be viewed as tempered distributions.

Theorem 2 *Let Y be a Gaussian ICFN process (5) with $G(L)$ causal, such that the kernel g is either Dirac at zero or is a function with domain $\mathbb{R}^+$. Let $q^{(j)}(x) = [I^{d-j} g](x)$ for $0 \leq j < d$ and any $x \geq 0$. Under Assumption A, the MMSE forecast filter $\Psi(L)$ has frf*

$$\Psi(e^{-i\lambda}) = \frac{\int_0^\infty g(x+h)e^{-i\lambda x}\,\mathrm{d}x + \sum_{j=0}^{d-1} q^{(j)}(h)(i\lambda)^{d-1-j}}{\int_0^\infty g(x)e^{-i\lambda x}\,\mathrm{d}x}.$$

Furthermore, suppose the CFN $\{W(t)\}$ is a CARMA process (2), and let $c^{(r)}(z)$ of (4) of Theorem 1 be defined. Then the filter frf is given by

$$q^{(j)}(h) = \sum_{r=1}^{P} \frac{1}{(k_r-1)!} \frac{\partial^{k_r-1}}{\partial y^{k_r-1}} \left\{ \frac{b(y)\,[I^{d-j} e^{y\cdot}](h)}{\Pi_{\ell\neq r}(y-\lambda_\ell)^{k_\ell}} \right\} |_{y=\lambda_r}$$
$$\Psi(e^{-i\lambda}) = \frac{a(i\lambda)}{b(i\lambda)} \left(\sum_{r=1}^{P} \frac{c^{(r)}(i\lambda)}{(k_r-1)!} + \sum_{j=0}^{d-1} q^{(j)}(h)(i\lambda)^{d-1-j} \right).$$

The spectral density of the error process is $|\int_0^h q(u)e^{-i\lambda u}\,\mathrm{d}u|^2 \sigma^2$, with minimal MSE of $(2\pi)^{-1}\int_0^h q^2(x)\,\mathrm{d}x\,\sigma^2$.

Remark 3 This is clearly a direct generalization of Theorem 1—essentially by setting $d = 0$ and collapsing empty summation formulas, the former result is recovered. For ease of reading we present the results separately. There are similarities to the discrete-time forecasting formulas, which are discussed for a quite general difference-stationary process in McElroy and Findley (2010).

Remark 4 For a CARIMA process, one needs to compute $[I^{d-j} e^{y\cdot}]$, which using induction is $[I^{d-j} e^{y\cdot}](h) = y^{j-d}[e^{yh} - \sum_{k=0}^{d-1-j} (yh)^k/k!]$.

Example 3 **CARIMA(2,1,1)** Let $a(z) = z^2 + a_1 z + a_2 = (z - \lambda_1)(z - \lambda_2)$ and $b(z) = b_0 + b_1 z$ with $\beta_1 = -b_0/b_1$. When the roots are distinct (they must be conjugate) we obtain

$$
\begin{aligned}
q(h) &= \frac{b(\lambda_1)(e^{\lambda_1 h} - 1)}{\lambda_1(\lambda_1 - \lambda_2)} - \frac{b(\lambda_2)(e^{\lambda_2 h} - 1)}{\lambda_2(\lambda_1 - \lambda_2)} \\
c^{(1)}(z) &= \frac{b(\lambda_1)e^{\lambda_1 h}}{(z - \lambda_1)(\lambda_1 - \lambda_2)} \\
c^{(2)}(z) &= -\frac{b(\lambda_2)e^{\lambda_2 h}}{(z - \lambda_2)(\lambda_1 - \lambda_2)} \\
\Psi(e^{-i\lambda}) &= \frac{a(i\lambda)}{b(i\lambda)} \left(c^{(1)}(i\lambda) + c^{(2)}(i\lambda) + q(h)\right).
\end{aligned}
$$

When there is a repeated root, say $\lambda_1 = \lambda_2$, we have

$$
\begin{aligned}
q(h) &= \frac{\dot{b}(\lambda_1)(e^{\lambda_1 h} - 1)}{\lambda_1} - \frac{b(\lambda_1)(e^{\lambda_1 h} - 1)}{\lambda_1^2} + \frac{b(\lambda_1) h e^{\lambda_1 h}}{\lambda_1} \\
c^{(1)}(z) &= \frac{\dot{b}(\lambda_1)e^{\lambda_1 h}}{z - \lambda_1} - \frac{b(\lambda_1)e^{\lambda_1 h}}{(z - \lambda_1)^2} + \frac{b(\lambda_1) h e^{\lambda_1 h}}{z - \lambda_1} \\
\Psi(e^{-i\lambda}) &= \frac{a(i\lambda)}{b(i\lambda)} \left(c^{(1)}(i\lambda) + q(h)\right).
\end{aligned}
$$

