RESEARCH REPORT SERIES
(Statistics #2006-13)

Continuous Time Extraction of a Nonstationary Signal with Illustrations in Continuous Low-pass and Band-pass Filtering

Tucker McElroy
Thomas Trimbur[1]

U.S. Census Bureau
Washington, DC 20233

Federal Reserve Board[1]

Report Issued: November 16, 2006

Continuous Time Extraction of a Nonstationary Signal with Illustrations in Continuous Low-pass and Band-pass Filtering

Tucker McElroy and Thomas Trimbur
U.S. Census Bureau and Federal Reserve Board

Abstract

This paper sets out the theoretical foundations for continuous-time signal extraction in econometrics. Continuous-time modeling gives an effective strategy for treating stock and flow data, irregularly spaced data, and changing frequency of observation. We rigorously derive the optimal continuous-lag filter when the signal component is nonstationary, and provide several illustrations, including a new class of continuous-lag Butterworth filters for trend and cycle estimation.

Keywords. Signal Extraction, Continuous Time Processes, Linear Filtering, Hodrick-Prescott Filter, Band-Pass Filter.

1 Introduction

The use of continuous-time models in econometrics has an extensive history; Bergstrom (1988) gives a comprehensive review of the subject. A continuous-time approach gives flexibility in a number of directions: in the treatment of stock and flow variables in economics, in working with missing data and with irregularly spaced data, and in handling a general frequency of observation. This last point is immediately practical; even when considering just a single economic variable, with different sampling frequencies (for instance, quarterly and annual), to preserve consistency in discrete trend estimates requires a unified basis for filter design. A continuous-time analysis also allows us to investigate more subtle models of irregular spacing and interval dependence in economic activity, as in Stock (1987, 1988). Overall, it gives a natural framework for exploring theoretical links with macroeconomic dynamics and for addressing a number of real-life issues that emerge in the analysis of economic data. These and other advantages are discussed in Bergstrom (1990).

Among recent developments, an approach to model estimation for discrete datasets is developed in Jones (1981), Harvey and Stock (1985), and Harvey (1989). The statistical treatment, based on the discrete state space form and the state space smoother, gives an efficient and general algorithm. Bergstrom (1988, p. 378-9) notes that the approach is readily adapted to handle missing values, changing intervals, and other data irregularities. Harvey and Stock (1993) build on this work by setting out methods for optimal interpolation and smoothing. Since this framework includes both stock and flow series and allows for an arbitrary sampling interval, the flexibility afforded by continuous-time can be brought to bear on the signal extraction problem. Note that these methods rely on the *discretization* of the underlying continuous-time model, so that the analysis is done in discrete-time.

In this paper, in contrast, we consider the signal extraction problem purely in continuous-time; in this way the frequency and time domain analysis is done independently of sampling type and interval. By investigating the properties of the optimal *continuous-lag filters,* we are able to make links with results from the engineering literature. For our purposes, however, interest centers on formulations appropriate for continuous-time econometrics.

A key result of the paper is a proof of the signal extraction formula for nonstationary models; this is crucial for many applications in economics. In discrete-time, methods for optimal smoothing and interpolation of nonstationary series rely on theoretical foundations, as set out in Bell (1984). In continuous-time, Whittle (1983) sketches an argument for stationary models; a satisfactory proof of the signal extraction formula in this case is provided by Kailath, Sayed, and Hassibi (2000). In this paper, we extend the proof to the case of a *trend* nonstationary signal – i.e., a process that is integrated of order $d > 0$ – which entails a careful discussion of assumptions on initial values. Similar considerations are needed in the discrete-time signal extraction theory as well, as demonstrated in Bell (1984). We also treat the case that the noise process is continuous-time white noise (e.g., the *derivative* of Brownian Motion); this is of interest, since such white noise processes arise in some of the more popular continuous-time econometric models.

A second result is the development of a class of continuous-lag Butterworth filters for economic data. In particular, we introduce low-pass and band-pass filters in continuous-time that are analogous to the filters derived by Harvey and Trimbur (2003) for the corresponding discrete-time models, and their properties are illustrated through plots of the continuous-time gain functions. One special case of interest is the derivation of a continuous-lag filter from the smooth trend model; this gives a continuous-time extension of the popular Hodrick-Prescott (HP) filter (Hodrick and Prescott, 1997). At the root of the model-based band-pass is a class of higher order cycles in continuous-time. This class generalizes the stochastic differential equation (SDE) model for a stochastic cycle developed in Harvey (1989) and Harvey and Stock (1993).

The study of business cycles has remained of interest to researchers and policymakers for some time. Some of the early work in continuous-time econometrics was geared toward this application; Kalecki (1935) and James and Belz (1936) used a model in the form of a differential-difference equation (DDE) to describe business cycle movements. The DDE form gives an alternative to the SDE form that is of some theoretical interest; see Chambers and McGarry (2002). Its usefulness for methodology seems, however, limited since the DDE admits no convenient representation in either the time or frequency domain. The SDE form, in contrast, has an intuitive structure. In introducing the class of higher order models, analytical expressions are derived for the spectral density and are plotted to illustrate the cyclical properties of the model.

Our formulation remains general so that, for instance, it includes the Continuous-Time Autoregressive Integrated Moving Average (CARIMA) processes of Brockwell and Marquardt (2005). These follow the SDE form and so can be handled analytically. Throughout the paper, examples are given to help explain the methodology.

Note that, in what follows, we omit any discussion of financial derivatives and pricing theory for contingent claims. This area has, of course, motivated a considerable amount of work involving continuous-time mathematics. In this paper, attention is devoted to macroeconomic series and to relevant financial indicators such as energy prices and interest rates. The valuation of derivative contracts, standard or idiosyncratic, whose return is determined by the evolution of such indicators, is a separate topic that does not concern us here. Our interest does, however, extend to the stochastic processes used to describe the dynamics of the indicators themselves, and to the signal extraction problem that emerges in this case.

This work is theoretical and does not consider real data examples. An accompanying study (McElroy and Trimbur, 2006) examines how the optimal continuous-lag filter may be discretized to yield a simple analytical formula for the discrete time weights, such that irregularly sampled data may be easily handled in a quasi-optimal fashion. For the sake of brevity and cohesion of the subject matter, we will not discuss discretization or discretely observed data any further in this paper. Section 2 discusses some basic concepts in continuous-time filtering, which may be new to some readers. We review some definitions in Priestley (1981), Hannan (1970), and Koopmans (1974), and cast the material in a notational framework that will facilitate a fluid exposition of the main results. The class of $CARMA$ and $CARIMA$ models of Brockwell and Marquardt (2005) is included, though our results are not contingent upon the data following this class. Section 3 gives the mathematical results for the signal extraction problem. Section 4 presents illustrations with well-known econometric models, and extends the methodology for estimating the growth rates and other dynamical properties of underlying components. Section 5 concludes, and the proofs are given in the Appendix.

