Thomas Trimbur* and Tucker McElroy

Signal Extraction for Nonstationary Time Series with Diverse Sampling Rules

DOI 10.1515/jtse-2014-0026

Abstract: This paper presents a flexible framework for signal extraction of time series measured as stock or flow at diverse sampling frequencies. Our approach allows for a coherent treatment of series across diverse sampling rules, a deeper understanding of the main properties of signal estimators and the role of measurement, and a straightforward method for signal estimation and interpolation for discrete observations. We set out the essential theoretical foundations, including a proof of the continuous-time Wiener-Kolmogorov formula generalized to nonstationary signal or noise. Based on these results, we derive a new class of low-pass filters that provide the basis for trend estimation of stock and flow time series. Further, we introduce a simple and accurate method for low-frequency signal estimation and interpolation in discrete samples, and examine its properties for simulated series. Illustrations are given on economic data.

Keywords: continuous time models, Hodrick-Prescott, low-pass filters, trends, turning points

1 Introduction

Interest among analysts and policy makers often centers around measuring various kinds of signals that represent the chief dynamics of the data. As a leading example, economists and statisticians often seek to estimate the stochastic trend that dominates the long-run evolution of economic time series. Trend and cycle estimation, as well as seasonal adjustment, are applications of signal extraction methodology that have been actively pursued by myriad practitioners. For additional background, see Bell and Hillmer (1984), Harvey (1985), Watson (1986), Clark (1987), Hodrick and Prescott (1997), Harvey and Trimbur (2003), and Alexandrov et al. (2012), among others. Most macroeconomic time series represent stock or flow measurement of some underlying process, which

***Corresponding author: Thomas Trimbur,** Federal Reserve Board – Research and Statistics, 20th and C St, NW 3003 Van Ness Street, NW, Washington, District of Columbia 20008, USA, E-mail: Thomas.Trimbur@census.gov
Tucker McElroy, U.S. Census Bureau – CSRM, 4600 Silver Hill Road, D.C., District of Columbia 20233, USA

may often be viewed as occurring in continuous-time (CT). For instance, transactions and production and distribution activities occur more or less continuously in the national economy.

A key challenge in the design of signal estimators is explicitly determining their dependence upon sampling frequency (e. g., whether monthly, quarterly, or annual, etc.) and sampling type (stock or flow). Harvey and Stock (1993) [henceforth HS] show the discrete models for structural time series models set up in CT. Ravn and Uhlig (2002) treat the case of the Hodrick-Prescott (HP) filter for stock and flow sampling of the observations. These and other papers address the topic for particular models or filters, and for specific filter types – that is, parametric or nonparametric. Also see McElroy and Trimbur (2011) and the references therein.

This paper addresses the challenge of estimator design in a new and general framework. For econometric analysis to be useful and informative across different time series, signal estimators should be designed so that the corresponding features of different series – measured with differing sampling frequency and/or sampling type – are still comparable in a meaningful way. Here, there are four essential contributions: (i) we provide new signal extraction theory and methodology covering the general case of nonstationary data; (ii) we introduce a discretization method that is adapted to sampling frequency and sampling type; (iii) we discuss the discretization of non-parametric filters; (iv) we provide a fast approach to the computation of interpolants.

Our results give the foundation for the strategy, of deriving a single filter expressed over continuous leads and lags, and then proceeding to the discrete filters consistent with the continuous-lag filter and with the sampling conventions. This approach was used in McElroy and Trimbur (2011) [henceforth MT]; it differs fundamentally from the standard approach, as in HS, of first discretizing the CT model according to sampling frequency/type, and then constructing the corresponding discrete-time signal estimators.

Figure 1 illustrates two pathways from the underlying continuous-time process to the discrete-lag filter used in practice on the available data (discrete and of finite length). A continuous-lag filter is derived from a CT process via the operation Ω, which in the case of a stationary process is described in Whittle (1983).

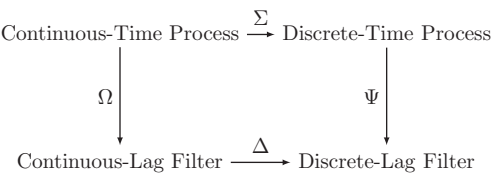

Figure 1: Diagram illustrating two pathways from a continuous-time process to a discrete-lag filter.

A discrete-lag filter is derived from a discrete-time process via Ψ, which is essentially given in the theory of Bell (1984). The discretization of a process is represented by Σ, described in HS and also in Brockwell (1995); the discretization of a filter is represented by Δ, as described in MT. The pathway $\Delta \circ \Omega$ is adopted in this paper, whereas HS follows the $\Psi \circ \Sigma$ strategy. In the case of nonparametric filtering, Ω is replaced by a pre-determined continuous-lag filter. Note that both Σ and Ω rely upon a knowledge of the original CT process' dynamics (say, described through its spectral density).