Example 4 **Integrated Brownian motions** Here we have $\{DY(t)\}$ is Gaussian white noise, so $\{Y(t)\}$ is a Brownian Motion. Taking higher order integration produces integrated Brownian Motions, twice integrated Brownian Motions, etc. Now $[I^d \Delta_0](x) = x^{d-1}/(d-1)!$, and $\int_h^\infty g(x)\,dx = 0$ for $h > 0$, so the forecast frf is $\sum_{j=0}^{d-1} (i\lambda h)^{d-1-j}/(d-1-j)!$. This is a linear combination of derivatives, and is analogous to the forecast function for cumulated discrete noise. When $d = 2$ the filter can be expressed $\Psi(L) = 1 + hD$, i.e., one determines velocity via applying D, and then one projects forward h units, adding the value of the current observation. Observe that in the repeated roots case of Example 2, we obtain this $d = 2$ forecast function as the limit as the root λ_1 tends to zero, which is intuitive.

4 Concurrent signal extraction

As discussed in McElroy and Trimbur (2006, 2011), it is possible to define MSE optimal signal extraction filters based on a bi-infinite sample. In this section we now write $Y = S + N$, where S and N are unobserved component stochastic processes, which themselves may be ICFN processes. More typically, at most one of the signal and noise components is non-stationary; see the discussion in McElroy and Trimbur (2006), and applications to the discretely observed context in McElroy and Trimbur

(2011). Essentially, if one component is integrated, then it may be associated with the trend—that is to say, any trend dynamics in the process may be associated with an integrated component process. The usual situation is that S is ICFN, while N is a stationary CFN process.

The classical Wiener-Kolmogorov (Wiener 1949) signal extraction problem is to compute the MMSE estimate of the unobserved signal S at time t, given a bi-infinite sample $\{Y(s)\}_{s\in\mathbb{R}}$. When the process is Gaussian, this estimate is linear and can be given as a linear filter operating on the data process. Denote this filter by $\Phi(L)$—with kernel $\phi(x)$—so that we have by definition

$$\widehat{S(t)}_{|_{-\infty}^{\infty}} = \Phi(L)Y(t).$$

The notation shows that the filter depends on past and future values of the data process. A precise formulation of the optimal estimator is given in Theorem 4 of the Appendix. In practical applications, it is of more interest to compute the concurrent—or asymmetric—signal estimate, which only depends on present and past data; i.e., the concurrent signal extraction problem restricts to the semi-infinite sample $\mathbf{Y}$. Even more generally, we may wish to compute the estimate of the signal at time t based on data up to time $t+h$. This is denoted by

$$\widehat{S(t)}_{|_{-\infty}^{t+h}} = \Theta(L)Y(t),$$

and $\theta(x)$ is supported on $[-h,\infty)$. When $h = 0$ we have the concurrent signal extraction problem. When $h > 0$ we have an asymmetric signal extraction problem, but when $h < 0$ we are forecasting the signal ahead $-h$ steps. We generally refer to these cases as the asymmetric signal extraction problem. This section sets out a formula for the kernel θ (that depends on h). As the result below shows, the optimal filter is obtained by applying $\Phi(L)$ to the forecast-extended process $Y(t)$, utilizing the multi-step ahead forecasting of Sects. 2 and 3.

Before stating the theorem, we must introduce a few concepts. While the signal extraction filter $\Phi(L)$ can be computed even when Y is not ICFN—it is only required that the signal be ICFN and the noise CFN (McElroy and Trimbur 2006)—in order to forecast Y we require (in this paper) that it be ICFN. In other words, the component processes must be defined such that their sum is an ICFN (that this need not always be true, just consider the case that S is Brownian Motion and N is continuous time white noise). Examples are given after Theorem 3. Furthermore, we must assume that the conditions described in Theorem 1 of McElroy and Trimbur (2006) hold; namely, that $\{D^j Y(0)\}_{j=0}^{d}$ are uncorrelated with $\{N(t)\}$ and $\{U(t)\}$, where $U(t) = D^d S(t)$—the noise is assumed to be stationary CFN and the signal is potentially integrated (though $d = 0$ is fine too). Classically and conceptually speaking (Bell 1984), signal and noise should be orthogonal, so we assume N and U are uncorrelated with one another. Furthermore, the spectral densities of U and W—defined by the magnitude squared of the Fourier Transform of the causal kernel of each process, multiplied by the innovation variance—should have ratio that is suitably integrable and smooth. In particular, say

$$D^d S(t) = U(t) = G_S(L)\xi(t) \qquad D^d Y(t) = W(t) = G(L)\epsilon(t)$$
$$f_U(\lambda) = |G_S(e^{-i\lambda})|^2 \sigma_\xi^2 \qquad f_W(\lambda) = |G(e^{-i\lambda})|^2 \sigma^2$$

for notation, the second line defining the spectral densities. Then define $\Phi(e^{-i\cdot})$ to be f_U/f_W. Under suitable conditions on this function, $\Phi(L)$ will be the desired MMSE signal extraction filter for a bi-infinite sample. We summarize this as the following condition:

Assumption S Suppose that $\{D^j Y(0)\}_{j=0}^d$ are uncorrelated with N and U, and that N is uncorrelated with U. Also suppose that $\Phi(e^{-i\lambda}) = f_U(\lambda)/f_W(\lambda)$ is integrable with $d-1$ continuous derivatives (though if $d = 0$, we only require continuity).

