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

Tucker S. McElroy and Thomas M. Trimbur

2007-68

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

Tucker S. McElroy, U.S. Census Bureau, Washington DC

Thomas M. Trimbur,* Federal Reserve Board, Washington DC

## 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:** Continuous Time Processes, Cycles, Hodrick-Prescott Filter, Linear Filtering, Signal Extraction, Turning points

**Disclaimer:** This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau or of the Federal Reserve Board.

*Corresponding author. Address: Division of Research and Statistics; 20th and C Street, NW; Federal Reserve Board; Washington, DC 20551. Email: Thomas.M.Trimbur@frb.gov.

# 1 Introduction

This paper concentrates on signal extraction in continuous-time. The goal is to set the stage for the development of coherent discrete filters in various applications. Thus, we start by setting up a fundamental signal extraction problem and by investigating the properties of the optimal *continuous-lag filters*. This part of the frequency and time domain analysis is done independently of sampling type and interval and so may be seen as fundamental to the dynamics of the problem.

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 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). Whittle (1983) also provides results for the nonstationary case, but omits proof, and in particular, fails to consider the initial value assumptions that are central to the problem.

In this paper, we extend the proof to the case of a nonstationary signal, that is, integrated of order $d$, where $d > 0$ is an integer. We also treat the case of a white noise irregular, which is frequently used in standard models in continuous-time econometrics. Since continuous-time white noise is essentially the first derivative of Brownian motion (which is nowhere differential), this requires a careful mathematical treatment to ensure that the signal extraction problem is well-defined.

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)

2

and Harvey and Stock (1993).

The study of business cycles has remained of interest to researchers and policy-makers 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, we derive analytical expressions for the spectral density; this gives a clear summary of 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.

In focussing on continuous-time foundations in this paper, we also note the clear practical motivations for this strategy. 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. See Bergstrom (1988, 1990) for further discussions of the advantages of continuous-time analysis.

The practical application of the continuous-time strategy is sketched at the end of this paper, and is set out in greater detail in a companion paper (McElroy and Trimbur, 2007); therein we examine how the optimal continuous-lag filter may be discretized to yield expressions for the discrete-time weights appropriate for data sampled under various conditions. The method is illustrated with a number of examples, including the continuous-time HP filter. In recent work, Ravn and Uhlig (2002), Maravall and del Rio (2001), and Harvey and Trimbur (2007) have investi-

gated how to adapt the HP filter to monthly and annual series, given that the filter was originally designed for quarterly US GDP. We show how the continuous-time analogue of the HP filter is discretized to yield a set of consistent discrete filters, thereby solving the problem of adapting the filter.

Most previous methods rely on the *discretization* of the underlying continuous-time model, so that the analysis is done in discrete-time. Our approach instead uses the continuous-time formulation of filtering more directly. Thus, the method centers on the use of the underlying *continuous-lag filters*, on their properties, and on the transformations needed for discrete datasets.

The theoretical results derived here set the foundation for broader applications. For instance, a new class of low-pass and band-pass filters are presented in Section 4 of this paper. Further, in considering the signal extraction problem in continuous-time, we derive filters to measure velocity and acceleration of a time series, which could be useful for the analysis of turning points.

The rest of the paper is organized as follows. Section 2 reviews continuous-time filtering, based on material from Priestley (1981), Hannan (1970), and Koopmans (1974). Section 3 sets out the signal extraction framework. In section 4, examples are given for economic series; an extension of the standard HP filter is derived from the smooth trend model, and a general class of band-pass filters is presented. Extensions to methodology are then described, specifically, the conversion of continuous-lag filters to estimate growth rates and other characteristics of an underlying component. Section 5 discusses the application of the method to real series, and Section 6 concludes. 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 $\psi$.

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** $\psi$ (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 $\psi$ is differentiable to a specified order, with integrable or square integrable derivatives.

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 $\psi$. 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 $\Psi(e^{-i\lambda}) = \mathcal{F}[\psi](\lambda)$.

**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

6

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

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

As in discrete time series signal processing, passing an input (stationary) 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 contributions to the variance at various frequencies 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.

**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, one can show using integral calculus that $\Psi(L) = 1/(1 - (\log L)^2)$, as a formal expression. 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 mean-square differentiation operator $D$, consider the limit of measuring the displacement of a continuous-time process, per unit of time, over an arbitrarily small interval $\delta$:

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

The limits are interpreted to converge in mean square. 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 mean-sqaure 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 $\psi$, 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

8

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 (formally)

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

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 $\psi$'s Laplace transform is a rational function.

**Example 3: Higher order stochastic cycles** We consider a general class of continuous-time stochastic cycles. These are indexed by a positive integer $n$ that denotes the order of the model.

Denote the cyclical process by $\psi_n(t)$, and let $\psi_n^*(t)$ represent an auxiliary process used in the construction of the model. Define $\boldsymbol{\psi}_n(t) = (\psi_n(t)\ \psi_n^*(t))'$. An $n-$th order stochastic cycle in continuous-time is given by

$$d\boldsymbol{\psi}_i(t)dt = \mathbf{A}\boldsymbol{\psi}_i(t)dt + \begin{bmatrix} \psi_{i-1}(t) \\ 0 \end{bmatrix} dt, \quad i = 2, ..., n \tag{13}$$

$$d\boldsymbol{\psi}_1(t)dt = \mathbf{A}\boldsymbol{\psi}_1(t)dt + \begin{bmatrix} \kappa(t) \\ 0 \end{bmatrix} dt, \quad \kappa(t) \sim WN\left(0, \sigma_\kappa^2\right)$$

9

where $\boldsymbol{\psi}_i(t) = (\psi_i(t)\ \psi_i^*(t))'$, $i = 1, ..., n-1$ represent additional auxiliary processes. The coefficient matrix in (13) is

$$\mathbf{A} = \begin{bmatrix} \log \rho & \lambda_c \\ -\lambda_c & \log \rho \end{bmatrix}$$

The parameter $\rho$ is called the damping factor; it satisfies $0 < \rho \le 1$. The stochastic variation in the cycle per unit time depends on the continuous-time variance parameter $\sigma_\kappa^2$. The parameter $\rho$ controls the persistence of cyclical fluctuations, and its specific role depends on $n$. Generally, for higher orders, the model generates smoother dynamics for the cycle.