2 Continuous-Time Processes and Filters

This section sets out the theoretical framework for the analysis of continuous-time signal processing and filtering. Much of the treatment follows Hannan (1970); also see Priestley (1981) and Koopmans (1974). Let $y(t)$ for $t \in \mathbb{R}$, the set of real numbers, denote a real-valued time series that is measurable and square-integrable at each time. The process is weakly stationary by definition if it has constant mean – set to zero for simplicity – and autocovariance function R_y given by

$$R_y(h) = \mathbb{E}[y(t)y(t+h)] \quad h \in \mathbb{R}. \tag{1}$$

Note that the autocovariances are defined for the continuous range of lags h. Thus if $y(t)$ is a Gaussian process, R_y completely describes the dynamics of the stochastic process. A convenient model for stationary continuous-time processes that is analogous to moving averages in discrete time series is given by

$$y(t) = (\psi * \epsilon)(t) = \int_{-\infty}^{\infty} \psi(x)\epsilon(t-x)\, dx \tag{2}$$

where $\psi(\cdot)$ is square integrable on $\mathbb{R}$, and $\epsilon(t)$ is continuous-time white noise (WN). In this case, $R_y(h) = (\psi * \psi^-)(h)$, where $\psi^-(x) = \psi(-x)$. If $y(t)$ is Gaussian, then $\epsilon(t) = dW(t)/dt$, the derivative of a standard Wiener process. Though $W(t)$ is nowhere differentiable, $\epsilon(t)$ can be defined using the theory of Generalized Random Processes, as in Hannan (1970, p. 23). It is convenient to work with models expressed in terms of the disturbance $\epsilon(t)$, because this makes it easy to see the connection with discrete models based on white noise disturbances.

As an example, Brockwell's (2001) Continuous-time Autoregressive Moving Average ($CARMA$) models can be written as

$$a(D)y(t) = b(D)\epsilon(t)$$

where $a(z)$ is a polynomial of order p, and $b(z)$ is a polynomial of order $q < p$, and D is the derivative operator. The condition for stationarity is analogous to the one for a discrete AR polynomial: the roots of the equation $a(z) = 0$ must all have strictly negative real part. It can be shown (Brockwell and Marquardt, 2005) that $y(t)$ following such a stationary $CARMA$ model can be re-expressed in the form (2), for an appropriate ψ.

Next, we define the continuous-time lag operator L via the equation

$$L^x y(t) = y(t-x) \tag{3}$$

for any $x \in \mathbb{R}$ and for all times $t \in \mathbb{R}$. We denote the identity element L^0 by 1, just as in discrete time. Then a **Continuous-Lag Filter** is an operator $\Psi(L)$ with associated **weighting kernel** ψ (an integrable function) such that

$$\Psi(L) = \int_{-\infty}^{\infty} \psi(x)\, L^x\, dx. \tag{4}$$

The effect of the filter on a process $y(t)$ is

$$\Psi(L)y(t) = \int_{-\infty}^{\infty} \psi(x)\, y(t-x)\, dx = (\psi * y)(t). \tag{5}$$

The requirement of integrability for the function $\psi(x)$ is a mild condition that is sufficient for many problems. However, when the input process is nonintegrable over t, an integrable $\psi(x)$ may become inadmissible as a kernel, i.e., it may fail to give a well-defined process as output. In such a case, we may need to assume that ψ is differentiable to a specified order, with integrable or square integrable derivatives; this naturally leads to consideration of weighting kernels in suitably regular Sobolev spaces (Folland, 1995).

This development parallels the discussion in Priestley (1981), where the filter is written as

$$\mathcal{L}[\psi](D) = \int_{-\infty}^{\infty} \psi(x) e^{-Dx}\, dx,$$

with $\mathcal{L}[\psi]$ denoting the Laplace transform of ψ. As will be discussed below, we can make the identification $D = -\log L$, which effectively maps Priestley's formulation into (4).

2.1 Continuous-lag filters in the Frequency Domain

In analogy with the discrete-time case, the frequency response function is obtained by replacing L by the argument $e^{-i\lambda}$:

$$\Psi(e^{-i\lambda}) = \int_{-\infty}^{\infty} \psi(x)\, e^{-i\lambda x}\, dx, \quad \lambda \in \mathbb{R} \tag{6}$$

Denoting the continuous-time Fourier Transform by $\mathcal{F}[\cdot]$, equation (6) can be written as $\Psi(e^{-i\lambda}) = \mathcal{F}[\psi]$.

Example 1 Consider a Gaussian kernel $\psi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. In this example, the inclusion of the normalizing constant means that the function integrates to one; since applying the filter tends to preserve the level of the process, it could be used as a simple trend estimator. The frequency response has the same form as the weighting kernel and is given by $\mathcal{F}[\psi](\lambda) = e^{-\frac{\lambda^2}{2}}$.

The power spectrum of a continuous time process $y(t)$ is the Fourier Transform of its autocovariance function R_y:

$$f_y(\lambda) = \mathcal{F}[R_y](\lambda), \quad \lambda \in \mathbb{R} \tag{7}$$

The gain function of a filter $\Psi(L)$ is the magnitude of the frequency response, namely

$$G_c(\lambda) = |\mathcal{F}[\psi](\lambda)|, \quad \lambda \in \mathbb{R}. \tag{8}$$

As in discrete time series signal processing, passing an input process through the filter $\Psi(L)$ results in an output process with spectrum multiplied by the squared gain; so the gain function gives information about how various frequencies in the data are attenuated or accentuated by the filter. Note that in contrast to the discrete case where the domain is restricted to the interval $[-\pi, \pi]$, the functions in (7) and (8) are defined over the entire real line. Given a candidate gain function $g(\lambda)$, taking the inverse Fourier Transform in continuous-time yields the associated weighting kernel:

$$\mathcal{F}^{-1}[g](x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(\lambda) e^{i\lambda x} \, dx, \quad x \in \mathbb{R} \tag{9}$$

This expression is well-defined for any integrable $g(\lambda)$. Integrability is a mild condition satisfied by nearly all filters of practical interest; it corresponds in discrete-time to the condition that the filter weights be absolutely summable.

Example 2 Weighting kernels that decay exponentially on either side of the observation point have often been applied in smoothing trends; this pattern arises frequently in discrete model-based frameworks, e.g., Harvey and Trimbur (2003). Similarly, in the continuous time setting, a simple example of a trend estimator is the double exponential weighting pattern $\psi(x) = \frac{1}{2}\, e^{-|x|}$, $x \in \mathbb{R}$. In this case, it is easy to show using integral calculus that $\Psi(L) = 1/(1 - (\log L)^2)$. The Fourier transform has the same form as a Cauchy probability density function, namely $\mathcal{F}[\psi](\lambda) = 1/(1+\lambda^2)$. This means that the gain of the low-pass filter $\Psi(L)$ decays slowly as $\lambda \longrightarrow \infty$.

2.2 The Derivative Filter and Nonstationary Processes

In (3), the extension of the lag operator L to the continuous-time framework is made explicit. In building models, we can treat L as an algebraic quantity as in the discrete-time framework. The extension of the differencing operator, $\Delta = 1 - L$, used to define nonstationary models, is discussed in Hannan (1970, p. 55) and Koopmans (1974).

To define the differentiation operator D, consider the limit of measuring the displacement of a continuous-time process, per unit of time, over an arbitrarily small interval δ:

$$\frac{d}{dt} y(t) = \lim_{\delta \to 0} \frac{y(t) - y(t-\delta)}{\delta} = \lim_{\delta \to 0} \frac{1 - L^{\delta}}{\delta} y(t) = (-\log L) y(t).$$

Thus, we see that taking the derivative d/dt has the same effect as applying the continuous lag filter $-\log L$. This holds for all differentiable processes $y(t)$, implying $D = -\log L$; note that Priestley (1981) derives the equivalent $L = \exp\{-D\}$ via Taylor series arguments. This operator D will be our main building block for nonstationary continuous-time processes. It will also be useful in thinking about rates of growth and rates of rates of growth – the velocity and acceleration of a process, respectively. We refer to $-\log L$ as the derivative filter; taking powers yields higher order derivative filters. For instance, $(\log L)^2$ gives a measure of acceleration with respect to time. We note that the frequency response of D^d is $(-i\lambda)^d$.