Note that the initial value condition in Assumption S implies Assumption A, since $W(t) = U(t) + D^d N(t)$. Below we express the general solution for an ICFN process, and also specialize to the CARIMA case.

Theorem 3 *Let Y be a Gaussian ICFN process (5) with $G(L)$ causal, such that the kernel g is either Dirac at zero or is a function with domain $\mathbb{R}^+$. Let $q^{(j)}(x) = [I^{d-j} g](x)$ for $0 \leq j < d$ and any $x \geq 0$. Under Assumption A, the MMSE signal extraction filter $\Theta(L)$ has frf*

$$\Theta(e^{-i\lambda}) = \int_{-h}^{\infty} \phi(x) e^{-i\lambda x}\, \mathrm{d}x$$
$$+ e^{i\lambda h} \frac{\int_0^\infty \int_0^\infty \phi(-u-h) g(u+x)\, \mathrm{d}u\, \mathrm{d}x + \sum_{j=0}^{d-1} \int_0^\infty \phi(-u-h) q^{(j)}(u)\, \mathrm{d}u\ (i\lambda)^{d-1-j}}{\int_0^\infty g(x) e^{-i\lambda x}\, \mathrm{d}x}.$$

Furthermore, suppose the CFN $\{W(t)\}$ is a CARMA process (2), and let $c_\phi^{(r)}(z)$ be defined via

$$c_\phi^{(r)}(z) = \frac{\partial^{k_r - 1}}{\partial y^{k_r - 1}} \left\{ \frac{b(y) \left[\int_0^\infty \phi(-x-h) \exp\{yx\}\, \mathrm{d}x\right]}{(z-y)\Pi_{\ell \neq r}(y - \lambda_\ell)^{k_\ell}} \right\} |_{y=\lambda_r} \tag{7}$$

for any complex $z \neq \lambda_r$, for $r = 1, \ldots, P$. Then the signal extraction filter frf is given by

$$q_\phi^{(j)}(h) = \sum_{r=1}^{P} \frac{1}{(k_r - 1)!} \frac{\partial^{k_r - 1}}{\partial y^{k_r - 1}} \left\{ \frac{b(y) \left[\int_0^\infty \phi(-x-h) [I^{d-j} e^{y\cdot}](x)\, \mathrm{d}x\right]}{\Pi_{\ell \neq r}(y - \lambda_\ell)^{k_\ell}} \right\} |_{y=\lambda_r}$$
$$\Theta(e^{-i\lambda}) = \int_{-h}^{\infty} \phi(x) e^{-i\lambda x}\, \mathrm{d}x + e^{i\lambda h} \frac{a(i\lambda)}{b(i\lambda)}$$
$$\times \left(\sum_{r=1}^{P} \frac{c_\phi^{(r)}(i\lambda)}{(k_r - 1)!} + \sum_{j=0}^{d-1} q_\phi^{(j)}(h) (i\lambda)^{d-1-j} \right).$$

Remark 5 Letting $h \to \infty$ should recover the classical Wiener–Kolmogorov filter formula. By the integrability of ϕ and the Dominated Convergence Theorem, this is indeed the case, i.e., $\lim_{h\to\infty} \Theta(e^{-i\lambda}) = \Phi(e^{-i\lambda})$ for each λ.

To use this result, we need to compute the various c_ϕ and q_ϕ quantities. Also, one needs to have the spectra of the signal and noise processes to construct the signal extraction filter $\Phi(L)$. In discrete time series signal extraction, one either assumes a known form for the signal and noise and constructs the so-called reduced form for the observed process (called the structural approach), or one assumes a known form for the observed process and decomposes it as desired into signal and noise (called the decomposition approach). These methods for discrete time series are discussed in an abundance of references; see McElroy (2008) and the cited literature therein. We will illustrate the application of Theorem 3 by developing two examples that elucidate, respectively, the structural and decomposition approaches for continuous time processes.