Since the parameter $\lambda_c$ corresponds roughly to a peak in the spectrum, it indicates a central frequency of the cycle, and $2\pi/\lambda_c$ is an average period of oscillation. As $\lambda_c$ is a frequency in continuous-time, it can be any positive real number, though for macroeconomic data, business cycle theory will usually suggest a value in some intermediate range.

To construct a Gaussian cyclical process, the increment $\kappa(t)$ can be derived from Brownian motion $W_\kappa(t)$, that is, $\kappa(t) \sim DW_\kappa(t)$.

In analyzing cyclical behavior, it is natural to consider the frequency domain properties. To derive the spectra for various $n$, start by rewriting (13) as a recursive formula:

$$\begin{bmatrix} D - \log \rho & -\lambda_c \\ \lambda_c & D - \log \rho \end{bmatrix} \begin{bmatrix} \psi_i(t) \\ \psi_i^*(t) \end{bmatrix} = \begin{bmatrix} \psi_{i-1}(t) \\ 0 \end{bmatrix}, \quad i = 1, ..., n$$

with the initialization $\psi_0(t) = \kappa(t)$. The solution is

$$((D - \log \rho)^2 + \lambda_c^2)^n \psi_n(t) = (D - \log \rho)^n \kappa(t)$$

so that the $n$th order cycle has a $CARMA(2n, n)$ form. The power spectrum is

$$\gamma_{\psi_n}(\lambda) = \sigma_\kappa^2 \left[ \frac{\lambda^2 + \log^2 \rho}{\left(\log^2 \rho + (\lambda - \lambda_c)^2\right)\left(\log^2 \rho + (\lambda + \lambda_c)^2\right)} \right]^n \tag{14}$$

One advantage of working with the cycle in continuous-time is the possibility of analyzing turning points instantaneously. The expected incremental change in the

10

cycle for $n \geq 2$ is

$$d\psi_n(t) = (\log \rho)\psi_n(t)dt + \lambda_c \psi_n^*(t)dt + \psi_{n-1}(t)dt \tag{15}$$

For $n = 1$, this reduces to

$$d\psi_1(t) = (\log \rho)\psi_1(t)dt + \lambda_c \psi_1^*(t)dt \tag{16}$$

Based on the smoothness of estimated cycles, the higher order models are likely to give more reliable indication of turning points.

Expression (16) is similar to the one in Harvey (1989, p. 487). There is, however, a slight difference because the form in (13) is analogous to the 'Butterworth form' used in Harvey and Trimbur (2003). The alternative form in Harvey (1989) and in Harvey and Stock (1985) is analogous to the 'balanced form' also considered in Harvey and Trimbur (2003). The key advantage of the Butterworth form, as used here, is its convenience for the analysis of spectra and gain functions.

We have shown that the class of models in (13) are equivalent to $CARMA$ processes whose parameters satisfy the conditions needed for periodic behavior. As an alternative, an openly specified model can be used to describe a general pattern of serial correlation; thus, a $CAR(1)$ could be used to capture first-order autocorrelation in the noise component, for instance, due to temporary effects of weather. When there is clear indication of cyclical dynamics, however, the models in (13) give a more direct analysis that is easier to interpret.

The properties of a class of stochastic cycles[1] are set out in Trimbur (2006) for the discrete-time case. The properties of the continuous-time models give a similar flexibility in describing periodic behavior.

Figure 1 shows the spectrum for parameter values $\lambda_c = \pi/4$. In particular, the function $(2\pi\sigma_\psi^2)^{-1}\gamma_{\psi_n}(\lambda;\rho, \lambda_c, n)$ is plotted, where the normalizing constant includes the unconditional variance $\sigma_\psi^2$ of the cycle; this is computed in Mathematica by numerical integration of the power spectrum for given values of $\rho$ and $n$. For

---

[1]Note that these models have a different form from the models that would result from the exact discretization of (13). In general, starting with a continuous-time model with uncorrelated components leads to a discretized model that has either correlated components or MA disturbances; one expects, however, that the basic structure of the discrete-time models should remain unchanged.
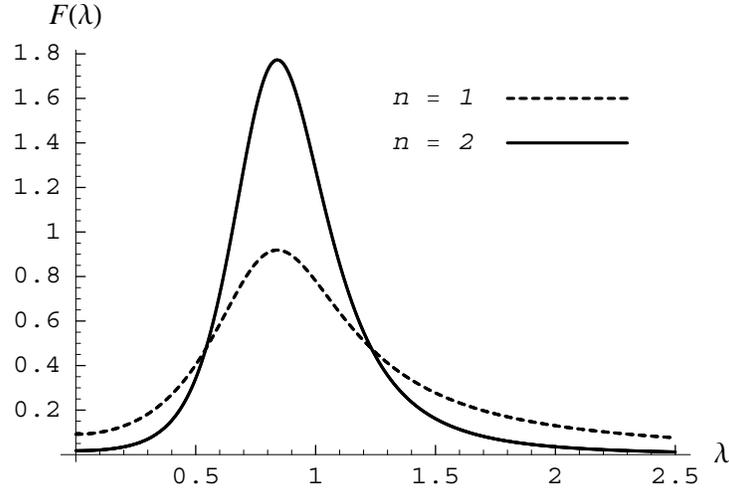
Figure 1: Spectral density of continuous-time cyclical process for $n = 1$ and 2 with $\rho = 0.7$ and $\lambda_c = \pi/4$.

$n = 1$, the damping factor is set to $\rho = 0.9$, and for $n = 2$, it is set to $\rho = 0.75$. Lower values of $\rho$ are appropriate for the higher order models because of their resonance property. Thus, the cyclical shocks reinforce the periodicity in $\psi_n(t)$, making the oscillations more persistent for given $\rho$. The difference in spectra in figure 1 indicates that the periodicity is more clearly defined for the second order cycle.