Standard discrete-time $ARIMA$ processes are written as difference equations, built on white noise disturbances. In analogy, continuous time processes can be written as differential equations, built on an extension of white noise to continuous-time. Thus a natural class of models is the Integrated Filtered Noise processes, which are given by

$$D^d y(t) = \Psi(L)\epsilon(t) \tag{10}$$

for some integrable ψ, and order of differential $d \geq 0$. This class will be denoted $y(t) \sim IFN(d)$; it encompasses a wide variety of linear continuous-time models. As an example, Brockwell and Marquardt (2005) define the class of Continuous-time Autoregressive Integrated Moving Average ($CARIMA$) models as the solution to

$$a(D)D^d y(t) = b(D)\epsilon(t). \tag{11}$$

Thus, applying the derivative filter d times transforms $y(t)$ into a stationary $CARMA(p,q)$ process. The autoregressive order p is the degree of the polynomial $a(D)$, and the moving average order q is the degree of the polynomial $b(z)$. The constraint $q < p$ is necessary to ensure the process is well-defined; this ensures that the spectral density of $D^d y(t)$ is an integrable function. This gives the $CARIMA(p,d,q)$ process. The original process $y(t)$ is nonstationary and is said to be integrated of order d in the continuous-time sense. Now this can be put into an $IFN(d)$ form: starting from (11), we can write

$$D^d y(t) = [b(D)/a(D)]\, \epsilon(t) \tag{12}$$

Using the definition of D, it follows that $y(t) \sim IFN(d)$ with $\Psi(L) = b(-\log L)/a(-\log L)$. Deriving the kernel $\psi(x)$ requires an expression for the rational function in terms of an integral over powers of L, namely $b(-\log L)/a(-\log L) = \int_{-\infty}^{\infty} \psi(x) L^x \, dx$. Using the formulation of Priestley (1981), we see that $CARIMA$ models can be equivalently expressed as IFN models where the kernel ψ's Laplace transform is a rational function.

Example 3 We consider a simple $CARMA(2,0)$ model that can be used to model periodic phenomena $\psi(t)$, such as a cycle or seasonal effect, with some stochastic frequency $\lambda_c \in \mathbb{R}$. Let

$$a(z) = (z + \rho - i\lambda_c)(z + \rho + i\lambda_c) \qquad b(z) = 1$$

and suppose $\epsilon(t)$ is WN with variance σ^2, in (11). Here ρ is a non-negative real number called the damping factor. Note that the roots of $a(z)$ are $\pm i\lambda_c - \rho$, with real part $-\rho \leq 0$. If ρ is strictly positive, we have a stationary process, but if $\rho = 0$ there is nonstationarity corresponding to an infinite peak in the spectrum. The spectral density is

$$f_\psi(\lambda) = \frac{\sigma^2}{(\rho^2 + (\lambda - \lambda_c)^2)(\rho^2 + (\lambda + \lambda_c)^2)}.$$

Note that one can construct higher-order cycles by considering $CARMA(2n, 0)$ models; the spectrum in this case is obtained by taking the nth power of the denominator of f_ψ above. Also note that when $\rho = 0$, the formula for f_ψ reduces to

$$f_\psi(\lambda) = \frac{\sigma^2}{(\lambda^2 - \lambda_c^2)^2},$$

which is minimized at $\lambda = \pm\lambda_c$, the spectral peaks. Of course, when $\lambda_c = 0$, this infinite spectral peak is associated with the trend frequency (and $a(z) = z^2$). Figure 1 illustrates the spectral density with the parameter values $\rho = .2$, $\lambda_c = \pi/3$, and $\sigma^2 = 1$. The maximum does not occur exactly at $\lambda = \lambda_c$, however, except as ρ tends to zero.

3 Signal Extraction in Continuous Time

This section sets out the signal extraction problem in continuous time. A new result with proof is given for the problem of extracting a nonstationary continuous-time signal from stationary noise. Whittle (1983) shows a similar result for continuous-time nonstationary processes, but omits the proof and does not discuss the importance of initial value assumptions. In fact, the filters that Whittle (1983) presents are *not* mean square error optimal unless Assumption A (see below) on the initial values holds true. Kailath, Sayed, and Hassibi (2000, p. 221 – 227) prove the formula for the special case of a stationary signal. We extend the treatment of Whittle (1983) by providing proofs, at the same time illustrating the importance of Assumption A to the result. We also handle the case that the differentiated signal and/or the noise process are WN. Suppose the following model for an underlying continuous time process $y(t)$:

$$y(t) = s(t) + n(t), \quad t \in \mathbb{R} \tag{13}$$

where the signal $s(t)$ is integrated of order d, and the noise process $n(t)$ is stationary. In general, d is any non-negative integer; for $d = 0$, $s(t)$ is stationary. For $d > 0$, $s(t)$ is nonstationary, but the dth derivative of $s(t)$, denoted by $u(t)$, is stationary. It is assumed that the differentiated signal $u(t)$ and the noise $n(t)$ are mean zero and uncorrelated with one another. For the basic case they have integrable autocovariance functions R_u and R_n, respectively. As an extension, we may also consider that one or both of R_u and R_n correspond to multiples of the Dirac delta function, and are viewed as tempered distributions with constant spectral densities (Folland, 1995). (In this case, the corresponding process is a WN with appropriate variance.) The underlying process $y(t) = s(t) + n(t)$ satisfies the differential equation

$$w(t) = D^d y(t) = D^d s(t) + D^d n(t) = u(t) + D^d n(t). \tag{14}$$

It follows from Section 2.2 that the spectral density of $w(t)$ satisfies

$$f_w(\lambda) = f_u(\lambda) + |\lambda|^{2d} f_n(\lambda). \tag{15}$$

Note that the nonstationary process $y(t)$ can be written in terms of some initial values plus a d-fold integral of the stationary $w(t)$ – see Hannan (1970, p. 81). For example, when $d = 1$ we can write

$$y(t) = y(0) + \int_0^t w(z)\, dz$$

for some initial value random variable $y(0)$; note that this is valid when $t > 0$ or when $t < 0$ as well. When $d = 2$ we can write

$$y(t) = y(0) + t\dot{y}(0) + \int_0^t \int_0^z w(x)\, dx\, dz$$

for initial position and velocity $y(0)$ and $\dot{y}(0) = \frac{dy}{dt}(0)$. In general, we can write

$$y(t) = \sum_{j=0}^{d-1} \frac{t^j}{j!} y^{(j)}(0) + [I^d w](t) \tag{16}$$

with the I operator defined by $[Iw](t) = \int_0^t w(z)\, dz$. Note that (16) holds for the signal $s(t)$ as well. In analogy with the decomposition assumptions of Bell (1984), we suppose that $u(t)$ and $n(t)$ are uncorrelated with all d initial values $y(0), \dot{y}(0), \cdots, y^{(d-1)}(0)$ (this set will be abbreviated by the vector y_*). This supposition will be referred to as Assumption A. (Since in our case nonstationarity results from integration, it is essentially a "trend" – or frequency zero – form of nonstationarity. Thus the analogous differencing operator in Bell (1984) would be $(1 - L)^d$ where L is the backshift operator.)