Example 5 **Cycle extraction** Let the signal be a CARMA(2,0) cycle process $S(t)$ satisfying

$$(D^2 - 2\log\rho D + \log^2\rho + \lambda_c^2)S(t) = \kappa(t),$$

where $\rho \in (0,1)$ and $\lambda_c > 0$, and $\{\kappa(t)\}$ is WN($q\sigma^2$). As discussed in McElroy and Trimbur (2006), ρ measures persistence and λ_c is the principal frequency of the business cycle. This differs from the CARMA(2,1) cycle described in Harvey (1989), but we prefer the CARMA(2,0) for our example, because it has more integrability, which in turn facilitates our calculations. Let the AR operator be denoted $a(D)$; the roots are $\log\rho + i\lambda_c$ and $\log\rho - i\lambda_c$, denoted as λ_1 and λ_2 respectively. Suppose that a transitory component $N(t)$ contaminates the cycle, given by a CAR(1), namely $(D+\mu)N(t) = \epsilon(t)$ is WN(σ^2). Call this AR operator $c(D)$. Here $q > 0$ is the signal-to-noise ratio, being the ratio of the innovation variances of signal and noise. We first derive the CARMA(3,2) process for $Y(t) = S(t)+N(t)$. By summing spectral densities, we have

$$f_Y(\lambda) = \frac{|c(i\lambda)|^2 q\sigma^2 + |a(i\lambda)|^2\sigma^2}{|a(i\lambda)|^2|c(i\lambda)|^2}.$$

Thus the AR polynomial for $Y(t)$ is simply $a(D)c(D)$, but some additional mathematics is required to find the MA polynomial. Upon simplification, the numerator of f_Y is found to be

$$\lambda^4 + (2\log^2\rho - 2\lambda_c^2 + q)\lambda^2 + (\log^2\rho + \lambda_c^2)^2 + \mu^2 q, \tag{8}$$

multiplied by σ^2. Suppose the MA polynomial of $Y(t)$ is some $\tau(D) = D^2 + \tau_1 D + \tau_2$. Equating $|\tau(i\lambda)|^2$ to (8) yields the equations

$$\begin{aligned}\tau_1^2 - 2\tau_2 &= 2\log^2\rho - 2\lambda_c^2 + q\\ \tau_2^2 &= (\log^2\rho + \lambda_c^2)^2 + \mu^2 q.\end{aligned}$$

This represents a spectral decomposition calculation, which is often computed in discrete time series analysis. Here we solve for τ_2, obtaining both a positive and negative solution, and plug into the first equation to obtain τ_1. Up to four solution pairs are obtained (though some may be discarded if they produce non-real values of τ_1). We retain any solutions that produce zeroes ζ_1, ζ_2 of $\tau(z)$ that have negative real part; for simplicity of exposition, we assume the roots have multiplicity one. For a more general treatment of the spectral factorization of CARMA processes, see the Appendix. Then $a(D)c(D)Y(t) = \tau(D)\xi(t)$, where ξ is WN(σ^2). The signal extraction filter has frf and kernel

$$\Phi(e^{-i\lambda}) = q\,\frac{|c(i\lambda)|^2}{|\tau(i\lambda)|^2} \qquad \phi(x) = \sum_{j=1}^{2} \frac{c(\overline{\zeta_j})c(\zeta_j)}{\tau(\overline{\zeta_j})\dot{\tau}(\zeta_j)}\,\exp\{\zeta_j|x|\},$$

using complex integration (see McElroy and Trimbur (2006)). Then we can apply Theorem 3, noting that $d = 0$, focusing on the case that $h \geq 0$. Computing $c_\phi^{(r)}(z)$ via (7) for the three roots $\lambda_1, \lambda_2, \lambda_3 = -\mu$ yields

$$\alpha_r = -\sum_{j=1}^{2} \frac{c(\overline{\zeta_j})c(\zeta_j)e^{\zeta_j h}}{(\lambda_r + \zeta_j)\tau(\overline{\zeta_j})\dot{\tau}(\zeta_j)} \qquad c_\phi^{(r)}(z) = \frac{\tau(\lambda_r)\alpha_r}{(z-\lambda_r)\prod_{\ell\neq r}(\lambda_r - \lambda_\ell)}$$

for $r = 1, 2, 3$. This will then produce the frf for the concurrent signal extraction filter given by

$$\int_{-h}^{\infty} \phi(x)\,e^{-i\lambda x}\,\mathrm{d}x + e^{i\lambda h}\,\frac{a(i\lambda)c(i\lambda)}{\tau(i\lambda)} \sum_{r=1}^{3} c_\phi^{(r)}(i\lambda),$$

which has poles at $\zeta_1, \zeta_2, \lambda_1, \lambda_2, \lambda_3$. This frf is integrable as a function of λ, and hence the kernel can be calculated explicitly.

$$c_r = \frac{\tau(\lambda_r)\alpha_r}{\prod_{\ell\neq r}(\lambda_r - \lambda_\ell)}$$
$$\theta(x) = \phi(x) + \sum_{j=1}^{2} e^{\zeta_j(x+h)}\,\frac{a(\zeta_j)c(\zeta_j)}{\dot{\tau}(\zeta_j)} \sum_{r=1}^{3} c_\phi^{(r)}(\zeta_j) + \sum_{r=1}^{3} e^{\lambda_r(x+h)}\,\frac{a(\lambda_r)c(\lambda_r)}{\tau(\lambda_r)}\,c_r,$$

for $x \geq -h$ (and the kernel is zero otherwise).