Discretization of a process (Σ) inexorably involves some loss of accuracy unless that original process is band-limited. For instance, if a process involves no high frequency content past some threshold η, then down-sampling the process – a method of discretization – at an interval δ does not dispense with any frequency content so long as $\pi/\delta \leq \eta$. The same result holds for discretizing filters (Δ): when the original filter corresponding to Ω only passes through frequency components on $[0, \eta]$, then the derivation of filters for discrete sampling at spacing δ, will give an estimate of underlying signal as accurate as the original filter when $\pi/\delta \leq \eta$. With economic data, typically the analyst or practitioner has little or no control over δ, this sampling interval being dictated by the financial or logistical burdens of survey design, weighed against the demands for timely release of information. Then we either make the unwarranted assumption that $\eta \geq \pi/\delta$ – questionable for processes like a production aggregate that can evolve at high frequencies – or accept that the discretization can introduce a degree of aliasing due to the unbounded bandwidth for the underlying variable's evolution. This paper adopts the latter perspective and explores methods of discretization that minimize error in this general framework.

Assuming the model for the observations is given, if the discretization follows the exact method in MT, then the two paths for estimating signals in discrete-time data yield identical results for any given model or filter. In general, the relative attractiveness of the methods depends on three points: (i) mathematical elegance, (ii) ease of application and computation, and (iii) versatility of the method – that is, what uses and perspectives on dynamics does it conveniently provide? Starting with uncorrelatedness in CT, compact expressions for continuous-lag filters arise. However, the model discretization in Σ, the first step of HS, must be deduced for each CT process, sampling type, and signal construct (e. g. point or time-averaged); this leads to signal-noise correlations and other complications in the discrete model that vary with sampling frequency. Further, the significant complexity in the form of the optimal filter in HS's method obscures the main aspects of the signal estimation, and a coherent contrast of filters for series sampled in different ways is impossible. In contrast,

the CL filter at the core of MT's strategy reflect the key signal-noise dynamics, separate from the sampling-induced dynamics. They allow for direct and transparent comparisons and have elegant gain and weight functions.

Regarding point (ii), below we present approximate discretizations of filters that significantly facilitate the implementation relative to MT. In separating out the impact of sampling construction, our approach (here and in MT) is conceptually more straightforward and conducive to analysis and comparison of signal estimators than the HS approach. Moreover, when signal interpolation is needed our approach preserves the basic properties of the estimator, which in a modelling framework incorporates signal-noise dynamics. In contrast, Σ smothers the dynamics of the process between observation times (via the spectral-folding phenomenon described in Koopmans 1974), so that Ψ cannot produce interpolants. In applying Δ, our approach allows for arbitrary time location in addition to a flexible sampling mechanism.

In order to make the approach of $\Delta \circ \Omega$ practical, one must be able to attain CT signal estimators (Ω), and then be able to conveniently discretize these estimators (Δ). Hence, following definitions and preliminary material in Section 2, our paper sets out the rigorous derivation of CT signal extraction formulas for the general case with stochastic trends or other kinds of nonstationarity in Section 3, with the first proof of this new generalization to nonstationary CT presented in the Appendix. Note that McElroy (2013), in contrast, gives a rigorous treatment of CT signal extrapolation and forecasting, as opposed to CT signal extraction.

Subsequent to Section 4, which describes some key trend estimators, Section 5 presents a new method of discretization that is simple and fast to compute, with the greatest accuracy achieved for small sampling intervals. Regarding the discretization of filters (Δ), the exact method of MT leads to mathematically sophisticated expressions that depend on full knowledge of the data process' spectral density; analytical solutions are given for basic trend plus noise models, yet the necessary calculations make the approach impractical for general problems. The approximate discretization method (Δ) of this paper involves greatly simplified mathematical expressions and only requires knowledge of the CT filter (without depending on the process' spectrum). This method is far easier to implement than MT and substantially broadens the scope of applicability of the approach. With the new discretization method, we can handle nonparametric filtering and can interpolate signal estimates in a flexible fashion, which the HS approach cannot do (except in *ad hoc* extensions, such as through splines).

The main objective of the paper is to enable signal estimation across diverse sampling schemes, with respect to a single series of interest, such as CPI (Consumer Price Index) inflation, or with respect to several series, such that

the resulting signals are coherent, in the sense that similar properties of the various time series are extracted for valid comparisons. In addition, the ancillary results may be of broader interest, namely the derivation of CT low-pass filters (Section 4) that generalize the HP (Hodrick and Prescott 1997) filters to CT. This allows us to rigorously examine the assumptions in Ravn and Uhlig (2002), and to quickly produce signal estimates at all time points, including interpolated estimates, that is, between sampling times.