In analogy with discrete-time results (Bell, 1984), we seek the minimum mean square error *linear* estimate of the signal $s(t)$ given a bi-infinite sample of data, i.e., we seek to minimize $\mathbb{E}[(\hat{s}(t) - s(t))^2]$ among all $\hat{s}(t)$ that are a linear function of the data $y(h)$ with $h \in \mathbb{R}$. Since the solution is a linear estimate, we can write it as a convolution of some weighting kernel ψ and the underlying process $y(t)$:

$$\hat{s}(t) = \Psi(L)y(t) = (\psi * y)(t)$$

Note that if the underlying process $y(t)$ is Gaussian, then the best estimate is the linear estimate given above. The signal extraction problem is to determine the weighting kernel ψ; the following theorem provides the solution.

Theorem 1 *Suppose that $y(t)$ and $s(t)$ have decompositions of the form (16) and that Assumption A holds. Also suppose that $u(t)$ and $n(t)$ are mean zero weakly stationary processes that are uncorrelated with one another, which are either WN processes or have integrable autocovariance functions. Let*

$$g(\lambda) = \frac{f_u(\lambda)}{f_w(\lambda)}.$$

If g is integrable with $d-1$ continuous derivatives (if $d = 0$, we only require that g be continuous), then the linear minimum mean square error signal extraction weighting kernel exists and is given by

$$\psi(x) = \mathcal{F}^{-1}[g](x). \tag{17}$$

The spectral density of the error process $e(t) = \hat{s}(t) - s(t)$ is

$$f_e(\lambda) = \frac{f_u(\lambda) f_n(\lambda)}{f_w(\lambda)};$$

hence the MSE is $\int_{-\infty}^{\infty} f_u f_n / f_w \, d\lambda$.

Remark 1 Such a filter $\Psi(L)$ will be referred to as a continuous-time Wiener-Kolmogorov (WK) filter. One of the important properties of these WK is that they pass polynomials, in exact analogy with how discrete-time filters can be constructed to pass polynomials. By *pass* we mean that

$$\Psi(L)p(t) = p(t)$$

for a polynomial $p(t)$ of sufficiently low degree. If we write out the action of the filter on $p(t)$ as a convolution against ψ, we find that the filter passes $p(t)$ iff

$$\int_{-\infty}^{\infty} x^j \psi(x) \, dx = \delta_{j,0},$$

where δ denotes the Kronecker delta. As demonstrated in the proof of Theorem 1, a WK filter passes polynomials of degree up to $2d-1$, so long as these moments exist.

Example 4 Taking $d = 0$, let $s(t)$ have autocovariance function $R_s(h) = \frac{1}{\sqrt{2\pi}} e^{-\frac{h^2}{2}}$, which will be abbreviated by $\phi(h)$. Suppose further that $n(t)$ has autocovariance function $R_n(h) = (1 - h^2)\phi(h)$. Then y is characterized by $R_y(h) = (2 - h^2)\phi(h)$, and the associated spectral densities are

$$f_s(\lambda) = e^{-\lambda^2/2}, \qquad f_n(\lambda) = \lambda^2 e^{-\lambda^2/2}, \qquad f_y(\lambda) = (1 + \lambda^2)e^{-\lambda^2/2}$$

The signal resembles a damped trend, whereas $n(t)$ is a pink noise process that incorporates pseudo-cyclical and irregular fluctuations. The ratio of spectra $f_s(\lambda)/f_y(\lambda) = \frac{1}{1+\lambda^2}$ is integrable, and from Example 2 the inverse Fourier Transform gives a simple filter with kernel $\psi(x) = \frac{1}{2} e^{-|x|}$.

Example 5 Suppose that with $d = 1$, the spectral densities of differentiated signal and noise have constant ratio q, with $f_u(\lambda) = q\phi(\lambda)$ where ϕ is the standard normal density function. The continuous-time frequency response of the signal extraction filter is then

$$g(\lambda) = \frac{q\phi(\lambda)}{q\phi(\lambda) + \lambda^2\phi(\lambda)} = \frac{1}{1 + \lambda^2/q},$$

which transforms to the double-exponential weighting kernel $\sqrt{q}\exp\{-\sqrt{q}|x|\}/2$. Such a kernel passes lines and constants, as mentioned in Remark 1.

4 Illustrations of Continuous-Lag Filtering

The class of $CARIMA$ models are particularly convenient for computing WK weighting kernels. This is because if we have a $CARIMA$ model for the signal and a $CARMA$ model for the noise, then the observed process will be $CARIMA$ as well. Then the spectral densities f_u and f_w will be rational functions in λ^2, and hence $g(\lambda)$ will be a rational function in λ^2. It is generally fairly straight-forward to compute inverse Fourier transforms of such functions via the calculus of residues (Ahlfors, 1979). By "straight-forward," we mean there is a well-known general procedure for calculating the residues of such rational functions, even though the computation may be somewhat tedious and time-consuming in particular cases. The first two illustrations below look at particulary simple $CARIMA$ models for the signal and noise; we have chosen these examples based on their popularity in the economics literature for modeling trends, and because the formulas for the resultant WK weighting kernels are particularly simple and elegant. The third illustration adds a continuous-time cycle component to the local level and smooth trend models, which results in both a low-pass and band-pass continuous-lag filter. This is similar in spirit to the work of Harvey and Trimbur (2003) on discrete-lag low-pass and band-pass filters. Finally, we consider the estimation of the velocity and acceleration of a signal, and illustrate the applications to turning point detection.

Illustration 1**: Local Level Model** The trend plus noise model (see Harvey (1989, p. 485) for terminology) is written as $y(t) = \mu(t) + \epsilon(t)$ where $\mu(t)$ denotes the stochastic level, and $\epsilon(t)$

is continuous-time white noise with variance parameter σ_ϵ^2, denoted by $\epsilon(t) \sim WN(0, \sigma_\epsilon^2)$. An interpretation of the variance σ_ϵ^2 is that $\Theta(L)\epsilon(t)$ has autocovariance function $(\theta * \theta^-)(h)\sigma_\epsilon^2$ for any (integrable) auxiliary weighting kernel θ. The local level model assumes $D\mu(t) = \eta(t)$, where $\eta(t) \sim WN(0, \sigma_\eta^2)$. The signal-noise ratio in the continuous-time framework is defined as $q = \sigma_\eta^2/\sigma_\epsilon^2$. So the observed process $y(t)$ requires one derivative for stationarity, and we write $w(t) = Dy(t)$. The spectral densities of the differentiated trend and observed process are

$$f_\eta(\lambda) = q\sigma_\epsilon^2 \qquad f_w(\lambda) = f_\eta(\lambda) + \lambda^2\sigma_\epsilon^2 = (q + \lambda^2)\sigma_\epsilon^2$$

since $\mathcal{F}[D] = i\lambda$. Though the constant function $f_\eta(\lambda)$ is nonintegrable over the real line, the frequency response of the signal extraction filter is given by the ratio $1/[1+\lambda^2/q]$, which is integrable. As in the previous example, the weighting kernel has the double exponential shape:

$$\psi(x) = \frac{\sqrt{q}}{2}\exp\{-\sqrt{q}|x|\}$$

The rate of decay in the tails now depends on the signal-noise ratio of the underlying continuous-time model.