Example 6 **Trend extraction** Suppose Y is a CARIMA(p,1,0) with AR polynomial $c(D)$ of degree $p \geq 0$, and we wish to decompose the process in terms of an integrated signal and a stationary noise process whose signal-to-noise ratio is a pre-determined $q > 0$ (decided upon by the practitioner). Letting $b(D) = \sqrt{q} + D$ and $W(t) = DY(t)$, we may write

$$c(D)W(t) = \epsilon(t) = \frac{\sqrt{q}}{\sqrt{q}+D}\,\epsilon(t) + \frac{D}{\sqrt{q}+D}\,\epsilon(t),$$

from which it is apparent that we can decompose Y into signal and noise processes given as follows: $c(D)b(D)DS(t) = \xi(t)$ which is WN($q\sigma^2$), and $c(D)b(D)N(t) = \eta(t)$ which is WN(σ^2), and with the two white noises independent of each other. Then the sum of these processes has the same second order structure as Y, and hence we can write $Y = S + N$. The signal extraction filter is easy to calculate (McElroy and Trimbur 2006):

$$\Phi(e^{-i\lambda}) = \frac{1}{1+\lambda^2/q} \qquad \phi(x) = \frac{\sqrt{q}}{2}\exp\{-\sqrt{q}|x|\},$$

which is the double-exponential weighting kernel. We will apply Theorem 3 for the case that $c(z) = 1$ (so for Gaussian processes, Y is a Brownian Motion), and $h \geq 0$: the frf is

$$\Theta(e^{-i\lambda}) = \int_{-h}^{\infty} \phi(x) e^{-i\lambda x}\, \mathrm{d}x + \int_{-\infty}^{-h} \phi(x)\, \mathrm{d}x\; e^{i\lambda h},$$

which has kernel (which acts upon $Y(t)$, although the data goes up through time $t+h$) given by

$$\frac{\sqrt{q}}{2}\, e^{-\sqrt{q}|x|} 1_{[-h,\infty)}(x) + \frac{1}{2} e^{-\sqrt{q}h}\; \Delta_{-h}(x).$$

Here Δ_{-h} denotes the Dirac delta function at $-h$, which when integrated against $Y(t-x)$ yields $Y(t+h)$. As h increases, the portion of the asymmetric filter due to the forecast is increasingly obsolete. The actual estimate therefore takes the form

$$\widehat{S(t)}_{|_{-\infty}^{t+h}} = \frac{\sqrt{q}}{2} \int_{-h}^{\infty} e^{-\sqrt{q}|x|} Y(t-x)\, \mathrm{d}x + \frac{1}{2} e^{-\sqrt{q}h}\; Y(t+h).$$

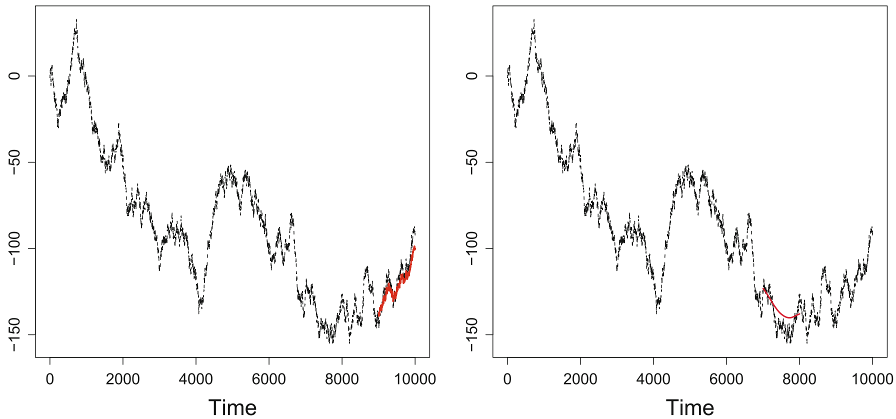

Fig. 1 Simulated Brownian Motion (*dotted*) with signal extraction estimate overlaid in *red* (*solid*). The *left panel* uses $h = 0$ (concurrent filter) and the *right panel* uses $h = 20$ (asymmetric filter)

If the process is sampled at suitably high rate, we might discretize the integral in this estimate accordingly, and thereby obtain a weighted sum of the data, plus a constant times the most current data point. Our data goes up through time $t+h$, but we seek to estimate the trend at time t. In Fig. 1, we have simulated a Brownian Motion of length 10000, and applied the above estimate for values of t equal to 9001 through 10000. Note that the exponential weights decay quite swiftly, so there is no loss in practice by taking only the past 9000 data points rather than an entire semi-infinite past (which is unavailable). The plots were computed using $q = .1$ and $h = 0$ and $h = 20$ (for left panel and right panel respectively), with a sampling frequency of 100. When h is close to zero, the forecasting portion of the filter is dominant, which essentially gives an attenuated present value; as $h \to \infty$, the asymmetric filter more closely resembles the symmetric filter, yielding an unbiased smooth trend estimate.