Section 6 compares the new discretization method – called "primary alias filter discretization" – with the optimal techniques of MT, demonstrating through numerical studies the favorable properties of the new approach. The coherency of these signal discretization methods are illustrated on some housing and inflation series in Section 7. Section 8 concludes. The proofs of the main theorems are given in the Appendix.

2 Background on Continuous-Time Signal Extraction

This section reviews the theoretical framework and notation for the analysis of CT signal processing and filtering. Also see Hannan (1970), Koopmans (1974), Priestley (1981) and Bergstrom (1988) for additional material. In this context of economic time series, measured as stock or flow at a certain timing interval, the treatment of signal estimators in CT provides a pure analysis of the estimation or filtering problem that abstracts from sampling construction. Second, it allows for direct derivation of all the discretized filters from a single source filter, thus ensuring their internal coherency.

Let $y = \{y(t),\ t \in \mathbb{R}\}$ denote a real-valued stochastic process that is measurable with finite second moment at each time t, i. e., $\mathbb{E}[y^2(t)] < \infty$ for all $t \in \mathbb{R}$. The process is weakly stationary by definition if it has constant mean – set to zero for simplicity – and autocovariance function (acf) 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 mean-zero Gaussian process, R_y completely describes the dynamics of the stochastic process. A useful class of stationary CT processes that is analogous to discrete-time moving averages is given by

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

where ψ is square integrable on $\mathbb{R}$, and $\{\epsilon(t)\}$ is CT white noise (WN)[1] with variance parameter σ_ϵ^2. In this case, $R_y(h) = (\psi * \psi^-)(h)\sigma_\epsilon^2$, where $\psi^-(x) = \psi(-x)$. It is convenient to work with models expressed in terms of the WN ϵ, partly because this makes it easy to see the connection with discrete models based on white noise disturbances.

In analogy with discrete-time ARMA processes, which can be written as difference equations, built on white noise disturbances, continuous time processes can be written as differential equations, built on CT WN. In this way, another example of a CT stationary process is given by the Continuous-time Autoregressive Moving Average (CARMA) process – see Brockwell (2001), for example – which can be written as

$$a(D)y(t) = b(D)\epsilon(t) \tag{3}$$

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. Brockwell (2004) derives the solution $\{y(t)\}$ to (3) under the stationarity condition – the roots of the equation $a(z) = 0$ must all have strictly negative real part. (This is a weak solution, in the sense that it satisfies an integral representation of (3).) Brockwell and Marquardt (2005) show that a stationary CARMA process can be re-expressed in the form (2) for ψ given by the inverse Fourier Transform of $b(i\lambda)/a(i\lambda)$.

We will be interested in filtering stationary processes (such as CARMA processes), and so we will define the CT lag operator L via the equation

$$L^x y(t) = y(t - x) \tag{4}$$

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

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

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

$$\Psi(L)y(t) = \int_{-\infty}^{\infty} \psi(x)y(t - x)dx = (\psi * y)(t) \tag{6}$$

by definition. The requirement of integrability for the function $\psi(x)$ is a mild condition that is sufficient for many problems. However, when the input process

1 If W is a Levy process, the integral $\int \psi(t - x)dW(x)$ is well-defined in $\mathbb{L}_2$ for each t; then $\epsilon(x)dx$ is used as a short-hand for the Levy increment $dW(x)$, so that CT WN can be formally defined – see Priestley (1981) – and the convolution expression $(y^\star\epsilon)(t)$ follows.

is nonintegrable over t, an integrable ψ 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 development parallels the discussion in Priestley (1981), where the filter is written as

$$\mathcal{L}[\psi](D) = \int_{-\infty}^{\infty} \psi(x) e^{-Dx} dx, \qquad [7]$$

with $\mathcal{L}[\psi]$ denoting the Laplace transform of ψ. Comparing (7) with (5) indicates the heuristic identification

$$D = -\log L, \qquad [8]$$

which is further discussed below. We prefer to describe filters via (5), in terms of L rather than D, because of the intuitive linkage with convolution (6) and discrete time filtering, although (7) has a longer history.