Illustration 2: Smooth Trend Model The local linear trend model (Harvey 1989, p. 485) has the following specification:

$$\begin{aligned} D\mu(t) &= \beta(t) + \eta(t), \qquad \eta(t) \sim WN(0, \sigma_\eta^2) \\ D\beta(t) &= \zeta(t), \qquad \zeta(t) \sim WN(0, \sigma_\zeta^2) \end{aligned}$$

where $\eta(t)$ and $\zeta(t)$ are uncorrelated. Setting $\sigma_\eta^2 = 0$ gives the smooth trend model for which noisy fluctuations in the level are minimized and the movements occur due to changes in slope. The data generating process is $y(t) = \mu(t) + \epsilon(t)$ where $\epsilon(t)$ is white noise uncorrelated with $\zeta(t)$. Now the signal-noise ratio is $q = \sigma_\zeta^2/\sigma_\epsilon^2$. Recall that the discrete-time smooth trend model underpins the well-known HP filter for estimating trends in discrete time series; see Hodrick and Prescott (1997), as well as Harvey and Trimbur (2003). Here we develop an analogous HP filter for the continuous-time smooth trend model. We may write the model as

$$\begin{aligned} u(t) &= D^2\mu(t) = \zeta(t) \\ w(t) &= D^2y(t) = \zeta(t) + D^2\epsilon(t). \end{aligned}$$

The spectral densities of the appropriately differentiated trend and series are

$$f_u(\lambda) = q\sigma_\epsilon^2 \qquad f_w(\lambda) = f_u(\lambda) + \lambda^4 f\sigma_\epsilon^2 = (q + \lambda^4)\sigma_\epsilon^2.$$

Hence the ratio $1/(1 + \lambda^4/q)$ gives the frequency response function of the filter; the error spectrum is $\sigma_\epsilon^2/(1 + \lambda^4/q)$. Taking the inverse Fourier transform of this function (see the appendix for a detailed derivation) yields the weighting kernel

$$\psi(x) = \frac{q^{1/4} \exp\{-|x| q^{1/4}/\sqrt{2}\}}{2\sqrt{2}} \left(\cos(|x| q^{1/4}/\sqrt{2}) + \sin(|x| q^{1/4}/\sqrt{2}) \right)$$

This gives the continuous-time extension of the HP filter; it follows from Remark 1 that it passes cubics. Figure 2 graphs this weighting function, using three values of q: 10^{-3}, 10^{-4}, and 10^{-5}, representing higher, medium, and lower signal-noise ratios. The lower values of q indicate more noise, which in turn requires more smoothing – hence the wider kernels. Note the negative side-lobes, which allow the filter to pass quadratics.

Illustration 3: Continuous-Lag Band-Pass Consider again the class of stochastic cycles in Example 3. A simple (nonseasonal) model for a continuous-time process in macroeconomics is given by

$$y(t) = \mu(t) + \psi(t) + \varepsilon(t),$$

where $\mu(t)$ is a trend component that accounts for long-term movements and the cyclical component is $\psi(t)$. The irregular $\varepsilon(t)$ is meant to absorb any random, or nonsystematic variation, and in direct analogy with discrete-time, it is assumed to be $WN(\sigma_\varepsilon^2)$. In formulating the estimation of $\psi(t)$ as a signal extraction problem, we set the nonstationary signal to $\mu(t) + \varepsilon(t)$ and the *noise* to $\psi(t)$. This is done just to map the estimation problem to the framework developed in the last Section; it is not intended to suggest any special importance of the *signal* as a target of extraction. Actually in this case, the *noise* part will usually be of greater interest in a business cycle analysis. Thus, the optimal filter is constructed for $\mu(t) + \varepsilon(t)$, and the complement of this filter (i.e., that filter obtained by subtraction from the identity filter) yields the band-pass (when it is integrable). Using the spectrum of the cycle, combined with the pseudo-spectra of $\mu(t)$, it follows that

$$\begin{aligned} f_\mu(\lambda) &= \frac{q_\zeta \sigma_\varepsilon^2}{\lambda^{2d}} \\ f_\psi(\lambda) &= \frac{q_\kappa \sigma_\varepsilon^2}{(\rho^2 + (\lambda - \lambda_c)^2)(\rho^2 + (\lambda + \lambda_c)^2)} \\ f_\varepsilon(\lambda) &= \sigma_\varepsilon^2, \end{aligned}$$

where $q_\zeta = \sigma_\zeta^2/\sigma_\varepsilon^2$ is the signal-noise ratio for the trend, $q_\kappa = \sigma_\kappa^2/\sigma_\varepsilon^2$ for the cycle. Consider the general case $d \geq 0$; the band-pass frequency response is obtained from the ratio of f_ψ divided by $f_\mu + f_\psi + f_\varepsilon$. The formula is

$$BP_m(\lambda) = \frac{\lambda^{2m} q_\kappa}{\lambda^{2m} q_\kappa + (q_\zeta + \lambda^{2m})(\rho^2 + (\lambda - \lambda_c)^2)(\rho^2 + (\lambda + \lambda_c)^2)},$$

where BP_m stands for band-pass of order pair m. Here the order $m = d$ denotes the order of integration, as determined by the stochastic trend model; this parallels the development of

Harvey and Trimbur (2003). Note that this function is always integrable, and hence the band-pass weighting kernel ψ exists. Calculation of this weighting kernel requires standard complex analysis computations, which are omitted here. The low-pass filter, i.e., the trend extraction filter, has gain function

$$LP_m(\lambda) = \frac{(\rho^2 + (\lambda - \lambda_c)^2)(\rho^2 + (\lambda + \omega)^2)q_\zeta}{(\rho^2 + (\lambda - \lambda_c)^2)(\rho^2 + (\lambda + \lambda_c)^2)(q_\zeta + \lambda^{2m}) + \lambda^{2m} q_\kappa},$$

which is only integrable for $m \geq 1$ (it is of little interest to consider the model with $m = 0$). Figure 3 illustrates the low-pass and band-pass gain functions for $m = 1$ (we use the same cycle model from Example 3, so $\rho = .2$ and $\lambda_c = \pi/3$). The signal-noise ratios were chosen to be $q_\zeta = .25$ and $q_\kappa = 4$. Here the approximate band-pass shape is clearly evident in BP_1, and the low-pass LP_1 is seen to have a slight dip at the cycle frequencies. Figure 4 illustrates similar features, but with $m = 2$, $q_\zeta = .04$, and $q_\kappa = 4$. Now the band-pass has a steeper incline on the left side, and the low-pass has an even steeper drop.

We now consider an application where the "signal" of interest is a linear transformation of the original signal $s(t)$. In particular, suppose that a linear operator H is applied to the signal, so that a new signal of interest is given by $Hs(t)$. Examples of such linear operators include D and convolution with some kernel θ. The new signal-noise decomposition is then given by

$$Hy(t) = Hs(t) + Hn(t),$$

and we wish to derive a minimum mean squared error linear estimate of $Hs(t)$ given the data $Hy(h)$ for $h \in \mathbb{R}$. Because H is a linear operator, we can compute the estimate via

$$\widehat{Hs}(t) = H\hat{s}(t) = H(\psi * y)(t),$$

where ψ is the WK weighting kernel for the original signal extraction problem. Again by linearity, the above can be calculated via $(H\Psi(L))y(t)$, so long as this new weighting kernel of $H\Psi(L)$ exists.