5 Appendix

5.1 Proofs

Proof of Theorem 1 We first need to show that the associated error process is orthogonal to the sample $\mathbf{Y}$. Let $G(L)$ be the filter associated with the causal kernel g. The filter is defined by its given frf, and substituting L for $e^{-i\lambda}$ produces $\Psi(L) = \int_0^\infty g(x+h)L^x \, \mathrm{d}x / G(L)$. Then the error is

$$
\begin{aligned}
\varepsilon(t) &= \widehat{Y(t+h)} - Y(t+h) = (\Psi(L) - L^{-h})Y(t) = (\Psi(L) - L^{-h})G(L)\epsilon(t) \\
&= \left(\int_0^\infty g(x+h)L^x \, \mathrm{d}x - L^{-h}G(L) \right) \epsilon(t) \\
&= \left(\int_h^\infty g(x)L^{x-h} \, \mathrm{d}x - \int_0^\infty g(x)L^{x-h} \, \mathrm{d}x \right) \epsilon(t) \\
&= - \left(\int_0^h g(x)L^{x-h} \, \mathrm{d}x \right) \epsilon(t).
\end{aligned}
$$

Hence $\varepsilon(t)$ is a linear combination of $\epsilon(s)$ for $t \leq s \leq t+h$. In particular, its covariance with $Y(s)$ for any $s \leq t$ equals

$$
\int_0^h g(x)g(x-h+s-t) \, \mathrm{d}x = 0.
$$

This proves optimality of the filter. Now the stated expression for $\Psi(L)$ in the CARMA case is obtained as follows. Using (3) we have

$$
\int_0^\infty g(x+h)e^{-i\lambda x} \, \mathrm{d}x = \sum_{r=1}^{P} \frac{1}{(k_r-1)!} \frac{\partial^{k_r-1}}{\partial y^{k_r-1}} \left\{ \frac{b(y)\exp\{yh\}}{(i\lambda - y)\Pi_{\ell \neq r}(y-\lambda_\ell)^{k_\ell}} \right\} |_{y=\lambda_r}.
$$

Hence with the definition of $c^{(r)}(z)$, we obtain the given frf. The error spectral density is identified immediately from the forecast error expression given above. □

Proof of Theorem 2 The strategy is similar to the stationary case. We will demonstrate that the forecast error process only depends on future innovations, and hence together with Assumption A and (6) must be orthogonal to the data, and hence MSE optimal. Let q be a function that solves the ordinary differential equation (ODE) $q^{(d)}(x) = g(x)$ for $x \geq 0$ with boundary conditions $q^{(j)}(0) = 0$ for $0 \leq j < d$, and consider the filter defined by

$$L^{-h} - \frac{\int_0^h q(x) L^{x-h} \, \mathrm{d}x \cdot D^d}{G(L)},$$

which has forecast error process only depending on future innovations. The frf of this filter is

$$\begin{aligned} &G^{-1}(e^{-i\lambda})\, e^{i\lambda h} \left[G(e^{-i\lambda}) - (i\lambda)^d \int_0^h q(x)\, e^{-i\lambda x} \, \mathrm{d}x \right] \\ &\quad = G^{-1}(e^{-i\lambda})\, e^{i\lambda h} \left[\int_0^\infty g(x) e^{-i\lambda x} \, \mathrm{d}x + (i\lambda)^{d-1} q(h) e^{-i\lambda h} \right. \\ &\qquad \left. + \cdots + q^{(d-1)} e^{-i\lambda h} - \int_0^h g(x) e^{-i\lambda x} \, \mathrm{d}x \right] \\ &\quad = G^{-1}(e^{-i\lambda}) \left[\int_0^\infty g(x+h) e^{-i\lambda x} \, \mathrm{d}x + \sum_{j=0}^{d-1} q^{(j)}(h) (i\lambda)^{d-1-j} \right]. \end{aligned}$$

The first equality utilizes integration by parts repeatedly, which produces various boundary terms, using the boundary conditions $q^{(j)}(0) = 0$. The final expression is a causal filter, and so it only remains to solve the ODE. Recursively integrating and using each boundary condition allows us to write $q^{(j)}(x) = [I^{d-j} g](x)$, and so we have proved that the given $\Psi(L)$ is MSE optimal. When the process is also CARIMA, we can calculate the $q^{(j)}(h)$ quantities by applying the I^{d-j} operator to the expression (3), which by linearity only affects the exponential term. Also the error spectral density follows immediately from the above forecast error decomposition. □