In analogy with the discrete-time case, the frequency response function (frf) of a filter $\Psi(L)$ is obtained by replacing L by $e^{-i\lambda}$:

$$\Psi(e^{-i\lambda}) = \int_{-\infty}^{\infty} \psi(x) e^{-i\lambda x} dx, \quad \lambda \in \mathbb{R}. \qquad [9]$$

This is discussed in Koopmans (1974), who also considers the special case of the derivative filter. Denoting the CT Fourier Transform by $\mathcal{F}[\cdot]$, eq. [9] 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 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 spectral density of a weakly stationary continuous time process $\{y(t)\}$ is the Fourier Transform of its acf R_y:

$$f_y(\lambda) = \mathcal{F}[R_y](\lambda), \quad \lambda \in \mathbb{R}. \qquad [10]$$

This case of a CT autoregression is discussed in Priestley (1981); also see eq. [7] of Brockwell (2004) for the CARMA case. The gain function of a filter $\Psi(L)$ is the magnitude of the frf, namely

$$G(\lambda) = |\mathcal{F}[\psi](\lambda)|, \quad \lambda \in \mathbb{R}. \qquad [11]$$

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 (10) and (11) are defined over the entire real line. Given a candidate frf g taking the inverse Fourier Transform in CT yields the associated weighting kernel:

$$\mathcal{F}^{-1}[g](x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} g(\lambda)e^{i\lambda x}d\lambda, \quad x \in \mathbb{R}. \tag{12}$$

This expression is well-defined for any integrable g. 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 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 $\lambda \to \infty$.

In (4), the extension of the lag operator L to the CT 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 $1–L$ used to define nonstationary models is discussed in Hannan (1970, 55) and Koopmans (1974).

The mean-square differentiation operator D is formally defined in Priestley (1981), and that work shows that $L = e^{-D}$ via Taylor series arguments. We provide an alternative derivation of the equivalent relation $D = -\log L$. We might consider the limit of measuring the displacement of a CT process, per unit of time, over an arbitrarily small interval δ:

$$\begin{aligned}\frac{d}{dt}y(t) &= \lim_{\delta\to 0}\frac{y(t+\delta)-y(t)}{\delta} = \lim_{\delta\to 0}\frac{L^{-\delta}-1}{\delta}y(t) = \lim_{\delta\to 0}\frac{e^{-\delta\log L}-1}{\delta}y(t)\\ &= \lim_{\delta\to 0} -\log L e^{-\delta\log L}y(t) = -\log L\, y(t).\end{aligned}$$

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-square differentiable processes y, implying (8).

This operator D will be our main building block for nonstationary CT 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 $(i\lambda)^d$, which is a high-pass filter; see Koopmans (1974).

Thus a natural class of models is the Integrated Filtered Noise processes, which are given by

$$D^d y(t) = \Theta(L)\epsilon(t) \tag{13}$$

for some integrable θ with $\Theta(L) = \int \theta(x)L^x\,dx$, and order of differential $d \geq 0$. Also $\{\epsilon(t)\}$ is WN. This class will be denoted $y \sim \text{IFN}(d)$; it encompasses a wide variety of linear CT models. As an example, Brockwell and Marquardt (2005) define the class of Continuous-time Autoregressive Integrated Moving Average (CARIMA) models as the solution to

$$\alpha(D)D^d y(t) = b(D)\epsilon(t). \tag{14}$$

Thus, applying the d-fold derivative filter transforms y into a stationary CARMA (p,q) process. This defines the CARIMA(p,d,q) process. The original process $\{y(t)\}$ is nonstationary and is said to be integrated of order d in the CT sense. Now this can be put into an IFN(d) form: starting from (14), we can write (formally)

$$D^d y(t) = \frac{b(D)}{\alpha(D)}\epsilon(t), \tag{15}$$

so that $\Theta(L) = b(D)/a(D)$, or $\theta = \mathcal{F}^{-1}[b(i\lambda)/a(i\lambda)]$. So we see that CARIMA processes can be expressed as IFN processes where the kernel θ's Laplace transform is a rational function.

3 Nonstationary 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, extending the discrete-time results of Bell (1984). Whittle (1983) shows a similar result for CT nonstationary processes, but omits the proof and in particular, does not describe the importance of initial conditions. Kailath, Sayed, and