The spectrum peaks at a period around $2\pi/\lambda_c$ for moderate values of $\lambda_c$, say $\pi/5 < \lambda_c < \pi$, which is the standard business cycle range. The maximum does not occur exactly at $\lambda = \lambda_c$, however, except as $\rho$ tends to unity. The case $\rho = 1$ gives a nonstationary cycle, where one could, in theory, forecast out to unlimited horizons. Thus, in economic modeling, attention is usually restricted to stationary models where $\rho < 1$.

# 3  Signal Extraction in Continuous Time

This section develops the signal extraction problem in continuous time. A new result with proof is given for estimating a nonstationary signal from stationary noise. Whittle (1983) shows a similar result for nonstationary processes, but omits the proof and in particular, fails to recognize the importance of initial conditions. 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 initial value assumptions to the result. Further, the cases where the differentiated signal or noise process or both are $WN$ are treated rigorously.

## 3.1  Nonstationary signal and initial conditions

Consider the following model for a continuous time process $y(t)$:

$$y(t) = s(t) + n(t), \quad t \in \mathbb{R} \tag{17}$$

where $n(t)$ is stationary. The aim is to estimate the underlying signal $s(t)$ in the presence of the noise, and it will be assumed that $s(t) \sim I(d)$, or integrated of order $d$.

In general, $d$ is any non-negative integer; the special case $d = 0$ reduces to stationary $s(t)$. In many applications of interest, we have $d > 0$, so that the $d$th derivative of $s(t)$, denoted by $u(t)$, is stationary. It is assumed that $u(t)$ and $n(t)$ are mean zero and uncorrelated with one another. In the standard case, both autocovariance functions, $R_u$ and $R_n$, are integrable. An extension could also be considered where $R_u$ or $R_n$ or both are represented by a multiple of the Dirac delta function, which gives rise to tempered distributions (see Folland, 1995); the associated spectral densities are flat, indicating a corresponding $WN$ process.

The process $y(t)$ satisfies the stochastic differential equation

$$w(t) = D^d y(t) = u(t) + D^d n(t). \tag{18}$$

From Section 2.2, the spectral density of $w(t)$ is

$$f_w(\lambda) = f_u(\lambda) + \lambda^{2d} f_n(\lambda). \tag{19}$$

13

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

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

for some initial value random variable $y(0)$. Note that this remains valid both for $t > 0$ and for $t < 0$. When $d = 2$,

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

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

$$y(t) = \sum_{j=0}^{d-1} \frac{t^j}{j!} y^{(j)}(0) + [I^d w](t) \tag{20}$$

with the $I$ operator defined by $[I^d w](t) = \int_0^t w(z)(t-z)^{d-1}\, dz/(d-1)!$. Note that (20) holds for the signal $s(t)$ as well.

For an $I(d)$ process, let $\mathbf{y}^*(\mathbf{0}) = \{y(0), \dot{y}(0), \cdots, y^{(d-1)}(0)\}$ denote the collection of $d$ values and higher order derivatives at time $t = 0$. It is assumed that $\mathbf{y}^*(\mathbf{0})$ is uncorrelated with both $u(t)$ and $n(t)$ for all $t$. This assumption is analogous to Assumption A in Bell (1984), except that now higher order derivatives are involved.

## 3.2   Formula for the optimal filter

Consider the theoretical signal extraction problem for a bi-infinite series $y(t)$ that follows (17). The optimal *linear* estimator of the signal $s(t)$ gives the minimum mean square error. Thus, the goal is to minimize $\mathbb{E}[(\hat{s}(t) - s(t))^2]$ such that $\hat{s}(t) = \Psi(L)y(t) = (\psi * y)(t)$ for some weighting kernel $\psi$. The notation $\Psi(L)$ for a continuous-lag filter was introduced earlier. The problem is to determine the optimal choice of $\Psi(L)$ for general nonstationary models of the form (17). The following theorem shows the main result.

**Theorem 1** *For the process in (17), suppose that $\mathbf{y}^*(\mathbf{0})$ is uncorrelated with both $u(t)$ and $n(t)$ for all $t$. Also, assume that $u(t)$ and $n(t)$ are mean zero weakly stationary processes that are uncorrelated with one another, with autocovariance*

14

*functions that are either integrable or given by constant multiples of the Dirac delta function, interpreted as a tempered distribution. 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 estimate of $s(t)$ is given by*

$$\hat{s}(t) = \Psi(L)y(t) \tag{21}$$

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

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

*The function $\psi(x)$ is the continuous weighting kernel of the optimal filter. 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 $\frac{1}{2\pi} \int_{-\infty}^{\infty} f_e(\lambda)d\lambda$.*

If $y(t)$ is Gaussian, then $\hat{s}(t)$ is optimal among all estimators. The filter $\Psi(L)$ will be referred to as a continuous-lag Wiener-Kolmogorov ($WK$) filter. This distinguishes $\Psi(L)$ from discrete-time model-based filters, which are only defined over a discrete set of lags. In contrast, here we focus on the model-based filters derived in continuous-time.

One of the important properties of the $WK$ filters is that they pass, or preserve, polynomials, in analogy to discrete-lag filters constructed to have this property in discrete-time. In particular,

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

for a polynomial $p(t)$ of sufficiently low degree. To make this explicit, the filter passes $p(t)$ when

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

15

for any $j$ up to the degree of $p$, with $\delta$ denoting the Kronecker delta. It is shown in the proof of Theorem 1 that, provided that the associated moments exist, a $WK$ filter passes polynomials of degree up to $2d - 1$.

Note that the noise and differentiated signal can be either $WN$ or can have integrable autocovariance functions. The signal extraction problem for different cases determines different classes of weighting kernels. We can now define continuous-lag filters that reflect the nonstationary component of a time series.

In particular, the nonstationarity means that the signal includes a stochastic trend and so is represented by an integrated process. First, we show a simple example of the case $d = 0$; this reduces to stationarity, so the only requirement on $g$ is continuity.

**Example 4** For $d = 0$, let $s(t)$ have autocovariance function $R_s(h)$ denoted by $\phi(h) = \frac{1}{\sqrt{2\pi}} e^{-\frac{h^2}{2}}$. 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 continuous, and from Example 2 the inverse Fourier Transform gives a simple filter with kernel $\psi(x) = \frac{1}{2} e^{-|x|}$.