Proof of Theorem 3 The optimal estimate is

$$\int_{-h}^\infty \phi(x) Y(t-x) \, \mathrm{d}x + \int_{-\infty}^{-h} \phi(x) \widehat{Y(t-x)}_{|_{-\infty}^{t+h}} \, \mathrm{d}x,$$

since the resulting error process is

$$[\widehat{S(t)}_{|_{-\infty}^{\infty}} - S(t)] + \int_{-\infty}^{-h} \phi(x) (\widehat{Y(t-x)} - Y(t-x)) \, \mathrm{d}x,$$

which is orthogonal to the sample $\{Y(s), s \leq t+h\}$; the first term is orthogonal to the whole data process, while the second term is orthogonal to the sample by Theorem 1. Now the action of this filter on $Y(t)$ is given by

$$\int_{-h}^{\infty} \phi(x) L^x \, \mathrm{d}x + \int_{-\infty}^{-h} \phi(x) \Psi_{-(x+h)}(L) \, \mathrm{d}x \, L^{-h},$$

where we have denoted the forecast filter operating on $Y(t+h)$ that forecasts $-(x+h)$ steps ahead by $\Psi_{-(x+h)}$. Next, insert the filter formulas from Theorem 2, and by linearity the integration against ϕ operates directly on g and the $q^{(j)}$ functions. Simplifying, the result is the stated frf $\Theta(e^{-i\lambda})$. In the CARIMA case, we can again insert the corresponding expressions from Theorem 2, and by linearity obtain the modified function $c_{\phi}^{(r)}$ and $q_{\phi}^{(j)}$ as stated. □

5.2 Spectral factorization for CARMA processes

When summing two independent CARMA processes, the aggregate will also be a CARMA process, whose spectral density can be computed as the sum of the corresponding spectra. Say the processes are written informally as $b^{(j)}(D)/a^{(j)}(D)\,\epsilon^{(j)}(t)$ for $j = 1, 2$ and two independent WN $(\sigma^2_{(j)})$ sequences, and the MAs have order q_j and the ARs have order p_j, each satisfying the stationarity conditions described in Sect. 2. Summing their spectra and finding a common denominator yields

$$\frac{|b^{(1)}(i\lambda)|^2 |a^{(2)}(i\lambda)|^2 \sigma^2_{(1)} + |b^{(2)}(i\lambda)|^2 |a^{(1)}(i\lambda)|^2 \sigma^2_{(2)}}{|a^{(1)}(i\lambda)|^2 |a^{(2)}(i\lambda)|^2}.$$

Now the polynomial in λ^2 in the denominator has order $p_1 + p_2$, whereas the polynomial in λ^2 in the numerator has order $q = \max\{q_1 + p_2, p_1 + q_2\}$. Of course the function is integrable. As shown below, we can find a polynomial c and constant σ^2 such that $|c(i\lambda)|^2 \sigma^2$ equals the numerator expression, with the result that the aggregate process can be represented as $c(D)/[a^{(1)}(D) a^{(2)}(D)]\,\epsilon(t)$, where ϵ is WN(σ^2).

First, the numerator polynomial can be written as some σ^2 times $\gamma(\lambda^2) = 1 + \gamma_1 \lambda^2 + \cdots + \gamma_q \lambda^{2q}$. The coefficients γ_k are determined by expanding and recombining the CARMA polynomials. The polynomial γ has q roots $\{\omega_k\}$, and hence $\gamma(\lambda^2) = \Pi_{k=1}^{q} (1 - \lambda^2 \omega_k^{-1})$. These roots may be real or complex; either way, we can write each factor above as $(1 - \lambda^2 \omega_k^{-1}) = (1 - \lambda \omega_k^{-1/2})(1 + \lambda \omega_k^{-1/2})$. Thus the full collection of roots of $\gamma(\lambda^2)$ is $\{\omega_k^{1/2}, -\omega_k^{1/2}\}$.

On the other hand, we wish to construct $c(z) = \Pi_{k=1}^{q} (1 - z \zeta_k^{-1})$ such that $|c(i\lambda)|^2 = \gamma(\lambda^2)$ and its roots ζ_k have negative real part. At once we have the identification $\zeta_k = i \omega_k^{1/2}$, by the uniqueness of factorization. That is, we consider the entire collection $\{\pm i \omega_k^{1/2}\}$ and retain those having negative real part. Due to the plus-and-minus structure, these quantities will always occur in pairs with non-negative and non-positive real portions. Now the real portion is equal to zero iff the corresponding

ω_k equals a non-negative real number. This in turn happens iff $\gamma(\lambda^2)$ takes on a zero value at $\lambda = \sqrt{\omega_k}$. In analogy to the case of discrete-time series spectral decompositions, this is a sort of "non-invertibility," in that the spectral density has a zero. So long as the MA polynomials $b^{(1)}$ and $b^{(2)}$ have roots with negative real part (and the AR polynomials always have this property by assumption), there can be no spectral zeroes, and c is well-defined.