Hassibi (2000, 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. Consider the following decomposition for a CT process $\{y(t)\}$:

$$y(t) = s(t) + n(t), \quad t \in \mathbb{R} \tag{16}$$

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 $s \sim I(d)$, or is 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 dth 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 acfs R_u and R_n are integrable. An extension can 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 constant, indicating a corresponding WN process. The process y satisfies the stochastic differential equation

$$w(t) = D^d y(t) = u(t) + D^d n(t). \tag{17}$$

From (1), (8), and (10) it follows – when the autocovariance function is well-defined – that the spectral density of $\{w(t)\}$ is

$$f_w(\lambda) = f_u(\lambda) + \lambda^{2d} f_n(\lambda). \tag{18}$$

More generally – even when it is a non-integrable function – we will make use of (18) as the definition of f_w, though it might not be interpretable as a spectral density. (Throughout the rest of the paper, we will nevertheless refer to such functions as actual spectral densities.) From Hannan (1970, 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

$$y(t) = \sum_{j=0}^{d-1} \frac{t^j}{j!} y^{(j)}(0) + \left[I^d w\right](t) \quad [19]$$

with the I operator defined by $[I^d w](t) = \int_0^t w(z)(t-z)^{d-1} dz/(d-1)!$, and $y^{(j)}(0) = [d^j y/dt^j](0)$. Note that (19) holds for the signal $\{s(t)\}$ as well, if we substitute s for y and u for w.

For an $I(d)$ process, let $\mathbf{y}^\star(\mathbf{0}) = \{y(0), \dot{y}(0), \cdots, y^{(d-1)}(0)\}$ denote the collection of d higher order derivatives at time $t = 0$. It is assumed that $\mathbf{y}^\star(\mathbf{0})$ 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 rather than actual lagged observed variables.

With these background concepts, the theoretical signal extraction problem for a bi-infinite series $\{y(t)\}$ that follows (16) is defined as follows. The optimal *linear* estimator of the signal $s(t)$, by definition, gives the minimum mean square error among all linear estimators. 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 ψ. The problem is to determine the optimal choice of $\Psi(L)$ for general nonstationary models of the form (16). The following theorem shows the main result.

Theorem 1: *For a weakly stationary process satisfying (16), suppose that* $\mathbf{y}^\star(\mathbf{0})$ *is uncorrelated with both* $\{u(t)\}$ *and* $\{n(t)\}$. *Also assume that* $\{u(t)\}$ *and* $\{n(t)\}$ *are mean zero weakly stationary processes that are uncorrelated with one another, with acfs 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*

$$\begin{aligned} \hat{s}(t) &= \Psi(L)y(t) \\ \Psi(L) &= \int_{-\infty}^{\infty} \psi(x) L^x dx \\ \psi(x) &= \mathcal{F}^{-1}[g](x). \end{aligned} \quad [20]$$

The function ψ *is the continuous weighting kernel of the optimal filter. The spectral density of the error process* $e(t) = \hat{s}(t) - s(t)$ *is*

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

and 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. One of the important properties of the WK filters is that they pass polynomials, in analogy with 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

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

for any j up to the degree of p, with δ 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 $2d-1$. In the next section we provide several detailed examples of WK filters.

4 Examples of Continuous-Lag Trend Filters

Compared to discrete-time analogues (with filters either set up directly for discretely sampled observations or deduced from underlying processes and sampling rules), the continuous-lag filters tend to have more compact forms for gain functions or time domain expressions, which can make the analysis of their properties simpler and more transparent. In this section we provide examples of continuous-lag WK filters that are based on the class of CARIMA models and that give CT versions of some well-known existing filters, as well as associated filters that target a function, or characteristic of signal. Using the CARIMA class to form WK weighting kernels offers flexibility for a range of applications and is also convenient for computational reasons, as the WK frf is given by a rational functions in λ^2, facilitating Fourier inversion via the calculus of residues.

Here we focus on trend signals, and the trend plus noise decomposition is written as

$$y(t) = \mu(t) + \epsilon(t), \qquad [21]$$

where $\{\mu(t)\}$ denotes the stochastic level, and $\{\epsilon(t)\}$ is WN with variance parameter σ_ϵ^2, denoted by $\epsilon(t) \sim WN(0, \sigma_\epsilon^2)$ See Harvey (1989) for discussion.

Illustration 1: Local Level Model The local level model assumes

$$D\mu(t) = \eta(t), \qquad \eta(t) \sim \text{WN}\left(0, \sigma_\eta^2\right). \tag{22}$$

The signal-to-noise ratio in the CT framework is defined as $q = \sigma_\eta^2 \Big/ \sigma_\epsilon^2$. So the observed process y requires one derivative for stationarity, and we write $w(t) = Dy(t)$. The spectral densities of the differentiated trend and observed process are

$$f_\eta(\lambda) = q\sigma_\epsilon^2 \quad f_w(\lambda) = f_\eta(\lambda) + \lambda^2\sigma_\epsilon^2 = (q + \lambda^2)\sigma_\epsilon^2. \tag{23}$$

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. 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-to-noise ratio of the underlying CT process.