**Example 5** Consider now the case $d = 1$, and suppose that the spectral density of the differentiated signal is $f_u(\lambda) = q\phi(\lambda)$, where $\phi$ has the form of the standard normal density function, and that $f_u(\lambda)/f_n(\lambda) = q$ for some constant $q$. The signal extraction filter has a continuous-time frequency response given by

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

which yields a double-exponential weighting kernel $\sqrt{q} \exp\{-\sqrt{q}|x|\}/2$. This kernel passes lines and constants and could be used as a simple device for trend smoothing.

16

# 4    Illustrations of Continuous-Lag Filtering

In this section, examples of continuous-lag $WK$ filters are given for economic time series. The filters are based on the class of $CARIMA$ models; this class is particularly convenient for computing $WK$ weighting kernels and offers flexibility for a range of applications. The spectral densities $f_u$ and $f_w$ that enter the formula for the gain are both rational functions in $\lambda^2$. Taking their ratio yields another rational function in $\lambda^2$ for $g(\lambda)$. As these analytical expressions summarize the comprehensive effects of the filter, they can be studied and used in filter design in different contexts.

We focus on examples where $CARIMA$ models are set up within different signal extraction problems. The specifications are guided by applications of interest in economics; their solutions rely on the theorem given in the last Section for handling nonstationary series. In the first example, we start with the simplest case, the local level model. The second example considers an extension of the well-known HP filter, which has been widely used in macroeconomics as a detrending method. We show that the expression for the continuous-lag HP filter is relatively simple, and this gives a basis for detrending data with different sampling conditions.

The third example extends the treatment with a derivation of continuous-lag low-pass and band-pass Butterworth filters. This class of filters represents the analogue of the discrete-time filters introduced in Harvey and Trimbur (2003). The band-pass filters arise naturally as cycle estimators in a well-defined model that jointly describes trend, cyclical, and noisy movements. The general cyclical processes in continuous-time are defined in an analogous way to the discrete-time models studied in Trimbur (2006). Note that the cyclical components are equivalent to certain $CARMA$ models. In the analysis of periodic behavior, it is more direct to work with the structural form; the frequency parameter, for instance, reflects the average, or central, periodicity.

The low-pass and band-pass filters we present have the property of mutual consistency. That is, they may be applied simultaneously. Other procedures, in contrast, do not preserve this property when the two filters are designed separately, or when they are based on different source models.

It is generally straightforward to investigate the weighting kernels of the $WK$ filters. This involves calculating the residues of rational functions, a standard problem for which well-known procedures are available. Still, the computations can become burdensome in particular cases, so in presenting illustrations, we restrict attention to some standard filters. In these cases, the derivation of analytical results is feasible, with the expressions simple enough to provide a clear interpretation.

In the framework of $WK$ filters, source models can be formulated to adapt to different situations. For instance, some series of Industrial Output are subject to weather effects that induce short-lived serial correlation. In estimating the trend, the base model can be set up with a low-order CAR or CARMA component.

The approach to filter design can also be adapted to focus on certain properties of the signal, such as rate of change. Thus, in a more general framework, our target of estimation becomes a functional of the signal. This opens the door to a number of potential applications, such as turning point analysis, where the interest centers on some aspect of the signal's evolution over an interval. After describing the basic principle, in a fourth example, we examine an application to measuring velocity and acceleration of signal.

**Illustration 1: Local Level Model** The trend plus noise model 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 $\sigma_\epsilon^2$, denoted by $\epsilon(t) \sim WN(0, \sigma_\epsilon^2)$. See Harvey (1989) for discussion. An interpretation of the variance $\sigma_\epsilon^2$ is that $\Theta(L)\epsilon(t)$ has autocovariance function $(\theta * \theta^-)(h)\sigma_\epsilon^2$ for any (integrable) auxiliary weighting kernel $\theta$.

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

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

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 $(1 + \lambda^2/q)^{-1}$, 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:

$$D\mu(t) = \beta(t) + \eta(t), \qquad \eta(t) \sim WN(0, \sigma_\eta^2)$$
$$D\beta(t) = \zeta(t), \qquad \zeta(t) \sim WN(0, \sigma_\zeta^2)$$

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 $q = \sigma_\zeta^2/\sigma_\epsilon^2$.
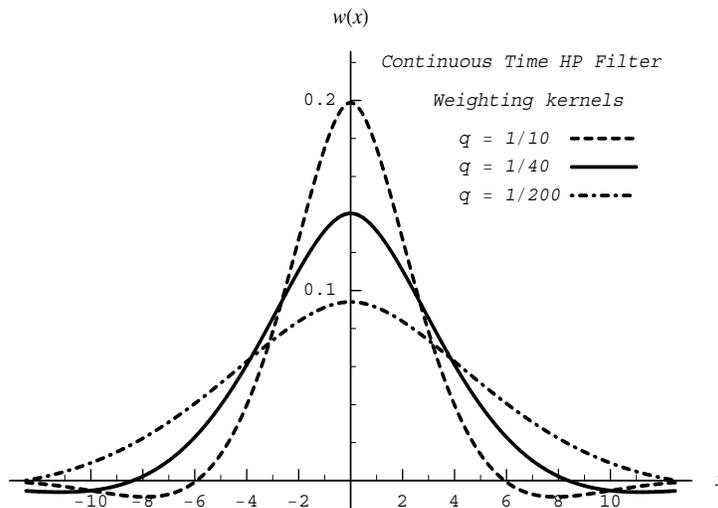


Figure 2: Weighting kernel for continuous-lag HP filter for $q = 1/10, 1/40$, and $1/200$.

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

$$u(t) = D^2\mu(t) = \zeta(t)$$
$$w(t) = D^2y(t) = \zeta(t) + D^2\epsilon(t).$$

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

$$\psi(x) = \frac{q^{1/4}\exp\{-|x|q^{1/4}/\sqrt{2}\}}{2\sqrt{2}}\left(\cos(|x|q^{1/4}/\sqrt{2}) + \sin(|x|q^{1/4}/\sqrt{2})\right) \qquad (22)$$

This gives the continuous-time extension of the HP filter. From the discussion following Theorem 1, the kernel in (22) passes cubics.