For example, if $H = D^d$ and $\lambda^d g(\lambda)$ is integrable – where g is the frequency response of the WK filter – then the dth derivative $\psi^{(d)}(x)$ exists and gives the appropriate weighting kernel to estimate $D^d s(t)$. But if $H = \int \theta(x) L^x \, dx$, then the weighting kernel is given by $\theta * \psi$, when this exists. The first derivative of the signal can be interpreted as a growth rate; more generally, the first and second derivatives give the velocity and acceleration of the signal. These can be used to give a mechanical interpretation of the signal's dynamics; for instance, the "energy" of the signal over a time interval $[t, t+h]$ could be computed as the integral of the square of the signal's velocity estimate over that interval. Alternatively, we may be interested in an aggregate forecast of the signal, given by $H = \int_{-1}^0 L^x \, dx$; the weighting kernel in this case is given by $(1_{(-1,0)} * \psi)(t)$, or

$$\int_{-1}^0 \psi(t - x) \, dx.$$

If we wish to know the mean squared error of $Hy(t)$ as an estimate of the new signal, we merely multiply the error spectral density of Theorem 1 by the magnitude squared of $\mathcal{F}[H](\lambda)$, viewing H as a linear filter (since it is a linear operator). This is valid so long as the resulting error spectral density is integrable; its integral is then the mean squared error. For example, if $H = D$ then we multiply the error spectrum by λ^2 and integrate the resulting function.

Illustration 2, continued Here we calculate the velocity and acceleration of the trend in the Smooth Trend model. The velocity is the first derivative, and is interpreted as yielding a growth rate for the trend. The acceleration is the second derivative, and provides additional information about the curvature of the trend line. Recall from elementary calculus, that a local maximum is indicated by zero velocity together with a negative acceleration; hence these so-called "indicators" may be used in conjunction to determine the onset of a recession or downturn in an economic time series. In this case, we find that $\lambda^2/(1+\lambda^4/q)$ is integrable, so both derivatives of ψ are well-defined. Direct calculation yields

$$\dot{\psi}(x) = -\frac{q^{1/2}}{2} e^{-q^{1/4}|x|/\sqrt{2}} \sin(q^{1/4}x/\sqrt{2})$$
$$\ddot{\psi}(x) = -\frac{q^{3/4}}{2\sqrt{2}} e^{-q^{1/4}|x|/\sqrt{2}} \left(\cos(q^{1/4}|x|/\sqrt{2}) - \sin(q^{1/4}|x|/\sqrt{2}) \right).$$

The error spectrum for the velocity WK filter is $\sigma_\epsilon^2 \lambda^2/(1+\lambda^4/q)$, which is proportional to the frequency response for the acceleration weighting kernel. Figures 5 and 6 display the WK weighting kernels for the velocity and acceleration. Note that the asymmetric shape of the velocity kernel is due to the oddness of its frequency response function. The acceleration kernel is much less smooth, which is due to the lessened integrability of its frequency response function.

5 Conclusion

In summary, we have considered the signal extraction problem for nonstationary signals in stationary noise. Essentially we considered "trend" nonstationarity, where the signal process is integrated. We have illustrated the signal extraction filter calculations for several examples, many of which correspond to popular continuous-time component models. The theory for $CARIMA$ models is particularly simple, since the WK frequency response will be a rational function in λ; the Fourier Transforms of such functions are fairly easy to compute, because the poles are easy to identify. However, for econometrics applications these examples may seem abstract, because we always observe data discretely; unlike engineering, filters in econometrics must be discrete because our data is not continuously sampled in practice. A related study (McElroy and Trimbur, 2006) addresses the topic of filter discretization at length, comparing and contrasting this approach with the more prevalent model discretization methodology. In particular, McElroy and Trimbur (2006) discusses many of the examples of this paper (including the Smooth Trend Model) and describes how the WK filters may be discretized such that any underlying polynomial trends are preserved.

Future research on this topic should address the question of a semi-infinite and/or compact sample (note that Whittle (1983) treats the former case, again without proof), so that concurrent filters and continuous State space smoothers can be constructed. In addition, it may be of interest to generalize the type of nonstationarity considered, e.g., poles in the pseudo-spectrum at nonzero frequencies. It would then be natural to consider signal extraction where the noise process is also nonstationary, but with spectral poles distinct from those of the signal (as discussed in Bell (1984), difficulties arise when there are common zeroes in the differencing operators for signal and noise). A possible application would be continuous-time seasonal adjustment, where the noise has spectral poles at annual frequencies and the signal has a spectral pole at the origin.

Acknowledgements. The authors thank David Findley for many stimulating conversations on this topic, and careful reading of the manuscript.

6 Appendix: Mathematical Proofs

Proof of Theorem 1. Throughout, we shall assume that $d > 0$, since the $d = 0$ case is essentially handled in Kailath et al. (2000). In order to prove the theorem, it suffices to show that the error process $e(t) = \hat{s}(t) - s(t)$ is orthogonal to the underlying process $y(h)$. By (16), it suffices to show that $e(t)$ is orthogonal to $w(t)$ and the initial values y_*. So we begin by analyzing the error process produced by the proposed weighting kernel $\psi = \mathcal{F}^{-1}[g]$. We first note the following interesting property of ψ. The moments of ψ

$$\int z^k \psi(z)\, dz = i^k \frac{d^k}{d\lambda^k} \frac{f_u(\lambda)}{f_w(\lambda)}|_{\lambda=0}$$

for $k < d$ exist by the smoothness assumptions on g, and are easily shown to equal zero if $0 < k < 2d$ (i.e., for $d \leq k < 2d$, the moments are zero so long as they exist – their existence is not guaranteed by the assumptions of the theorem). Moreover, the integral of ψ is equal to 1 if $d > 0$. These properties ensure (when $d > 0$) that the filter $\Psi(L)$ *passes* polynomials of degree less than d. This is because $\Psi(L)t^j = t^j$ for $j < d$. Now we will utilize the orthogonal increments representation of a stationary time series (Hannan, 1970) to write

$$[I^d u](t) = \int_{-\infty}^{\infty} (i\lambda)^{-d} \left(e^{i\lambda t} - \sum_{j=0}^{d-1} \frac{(i\lambda t)^j}{j!} \right) \mathbb{Z}_u(d\lambda), \tag{18}$$

where $\mathbb{Z}_u(d\lambda)$ is the orthogonal increments spectral measure associated with $u(t)$. The above representation is obtained by computing the frequency response $[I^d e^{i\cdot}](t)$ via induction on d. Together with (16) applied to $s(t)$, (18) yields a full description of the signal. Then the error process can be written as

$$e(t) = (\psi * y)(t) - s(t) = (\psi * s)(t) - s(t) + (\psi * n)(t).$$

Using the property that $\Psi(L)$ passes polynomials on (16) and (18), we obtain

$$(\psi * s)(t) = \sum_{j=0}^{d-1} \frac{t^j}{j!} s^{(j)}(0) + \int_{-\infty}^{\infty} (i\lambda)^{-d} \left(g(\lambda) e^{i\lambda t} - \sum_{j=0}^{d-1} \frac{(i\lambda t)^j}{j!} \right) \mathbb{Z}_u(d\lambda).$$

Subtracting $s(t)$ leaves

$$(\psi * s)(t) - s(t) = \int_{-\infty}^{\infty} (i\lambda)^{-d} (g(\lambda) - 1) e^{i\lambda t}\, \mathbb{Z}_u(d\lambda).$$