Illustration 2: Smooth Trend Model Even when considering an economic series with regularly spaced observations, the use of different sampling frequencies requires the careful design of filters to ensure consistency. For instance, Ravn and Uhlig (2002) consider the well-known HP filter (Hodrick and Prescott 1997), while Harvey and Trimbur (2007) discuss trend estimation in a more general setting. The local linear trend model (Harvey 1989, p. 485) has the following specification:

$$D\mu(t) = \beta(t) + \eta(t), \quad \eta(t) \sim \text{WN}(0, \sigma_\eta^2)$$

$$D\beta(t) = \zeta(t), \quad \zeta(t) \sim \text{WN}(0, \sigma_\zeta^2)$$

where $\{\eta(t)\}$ and $\{\zeta(t)\}$ are uncorrelated with each other. 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 ϵ is WN uncorrelated with ζ. Here the signal-to-noise ratio is $q = \sigma_\zeta^2 \Big/ \sigma_\epsilon^2$.

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 CT smooth trend model. We may write the model as

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

The spectral densities of the appropriately differentiated trend and series are

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

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

$$\psi(x) = \frac{q^{1/4}\exp\{-|x|q^{1/4}/\sqrt{2}\}}{2\sqrt{2}}\left(\cos(|x|q^{1/4}/\sqrt{2}) + \sin(|x|q^{1/4}/\sqrt{2})\right) \qquad [24]$$

This gives the CT extension of the HP filter. From the discussion following Theorem 1, the kernel in (24) passes cubics.

Figure 2 shows the weighting function for three different values of q. As the signal-to-noise ratio increases, the trend becomes more variable relative to

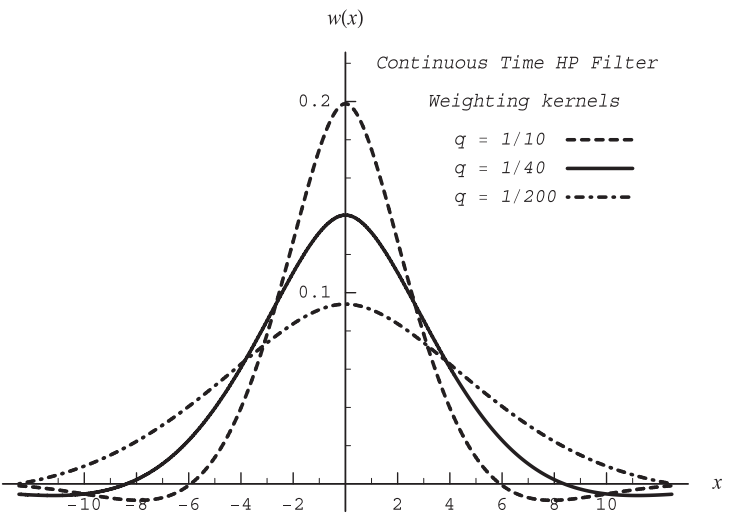

Figure 2: Weighting kernel for continuous-lag HP filter for $q = 1/10$, $1/40$, and $1/200$.

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 and cubics.

Illustration 3: Continuous-Lag Low-Pass Here we introduce low-pass filters in CT that are analogous to the filters derived by Harvey and Trimbur (2003) for the corresponding discrete-time models. A simple nonseasonal model for a CT process in macroeconomics is given by

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

where $\{\mu_m(t)\}$ is a trend component that accounts for long-term movements. The irregular $\{\varepsilon(t)\}$ is meant to absorb any random, or nonsystematic variation, and is assumed to be $\mathrm{WN}(0, \sigma_\varepsilon^2)$. The definition of the mth order trend is

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

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

In formulating the trend estimation as a signal extraction problem (Section 3), we set the nonstationary signal to $\{\mu_m(t)\}$ and the noise term to $\{\varepsilon(t)\}$. Then the optimal filter $\Psi(L)$ can be constructed from Theorem 1, and its frf is given by

$$LP_m(\lambda) = \frac{q}{q + \lambda^{2m}}, \qquad [25]$$

where LP_m stands for low-pass filter of order m, and $q = \sigma_\zeta^2 \big/ \sigma_\varepsilon^2$ is the signal-to-noise ratio for the trend. This definition (25) parallels the development of Harvey and Trimbur (2003) for the discrete-time case. The formula for the kernel is more complicated than (24), and is discussed in the Appendix.

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 $H\, s(t)$ 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 mth derivative of ψ, 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 $H\,s(t)$ with H a linear operator. To compute the mean squared error of $(H\Psi(L))y(t)$ as an estimate of the target $H\,s(t)$, 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 λ^2, and the result is then integrated over $\mathbb{R}$.

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.