Figure 2 shows the weighting function for three different values of $q$. As the signal-noise ratio increases, the trend becomes more variable relative to noise, so the resulting kernel places more emphasis on nearby observations. Similarly, as $q$ decreases, the filter adapts by smoothing over a wider range. The negative side-lobes, apparent in the figure for $q = 1/10$, enable 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

$$y(t) = \mu_m(t) + \psi_n(t) + \varepsilon(t)$$

where $\mu_m(t)$ is a trend component that accounts for long-term movements and the cyclical component $\psi_n(t)$ follows (13) for index $n$. The irregular $\varepsilon(t)$ is meant to

20

absorb any random, or nonsystematic variation, and in direct analogy with discrete-time, it is assumed that in continuous-time, $\varepsilon(t) \sim WN(0, \sigma_\varepsilon^2)$. The definition of the $m-$th order trend is

$$D^m \mu_m(t) = \zeta(t), \quad \zeta(t) \sim WN(0, \sigma_\zeta^2)$$

for integer $m > 0$. For $m = 1$, this gives standard Brownian motion. For $m = 2$, $\mu_m(t)$ is integrated Brownian motion, the continuous-time analogue of the smooth trend, as in the previous illustration.

In formulating the estimation of $\psi_n(t)$ as a signal extraction problem, we set the nonstationary signal to $\mu_m(t) + \varepsilon(t)$ and the 'noise' to $\psi_n(t)$. 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 'signal' as a target of extraction. Actually in this case, the 'noise' part will usually be of greatest interest in a business cycle analysis. Thus, the optimal filter $F(L)$ is constructed for $\mu_m(t) + \varepsilon(t)$, and the complement of this filter, $(1 - F(L))$, yields the band-pass. Similarly, to formulate the estimation of $\mu_m(t)$, take $s(t) = \mu_m(t)$ and $n(t) = \psi_n(t) + \varepsilon(t)$.

Thus, the class of continuous-lag Butterworth filters are given by

$$LP_{m,n}(\lambda) = \frac{\sigma_\zeta^2/\lambda^{2m}}{\frac{\sigma_\zeta^2}{\lambda^{2m}} + \sigma_\kappa^2 \left[\frac{\lambda^2 + \log^2 \rho}{\left(\log^2 \rho + (\lambda - \lambda_c)^2\right)\left(\log^2 \rho + (\lambda + \lambda_c)^2\right)}\right]^n + \sigma_\varepsilon^2},$$

$$BP_{m,n}(\lambda) = \frac{\sigma_\kappa^2 \left[\frac{\lambda^2 + \log^2 \rho}{\left(\log^2 \rho + (\lambda - \lambda_c)^2\right)\left(\log^2 \rho + (\lambda + \lambda_c)^2\right)}\right]^n}{\frac{\sigma_\zeta^2}{\lambda^{2m}} + \sigma_\kappa^2 \left[\frac{\lambda^2 + \log^2 \rho}{\left(\log^2 \rho + (\lambda - \lambda_c)^2\right)\left(\log^2 \rho + (\lambda + \lambda_c)^2\right)}\right]^n + \sigma_\varepsilon^2},$$

where $LP_{m,n}(\lambda)$ stands for low-pass filter and $BP_{m,n}(\lambda)$ stands for band-pass filter, both of order pair $(m, n)$. These expressions follow from combining the power spectrum of the cycle with the pseudo-spectrum of $\mu(t)$. Defining $q_\zeta = \sigma_\zeta^2/\sigma_\varepsilon^2$ as the signal-noise ratio for the trend and $q_\kappa = \sigma_\kappa^2/\sigma_\varepsilon^2$ as the signal-noise ratio for the cycle, it follows that

$$LP_{m,n}(\lambda; q_\zeta, q_\kappa) = \frac{q_\zeta}{q_\zeta + q_\kappa \lambda^{2m} \left[ \frac{\lambda^2 + \log^2 \rho}{\left(\log^2 \rho + (\lambda - \lambda_c)^2\right)\left(\log^2 \rho + (\lambda + \lambda_c)^2\right)} \right]^n + \lambda^{2m}}, \tag{23}$$

$$BP_{m,n}(\lambda; q_\zeta, q_\kappa) = \frac{q_\kappa \lambda^{2m} \left[ \frac{\lambda^2 + \log^2 \rho}{\left(\log^2 \rho + (\lambda - \lambda_c)^2\right)\left(\log^2 \rho + (\lambda + \lambda_c)^2\right)} \right]^n}{q_\zeta + \sigma_\kappa^2 \left[ \frac{\lambda^2 + \log^2 \rho}{\left(\log^2 \rho + (\lambda - \lambda_c)^2\right)\left(\log^2 \rho + (\lambda + \lambda_c)^2\right)} \right]^n + \lambda^{2m}}, \tag{24}$$

Here the order $m = d$ denotes the order of integration as determined by the stochastic trend model. The definitions in (23) and (24) parallel the development of Harvey and Trimbur (2003) for the discrete-time case. Note that in the continuous-time case, we must have integrability of the frequency response function over the entire real line rather than over a restricted interval.
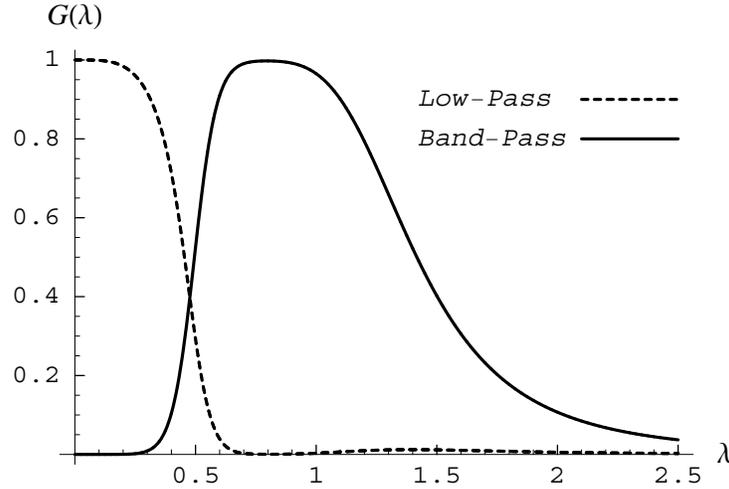


Figure 3: Gain functions of a pair of consistent low-pass and band-pass filters. The underlying model has $m = 2$, $n = 2$, with cyclical parameters $\rho = 0.9$, $\lambda_c = \pi/4$ and signal-noise ratios $q_\zeta = 0.1$, $q_\kappa = 1$.