Of course $g(\lambda) - 1 = -\lambda^{2d} f_n(\lambda)/f_w(\lambda)$, utilizing (15). The other portion of the error process is $(\psi * n)(t)$, which can be re-expressed in terms of the orthogonal increments representation of $n(t)$. Altogether, the error process is

$$e(t) = \int_{-\infty}^{\infty} (i\lambda)^{-d} (g(\lambda) - 1) e^{i\lambda t}\, \mathbb{Z}_u(d\lambda) + \int_{-\infty}^{\infty} g(\lambda) e^{i\lambda t}\, \mathbb{Z}_n(d\lambda),$$

which is stationary (its spectral density is computed below). Now again using (16) applied to $y(t)$, we have

$$\mathbb{E}[e(t) y(t+h)] = \mathbb{E}[e(t) \sum_{j=0}^{d-1} \frac{(t+h)^j}{j!} y^{(j)}(0)] + \mathbb{E}[e(t) [I^d w](t+h)]$$

for all $h \in \mathbb{R}$. The first term on the right hand side is zero due to Assumption A, since the terms in the error $e(t)$ are merely linear filters of the processes $u(t)$ and $n(t)$. The second term on the right is

$$\int_{-\infty}^{\infty} (i\lambda)^{-d} \left(\frac{-\lambda^{2d} f_n(\lambda)}{f_w(\lambda)} \right) e^{i\lambda t} (-i\lambda)^{-d} \left(e^{-i\lambda(t+h)} - \sum_{j=0}^{d-1} \frac{(-i\lambda(t+h))^j}{j!} \right) \mathbb{E}[\mathbb{Z}_u \overline{\mathbb{Z}_w}](d\lambda) \qquad (19)$$

$$+ \int_{-\infty}^{\infty} \frac{f_u(\lambda)}{f_w(\lambda)} e^{i\lambda t} (-i\lambda)^{-d} \left(e^{-i\lambda(t+h)} - \sum_{j=0}^{d-1} \frac{(-i\lambda(t+h))^j}{j!} \right) \mathbb{E}[\mathbb{Z}_n \overline{\mathbb{Z}_w}](d\lambda).$$

Here $\mathbb{Z}_w$ denotes the orthogonal increments process, which can be expressed as

$$\mathbb{Z}_w(d\lambda) = \mathbb{Z}_u(d\lambda) + (i\lambda)^d \mathbb{Z}_n(d\lambda)$$

(this follows from $w(t) = u(t) + D^d n(t)$. Therefore

$$\mathbb{E}[\mathbb{Z}_u \overline{\mathbb{Z}_w}](d\lambda) = f_u(\lambda) d\lambda \qquad \mathbb{E}[\mathbb{Z}_n \overline{\mathbb{Z}_w}](d\lambda) = (-i\lambda)^d f_n(\lambda) d\lambda.$$

Substituting into (19) yields zero after simplification. This holds for all $t, h \in \mathbb{R}$, which proves that $\Psi(L)$ is MSE optimal. Note that in order for the various integrals in (19) to exist, we require that $f_n(\lambda) f_u(\lambda)/f_w(\lambda)$ be integrable when multiplied by the polynomial λ^{d-1}. But since

$$\frac{f_n(\lambda) f_u(\lambda)}{f_w(\lambda)} = \frac{1}{1/f_n(\lambda) + \lambda^{2d}/f_u(\lambda)},$$

the integrability conditions hold so long as f_n and f_u are bounded. Finally, we compute the error spectral density:

$$\begin{aligned}\mathbb{E}[e(t)e(t+h)] &= \int_{-\infty}^{\infty} |(i\lambda)|^{-2d} \left| \frac{-\lambda^{2d} f_n(\lambda)}{f_w(\lambda)} \right|^2 e^{-i\lambda h} f_u(\lambda) d\lambda \\ &+ \int_{-\infty}^{\infty} \left| \frac{f_u(\lambda)}{f_w(\lambda)} \right|^2 e^{-i\lambda h} f_n(\lambda) d\lambda \\ &= \int_{-\infty}^{\infty} \frac{f_n(\lambda) f_u(\lambda)}{f_w(\lambda)} e^{-i\lambda h} \, d\lambda,\end{aligned}$$

which is integrable under the same boundedness conditions on f_n and f_u. Thus the error spectral density is $f_n f_u / f_w$, which completes the proof. □

Derivation of the Weighting Kernel in Illustration 2. We compute the Fourier Transform via the Cauchy Integral Formula (Ahlfors, 1979), letting $q = 1$ for simplicity:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{1+\lambda^4} e^{-i\lambda x} d\lambda$$

We can replace x by $|x|$ because the integrand is even. The standard approach is to compute the integral of the complex function

$$f(z) = \frac{e^{iz|x|}}{1+z^4}$$

along the real axis by computing the sum of the residues in the upper half plane, and multiplying by $2\pi i$ (since f is bounded and integrable in the upper half plane). It has two simple poles there: $e^{i\pi/4}$ and $e^{i3\pi/4}$. The residues work out to be

$$\begin{aligned}(z - e^{i\pi/4}) f(z)|_{e^{i\pi/4}} &= \frac{e^{-|x|(1-i)/\sqrt{2}}}{4i(1+i)/\sqrt{2}} \\ (z - e^{i3\pi/4}) f(z)|_{e^{i3\pi/4}} &= \frac{e^{-|x|(1+i)/\sqrt{2}}}{4i(1-i)/\sqrt{2}}\end{aligned}$$

respectively. Summing these and multiplying by i gives the desired result, after some simplification. To extend beyond the $q = 1$ case, simply let $x \mapsto q^{1/4} x$ and multiply by $q^{1/4}$ by change of variable. □

References

[1] Ahlfors, L. (1979) *Complex Analysis.* New York, New York: McGraw-Hill.

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

[3] Bergstrom, A. R. (1988) The History of Continuous-Time Econometric Models. *Econometric Theory* **4**, 365-383.

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

[5] Brockwell, P. (2001) Lévy-Driven CARMA Processes. *Annals of the Institute of Statistical Mathematics* **53**, 113–124.

[6] Brockwell, P. and Marquardt, T. (2005) Lévy-Driven and Fractionally Integrated ARMA Processes with Continuous Time Parameter. *Statistica Sinica* **15**, 477-494.

[7] Chambers, M. and McGarry, J. (2002) Modeling Cyclical Behavior With Differential-Difference Equations in an Unobserved Components Framework. *Econometric Theory* **18**, 387–419.

[8] Folland, G. (1995) *Introduction to Partial Differential Equations.* Princeton: Princeton University Press.

[9] Hannan, E. (1970) *Multiple Time Series.* New York, New York: Wiley.

[10] Harvey, A. and Stock, J. (1985) The Estimation of Higher-Order Continuous-Time Autoregressive Models. *Econometric Theory* **1**, 97-117.

[11] Harvey, A. and Stock, J. (1993) Estimation, Smoothing, Interpolation, and Distribution for Structural Time-Series Models in Continuous Time. In P.C.B. Phillips (ed.), *Models, Methods and Applications of Econometrics*, 55-70.

[12] Harvey, A. and Trimbur, T. (2003) General Model-Based Filters for Extracting Cycles and Trends in Economic Time Series. *Review of Economics and Statistics* **85**, 244-255.

[13] Hodrick, R., and Prescott, E. (1997) Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit, and Banking* **29**, 1 - 16.

[14] James, R., and Belz, M. (1936) On a Mixed Difference and Differential Equation. *Econometrica* **4**, 157-160.

[15] Jones, R. (1981) Fitting a Continuous-Time Autoregression to Discrete Data. In D.F. Findley (ed.), *Applied Time Series Analysis*, 651-674. New York: Academic Press.

[16] Kailath, T., Sayed, A., and Hassibi, B. (2000) *Linear Estimation.* Upper Saddle River, New Jersey: Prentice Hall.

[17] Kalecki, M. (1935) A Macrodynamic Theory of Business Cycles. *Econometrica* **3**, 327-344.

[18] Koopmans, L. (1974) *The Spectral Analysis of Time Series.* New York, New York: Academic Press, Inc.