We compute the frfs and kernels for the Smooth Trend Model of Illustration 2. Since $\lambda^2(1+\lambda^4/q)^{-1}$ is integrable, 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 velocity filter, or first derivative with respect to time, has the interpretation of a growth rate for the trend. The weighting kernel in Figure 3 shows how the

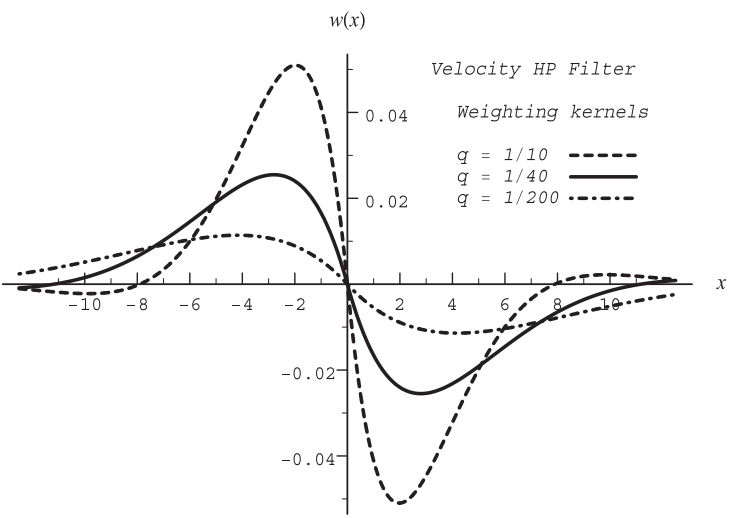

Figure 3: Weighting Kernels for Velocity WK filter based on Smooth Trend model for $q = 1/10,\ 1/40,$ and $1/200$.

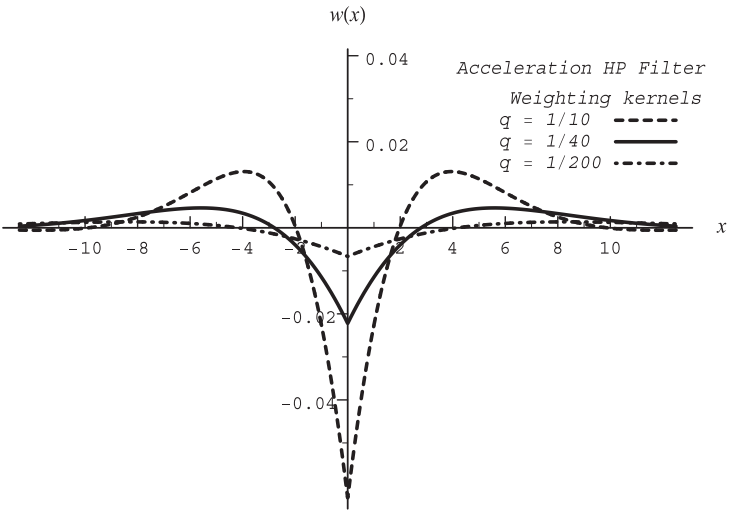

Figure 4: Weighting Kernels for Acceleration WK filter based on Smooth Trend model for $q = 1/10$, $1/40$, and $1/200$.

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 4 has a characteristic sharp decline around the origin, so that contemporaneous and nearby values are subtracted in estimating changes in growth.

5 Primary Alias Filter Discretization

Here we introduce a simple method for accomplishing Δ in Figure 1, i. e., filter discretization. In doing so, we provide a decomposition of the discrete signal estimation error based on intuitive ideas about signal-to-noise relationships and measurement.

Given a continuous-lag filter $\Psi(L)$ with frf g, we have a theoretical signal estimate of $\Psi(L)y(t)$ for any time t. Suppose we have a stock series $y_\tau = y(\delta\tau)$ available to us, for sampling frequency δ and $\tau \in \mathbb{Z}$ (we consider the case of flows below). But the target time t need not fall on the discrete grid $\delta\mathbb{Z}$; rather, we may write $t = \delta\tau + \delta c$ for some $c \in [0, 1)$ (any $t \in \mathbb{R}$ may be decomposed in this way). The number c is called the interpolant. The interpolated discretization problem is to compute a discrete filter $\Psi_\delta(B) = \Sigma_j \psi_j B^j$ such that $\Psi_\delta(B)y_\tau$ is close

to $\Psi(L)y(\delta(\tau+c))$. Here $B=L^{\delta}$, because the spacing between discrete observations is δ. The discretization error is $\Psi_{\delta}(B)y_t-\Psi(L)y(t)$. The discrete filter should produce discretization MSE that tends to zero as $\delta\to 0$, is tolerably low when $\delta>0$, does not depend upon the data process, and is simple to compute.