Figure 3 illustrates the low-pass and band-pass gain functions for trend order $m = 2$ and cycle orders $n = 2$. The low-pass dips at intermediate frequencies to

22

accommodate the presence of the cycle in the model. Figure 4 shows a comparison of the band-pass filter for $n = 2$ and 4. The other parameters are the same as in the previous figure. Note the increased sharpness produced by the higher order model.
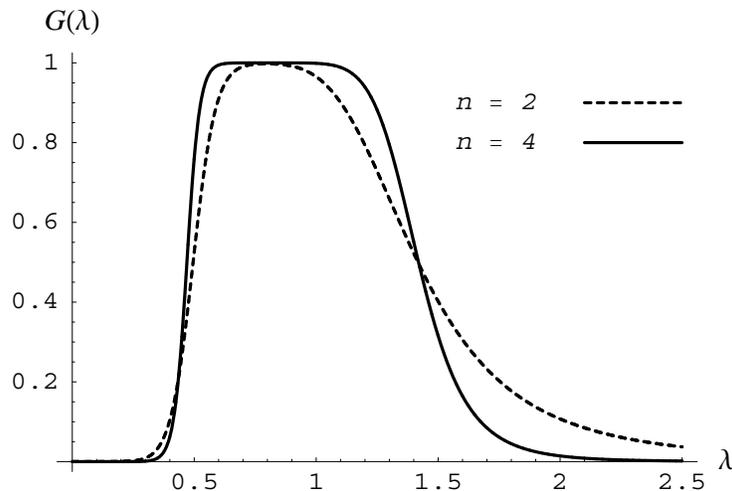


Figure 4: Gain function of band-pass filters for $n = 2$ and 4. The other parameters are the same as in the previous figure.

**Illustration 4: Smooth Trend Velocity and Acceleration** In some applications, interest centers on some property of the signal, such as its growth rate, rather than on the value of the signal itself. In particular, consider the linear operator $H = D^m$. The conditional expectation of $Hs(t)$ is equal to $H$ applied to the conditional expectation of $s(t)$, since $H$ is linear. So for Gaussian processes, assuming that $\lambda^m g(\lambda)$ is integrable – where $g$ is the frequency response of the original $WK$ filter – the weighting kernel for estimating $D^m s(t)$ is given by the $m$th derivative of $\psi$, that is, $\psi^{(m)}(x)$. The first derivative of the signal indicates a velocity, or growth rate. The second derivative indicates acceleration, or variation in growth rate.

More generally, we can consider signals $Hs(t)$ with $H$ a linear operator. To compute the mean squared error of $(H\Psi(L))y(t)$ as an estimate of the target $Hs(t)$,

23

multiply the error spectrum from Theorem 1 by the squared magnitude of $\mathcal{F}[H](\lambda)$. This results in the spectrum of the new error, whose integral equals the mean squared error. If, for example, $H = D$ then the error spectrum is multiplied by by $\lambda^2$, and the result is then integrated over $(-\infty, \infty)$.

Velocity and acceleration estimates can be computed for the HP filtered signal. The filters are constructed directly from the Smooth Trend model. In Newtonian mechanics, a local maximum in a particle's trajectory is indicated by zero velocity together with a negative acceleration; similarly, velocity and acceleration indicators may be used to discern a downturn or recession in a macroeconomic series.
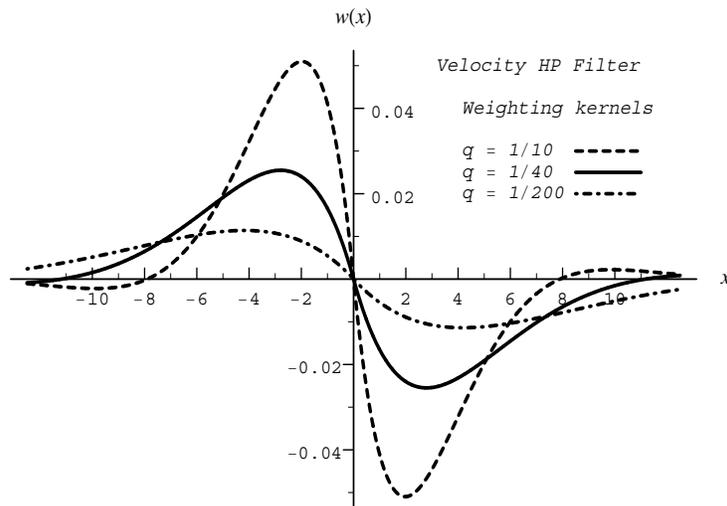


Figure 5: Weighting Kernels for Velocity $WK$ filter based on Smooth Trend model for $q = 1/10, 1/40$, and $1/200$.

Since $\lambda^2(1 + \lambda^4/q)^{-1}$ is integrable, both derivatives of $\psi$ are well-defined. Direct calculation yields

$$\dot{\psi}(x) = -\frac{q^{1/2}}{2}e^{-q^{1/4}|x|/\sqrt{2}}\sin(q^{1/4}x/\sqrt{2})$$

$$\ddot{\psi}(x) = -\frac{q^{3/4}}{2\sqrt{2}}e^{-q^{1/4}|x|/\sqrt{2}}\left(\cos(q^{1/4}|x|/\sqrt{2}) - \sin(q^{1/4}|x|/\sqrt{2})\right).$$
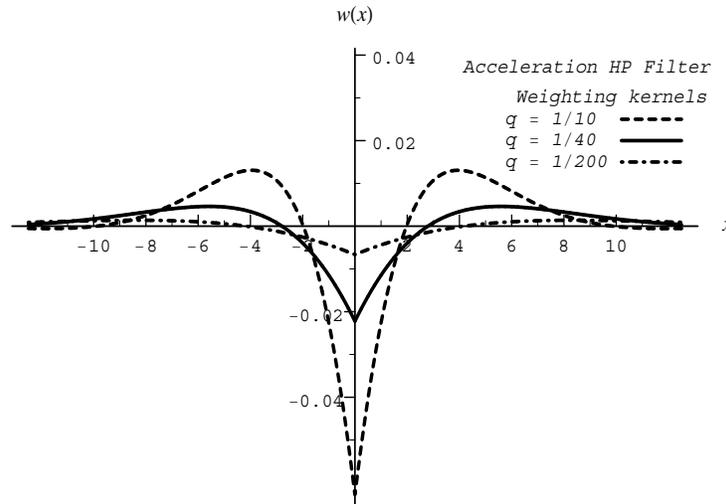
24

Figure 6: Weighting Kernels for Acceleration $WK$ filter based on Smooth Trend model for $q = 1/10, 1/40$, and $1/200$.