[19] McElroy, T., and Trimbur, T. (2006) Continuous-lag Filters and Coherent Signal Extraction for Time Series Observed at Diverse Sampling Frequencies. *mimeo*

[20] Priestley, M. (1981) *Spectral Analysis and Time Series.* London: Academic Press.

[21] Stock, J. (1987) Measuring Business Cycle Time. *The Journal of Political Economy* **95**, 1240-1261.

[22] Stock, J. (1988) Estimating Continuous-Time Processes Subject to Time Deformation: An Application to Postwar U.S. GNP. *Journal of the American Statistical Association* **83**, 77-85.

[23] Whittle, P. (1983) *Prediction and Regulation.* Oxford: Blackwell Publishers.

Figure 1: Spectral Density for Cycle ($\rho = .2$, $\lambda_c = \pi/3$, $\sigma^2 = 1$)

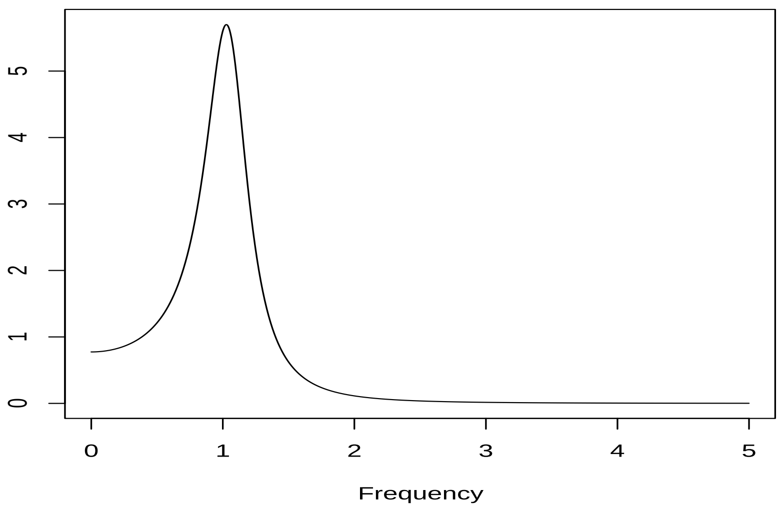

Figure 2: Smooth Trend WK Weighting Kernels

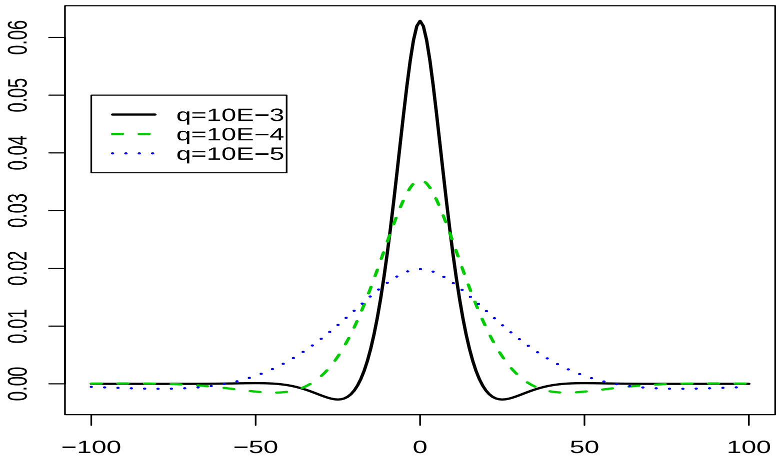

Figure 3: Low-Pass and Band-Pass Gain Functions ($m = 1, q_\zeta = .25, q_\kappa = 4$)

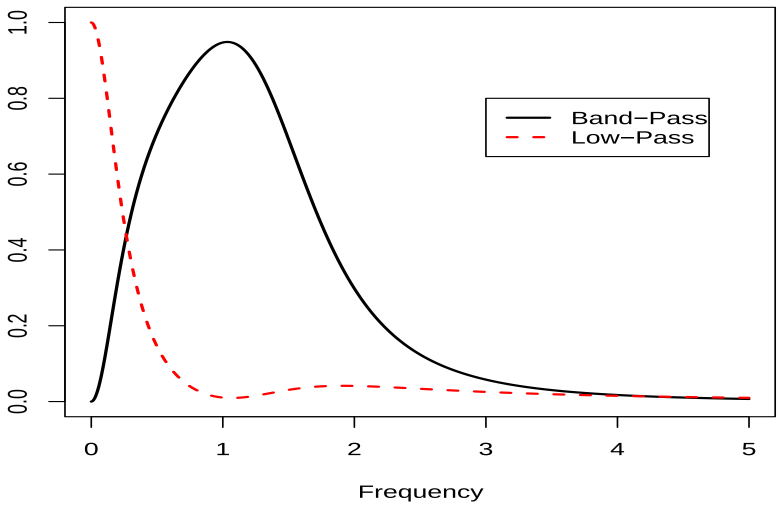

Figure 4: Low-Pass and Band-Pass Gain Functions ($m = 2$,$q_\zeta = .04$,$q_\kappa = 4$)

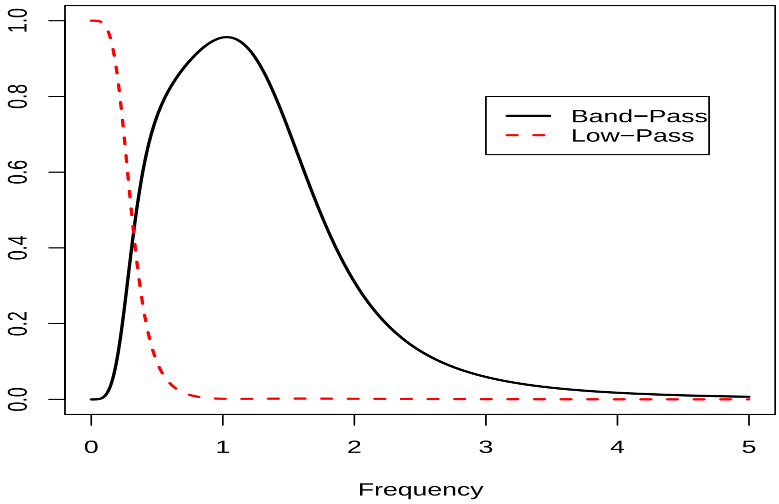

Figure 5: Smooth Trend Velocity WK Weighting Kernels

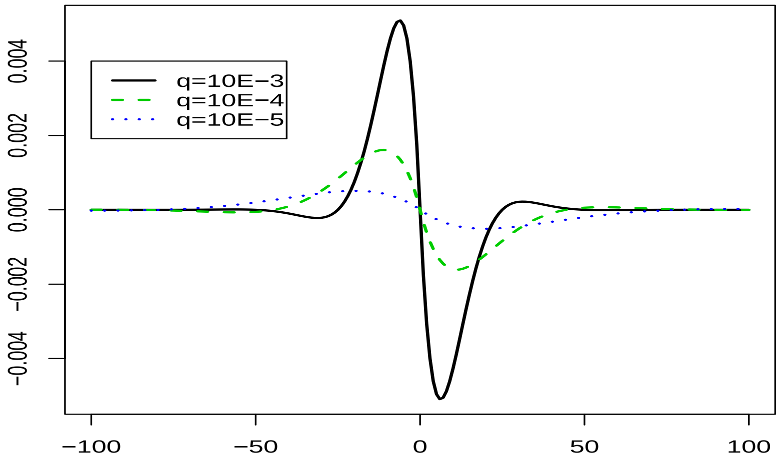

Figure 6: Smooth Trend Acceleration WK Weighting Kernels