In summary, the algorithm involves first computing $\gamma(\lambda^2)$ and σ^2 by simple algebra; secondly, we compute the roots $\{\omega_k\}$ of γ; thirdly, compute the collection $\{\pm i\omega_k^{1/2}\}$ and retain the subset with negative real part, designating these to be the $\{\zeta_k\}$. Finally, construct c via multiplication.

5.3 Symmetric signal extraction

For ease of reference, we repeat the main theorem of McElroy and Trimbur (2006) here. We take the notation and concepts of Sect. 4 as given, but can relax the assumption that the signal and noise processes have a Wold form. Rather, we assume only that U and N have autocovariance functions that are either integrable, or are interpreted as a multiple of the Dirac function. This includes the case considered in Sect. 4.

Theorem 4 (McElroy and Trimbur 2006) *Suppose that U and N have autocovariance functions that are either integrable or are multiples of the Dirac function at the origin, and suppose that Assumption B holds. Then the MMSE signal extraction filter $\Phi(L)$ has frf given by $\Phi(e^{-i\lambda}) = f_U(\lambda)/f_W(\lambda)$. Because this is integrable by assumption, the kernel ϕ exists. The spectral density of the error process is $f_U f_N/f_W$.*

A proof of this result is given in McElroy and Trimbur (2006).

References

Andrei, N. (2006). Modern control theory—a historical perspective. *Studies in Informatics and Control, 10*, 51–62.

Astrom, K., Murray, R. (2008). *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton: Princeton University Press.

Barndorff-Nielsen, O., Shepherd, N. (2001). Non-Gaussian Ornstein-Uhlenback-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B, 63*, 167–241.

Bell, W. (1984). Signal extraction for nonstationary time series. *Annals of Statistics, 12*, 646–664.

Bergstrom, A. R. (1988). The history of continuous-time econometric models. *Econometric Theory, 4*, 365–383.

Bergstrom, A. R. (1990). *Continuous Time Econometric Modelling*. New York: Oxford University Press.

Brockwell, P. (2000). Heavy-tailed and non-linear continuous-time ARMA models for financial time series. In W. S. Chan, W. K. Li, H. Tong (Eds.), *Statistics and Finance: An Interface* (pp. 3–22). London: Imperial College Press.

Brockwell, P. (2001). Lévy-driven CARMA processes. *Annals of the Institute of Statistical Mathematics, 53*, 113–124.

Brockwell, P. (2004). Representations of continuous-time ARMA processes. *Journal of Applied Probability, 41*, 375–382.

Brockwell, P., Lindner, A. (2009). Existence and uniqueness of stationary Lévy-driven CARIMA processes. *Stochastic Processes and their Applications, 119*, 2660–2681.

Brockwell, P., Farrazzano, V., Klüppelberg, C. (2012). High-frequency sampling of a continuous-time ARMA process. *Journal of Time Series Analysis, 33*, 152–160.

Folland, G. (1995). *Introduction to Partial Differential Equations*. Princeton: Princeton University Press.
Harvey, A. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
Jones, R. (1981). Fitting a continuous-time autoregression to discrete data. In D. F. Findley (Ed.), *Applied Time Series Analysis* (pp. 651–674). New York: Academic Press.
Karatzas, I., Shreve, S. (1998). *Methods of Mathematical Finance*. New York: Springer.
Kilian, C. (2005). *Modern Control Technology*. Clinton Par New York: Thomson Delmar Learning.
McElroy, T. (2008). Matrix formulas for nonstationary ARIMA signal extraction. *Econometric Theory, 24*, 1–22.
McElroy, T., Findley, D. (2010). Discerning between models through multi-step ahead forecasting errors. *Journal of Statistical Planning and Inference, 140*, 3655–3675.
McElroy, T., Trimbur, T. (2006). Continuous time extraction of a nonstationary signal with illustrations in continuous low-pass and band-pass filtering. Census Bureau Research Report, RRS2006/13.
McElroy, T., Trimbur, T. (2011). On the discretization of continuous-time filters for nonstationary stock and flow time series. *Econometric Reviews, 30*, 475–513.
Priestley, M. (1981). *Spectral Analysis and Time Series*. London: Academic Press.
Tsai, H., Chan, K. (2009). A note on the non-negativity of continuous-time ARMA and GARCH processes. *Statistics and Computing, 19*, 149–153.
Wiener, N. (1949). *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*. New York: Wiley.