The velocity filter, or first derivative with respect to time, has the interpretation of a growth rate for the trend. The weighting kernel in Figure 5 shows how the growth in the signal is assessed by comparing forward-looking displacements with recent displacements. Likewise, the acceleration indicates the second derivative, or curvature. The weighting kernel in Figure 6 has a characteristic sharp decline around the origin, so that contemporaneous and nearby values are subtracted in estimating changes in growth.

# 5 Application: unequally sampled data

In this Section, we outline a procedure for the practical application of the method. The discussion of discretization is kept concise, as this material is set out in detail in McElroy and Trimbur (2007).

Starting with a base model incontinuous-time, the classification into stock and

flow is reflected in the measurement of observations. Denote the underlying process by $y(t)$. Given the sampling interval $\delta > 0$, a stock observation at the $\tau$th time point is defined as

$$y_\tau = y(\delta \tau). \tag{25}$$

The times of observations correspond to $t_\tau = \delta \tau$ for integer $\tau$. The discrete stock time series is then the sequence of values $\{y_\tau\}_{\tau=-\infty}^{\infty}$.

A series of flow observations has the form

$$y_\tau = \int_{\delta(\tau-1)}^{\delta\tau} y(v)\, dv. \tag{26}$$

where $\delta$ is both the interval of cumulation and the interval separating successive observation points. Note that, more generally, the times $...,t_1, t_2, \cdots, t_\tau,...$ need not be equally spaced, but for now we assume for simplicity that the spacing is constant at $t_\tau - t_{\tau-1} = \delta$.

For a finite sample, let $\mathbf{y} = (y_1, y_2, \cdots, y_T)'$ be the column vector of observations. Then for a Gaussian process $y(t)$, the law of iterated conditional expectations yields

$$\mathbb{E}[s(t)|\mathbf{y}] = \mathbb{E}\left[(\psi * y)(t)|\mathbf{y}\right] = \int \psi(t - z)\mathbb{E}[y(z)|\mathbf{y}]\, dz. \tag{27}$$

That is, our best estimate of the signal at any time $t$, given the observations $\mathbf{y}$, is a convolution of the weighting kernel $\psi$ with an interpolated and extrapolated estimates $\mathbb{E}[y(z)|\mathbf{y}]$. Thus the estimate of the signal is computed as follows:

1. Determine $\psi$ from a fitted continuous-time model.

2. Compute $\mathbb{E}[y(z)|\mathbf{y}]$ for a set of $z$ values on a fine mesh.

3. Compute a numerical approximation to the integral in (27) and in this way approximate the signal estimate.

Explicit worked examples are beyond the scope of this paper, but the above suggests how the continuous-lag filter could be directly applied for a general problem. Another approach, which we prefer, is to first discretize $\psi$ and then apply the appropriate discrete filter to the observed data $\mathbf{y}$, and this approach is set out in detail in McElroy and Trimbur (2007). For now, it should be emphasized that our

approach, based on a continuous-lag kernel, can handle signal estimation in a rather general context, that is, with an unequally spaced series of stock or flow data, and a with a signal time point lying in between observation times. This generality of the signal extraction problem cannot be achieved in a purely discrete-time setting and could only possibly be achieved, with great difficulty, in an approach requiring full model discretization.

# 6   Conclusion

This paper has solved the problem of establishing a theoretical foundation for continuous-time signal extraction with nonstationary models. Economic series commonly show some kind of stochastic trend. The rigorous treatment of nonstationarity given in this paper thus paves the way for the design of filters appropriate in economics. As examples we have presented a new class of low-pass and band-pass filters for time series.

In further work (see McElroy and Trimbur, 2007), we show how such continuous-lag filters may be discretized for application to real data. One special case of interest is how to adapt the HP filter to changing observation interval and to different modes of measurement, such as stock and flow sampling. More generally, a broad range of filters may be used as the basis for analysis. In addition to model flexibility, there is also flexibility how the $WK$ filters may be adapted to the target of estimation. This has been illustrated with velocity and acceleration filters that could be used in applications where turning point indicators are of interest.

A key aspect of our approach is the rigor and generality of the treatment. A continuous-lag filter gives the basis for estimating a flexible target signal when the sample has various properties, such as unequal spacing, stock or flow sampling, missing data, and mixed frequency data.

# Appendix: 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 (20), 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 $\psi$. The moments of $\psi$

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

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 $\psi$ 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 $j < d$. We first note that representation (20) also extends to the signal: $s(t) = \sum_{j=0}^{d-1} \frac{t^j}{j!} s^{(j)}(0) + [I^d u](t)$. Then the error process is

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

Since $\Psi(L)$ passes polynomials, $(\psi * s)(t) - s(t) = \int (\psi(x) - \Delta_0(x))[I^d u](t - x)\, dx$, where $\Delta_0$ is the Dirac delta function. Note that any filter that does not pass polynomials cannot be MSE optimal, since the error process will grow unboundedly with time. So we have

$$\epsilon(t) = \int (\psi(x) - \Delta_0(x))[I^d u](t - x)\, dx + \int \psi(x) n(t - x)\, dx,$$

which is orthogonal to $y_*$ by Assumption A. Due to the representation (20), it is sufficient to show that the error process is uncorrelated with $[I^d w](t)$. For any real $h$

$$\mathbb{E}[\epsilon(t)w(t + h)] = \int (\psi(x) - \Delta_0(x)) \mathbb{E}\left([I^d u](t - x)[I^d u](t + h)\right)\, dx \qquad (A.1)$$

$$+ \int \psi(x) \mathbb{E}\left[n(t - x)\left(n(t - h) - \sum_{j=0}^{d-1} \frac{(t + h)^j}{j!} n^{(j)}(0)\right)\right]\, dx,$$

28

which uses the fact that $[I^d w](t) = [I^d u](t) + n(t) - \sum_{j=0}^{d-1} \frac{t^j}{j!} n^{(j)}(0)$. Now we have

$$\mathbb{E}\left[[I^d u](t-x)[I^d u](t+h)\right] = \int_0^{t-x} \int_0^{t+h} \frac{(t-x-r)^{d-1}(t+h-z)^{d-1}}{(d-1)!^2} R_u(r-z) \, dz \, dr.$$

(A.2)

If $f_u$ is integrable, we can write $R_u(h) = \frac{1}{2\pi} \int f_u(\lambda) e^{i\lambda h} \, d\lambda$. If $R_u \propto \Delta_0$ instead, then $f_u \propto 1$; we can still use the above Fourier representation of $R_u$ in (A.2), because the various integrals will take care of the non-integrability of $f_u$ automatically. Since $\int_0^x e^{i\lambda y} dy = (1 - e^{i\lambda x})/(i\lambda)$, we obtain that (A.2) is equal to

$$\frac{1}{2\pi} \int f_u(\lambda) \lambda^{-2d} \left( e^{-i\lambda(t-h)} - \sum_{j=0}^{d-1} \frac{(-i\lambda)^j}{j!} (t-h)^j \right) \left( e^{i\lambda(t-x)} - \sum_{j=0}^{d-1} \frac{(i\lambda)^j}{j!} (t-x)^j \right) \, d\lambda.$$

When integrated against $\psi(x) - \Delta_0(x)$, we use the moments property of $\psi$ to obtain

$$\frac{1}{2\pi} \int f_u(\lambda) \lambda^{-2d} \left( e^{-i\lambda(t-h)} - \sum_{j=0}^{d-1} \frac{(-i\lambda)^j}{j!} (t-h)^j \right) \left( e^{i\lambda t} \Psi(e^{-i\lambda}) - e^{i\lambda t} \right) \, d\lambda$$

$$= \frac{1}{2\pi} \int f_u(\lambda) \frac{-f_n(\lambda)}{f_w(\lambda)} \left( e^{i\lambda h} - \sum_{j=0}^{d-1} \frac{(-i\lambda)^j}{j!} (t-h)^j e^{i\lambda t} \right) \, d\lambda.$$

This uses $\Psi(e^{-i\lambda}) - 1 = -\lambda^{2d} f_n(\lambda)/f_w(\lambda)$, which is not integrable if $f_n \propto 1$; yet $f_u f_n / f_w$ will be integrable under the conditions of the theorem. As for the noise term in (A.1), we first note that $n^{(j)}(t)$ exists for each $j < d$ since $w(t)$ exists by assumption; this existence is interpreted in the sense of Generalized Random Processes (Hannan, 1970). In particular

$$\int \psi(x) \mathbb{E}[n(t-x)n(t-h)] \, dx = \int \psi(x) R_n(x-h) \, dx = \frac{1}{2\pi} \int f_n(\lambda) e^{i\lambda h} \Psi(e^{-i\lambda}) \, d\lambda.$$

This Fourier representation is valid even when $f_n \propto 1$, since $\Psi(e^{-i\lambda})$ is integrable by assumption. Similarly,

$$\mathbb{E}[n(t-x)n^{(j)}(0)] = \frac{\partial^j}{\partial z^j} \mathbb{E}[n(t-x)n(z)]|_{z=0} = \frac{\partial^j}{\partial z^j} R_n(t-x-z)|_{z=0} = \frac{\partial^j}{\partial x^j} R_n(t-x)$$

where the derivatives are interpreted in the sense of distributions – i.e., when this quantity is integrated against a suitably smooth test function, the derivatives are passed over via integration by parts:

$$\int \psi(x) \mathbb{E}[n(t-x)n^{(j)}(0)] \, dx = (-1)^j \int \psi^{(j)}(x) R_n(t-x) \, dx.$$

29

Since $\lambda^j \Psi(e^{-i\lambda})$ for $j < d$ is integrable by assumption, we have $\psi^{(j)}(x) = \frac{1}{2\pi} \int (i\lambda)^j \Psi(e^{-i\lambda}) e^{i\lambda x}\, d\lambda$, and the second term in (A.1) becomes

$$\frac{1}{2\pi} f_n(\lambda) \Psi(e^{-i\lambda}) \left( e^{i\lambda h} - \sum_{j=0}^{d-1} \frac{(-i\lambda)^j (t-h)^j}{j!} e^{i\lambda t} \right) d\lambda.$$

This cancels with the first term of (A.1), which shows that $\Psi(L)$ is MSE optimal. Using similar techniques, the error spectral density is obtained as well. $\square$

**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

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

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

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. $\square$

# 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] Gandolfo, G. (1993) *Continuous Time Econometrics.* London, England: Chapman and Hall.

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

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

[12] 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.

[13] 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.

31

[14] Harvey, A. and Trimbur, T. (2007) Trend Estimation, Signal-Noise Rations and the Frequency of Observations. In G. L. Mazzi and G. Savio (ed.), *Growth and Cycle in the Eurozone*, 60-75. Basingstoke: Palgrave MacMillan.

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

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

[17] 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.

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

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

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

[21] Maravall, A. and del Rio, A. (2001). Time Aggregation and the Hodrick-Prescott Filter. *Bank of Spain Working Paper* 0108.

[22] McElroy, T., and Trimbur, T. (2007) A coherent approach to filter design and interpolation for nonstationary stock and flow time series observed at a variable sampling frequency. *mimeo*

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

[24] Ravn, M. and Uhlig, H. (2002). On Adjusting the HP Filter for the Frequency of Observation. *Review of Economics and Statistics* **84**, 371 – 380.

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

[26] 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.

[27] Trimbur, T. (2006) Properties of higher order stochastic cycles. Journal of Time Series Analysis **27**, 1-17.

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