We begin with a basic decomposition of the discretization error, supposing that y is an $I(d)$ CT stochastic process. We use the abbreviation $e_c(\lambda)=e^{i\lambda\delta c}$, and interest will focus on the frf ge_c (the dependency of e_c on δ is suppressed in the notation). Let the frf of the discretized approximate filter Ψ_{δ} be denoted by $g_{\delta}(\lambda)=\Psi_{\delta}(e^{-i\lambda\delta})$ for $\lambda\in[-\pi/\delta,\pi/\delta]$; this is the appropriate band of frequencies for the sampled process, in terms of units for the original CT process' frequencies (see MT). Note that π/δ is the so-called Nyquist frequency.

By $\bar{g}_{\delta}$, we refer to the periodic extension of g_{δ} from the domain $[-\pi/\delta,\pi/\delta]$ to the entire real line. A key condition, which ensures that there is no bias in the discretization of $\Psi(L)$, is that

$$(ge_c)^{(j)}(0)=g_{\delta}^{(j)}(0). \tag{26}$$

for $j=0, 1, 2, \ldots, d$. This will guarantee that $\Psi_{\delta}(B)$ treats discrete time polynomial effects in exactly the same way as the continuous time treatment, so that if y were a degree d polynomial, the discretization error would be identically zero. The following result explicitly determines the discretization error.

Proposition 1: *Let $\Psi(L)$ be a filter with frf g that satisfies (26) for $j\le d$, where y is an I(d) process such that $D^d y=w$ has orthogonal increments process $d\mathbf{Z}$. Let $a_{\delta}(\lambda)$ be identically one for a stock sampling, but equal to $(1-e^{-i\lambda\delta})/(i\lambda)$ for a flow sampling. Then the discretization error is*

$$\Psi(L)y(\delta\tau+\delta c)-\Psi_{\delta}(B)y_{\tau}=\int_{-\infty}^{\infty}e^{i\lambda\delta\tau}[g(\lambda)e_c(\lambda)-\bar{g}_{\delta}(\lambda)a_{\delta}(\lambda)](i\lambda)^{-d}d\mathbf{Z}(\lambda),$$

and when the mean square of the discretization error exists, it is equal to

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}|g(\lambda)e_c(\lambda)-\bar{g}_{\delta}(\lambda)a_{\delta}(\lambda)|^2 f_w(\lambda)\lambda^{-2d}d\lambda. \tag{27}$$

The discretization MSE (27) can be decomposed into the sum of two positive terms: the minimal possible discretization MSE, plus the extra error due to using the filter $\Psi_{\delta}(B)$ – which need not be optimal. Theorem 1 of MT provides an explicit formula for the discrete filter with minimal discretization MSE. This Optimal Discrete Filter (ODF) has frf u_{δ} (its formula is discussed in the Appendix) defined on $[-\pi/\delta,\pi/\delta]$, with periodic extension $\bar{u}_{\delta}$. Then in the stock sampling case, (27) decomposes into

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}|g(\lambda)e_c(\lambda)-\bar{u}_\delta(\lambda)|^2 f_w(\lambda)\lambda^{-2d}d\lambda+\frac{1}{2\pi}\int_{-\infty}^{\infty}|\bar{u}_\delta(\lambda)-\bar{g}_\delta(\lambda)|^2 f_w(\lambda)\lambda^{-2d}d\lambda, \quad [28]$$

which is established in the proof of Proposition 2 below. The right-hand term in (28) can be made larger or smaller by taking different choices of g_δ, while the left-hand term cannot be altered by our choices. Hence taking g_δ equal to u_δ gives the lowest possible discretization MSE, i. e., the ODF choice. But other choices may also provide low MSE. In particular, consider $g_\delta = ge_c 1_{[-\pi/\delta,\pi/\delta]}$, i. e., restriction of ge_c to the primary band of aliases.[2] This choice will be called the Primary Alias Discrete Filter (PADF). Note that (26) is easily seen to be satisfied. Also the discretization MSE (for the stock case) can then be expressed as

$$\frac{1}{2\pi}\int_{|\lambda|>\pi/\delta}|g(\lambda)e_c(\lambda)-\bar{g}_\delta(\lambda)|^2 f_w(\lambda)\lambda^{-2d}d\lambda, \quad [29]$$

which has an integrand bounded above by an integrable function for all δ; then by the Dominated Convergence Theorem this quantity tends to zero as $\delta \to 0$, i. e., discretization MSE vanishes with sampling frequency for the PADF. As a special case, if f_w is band-limited, e. g. restricted to an interval $[-\eta, \eta]$ for some $\eta > 0$, then for $\delta \le 2\pi/\eta$ the discretization MSE (29) is identically zero.

Returning to the general case, the extra MSE arising from the PADF over the ODF is negligible relative to the ODF MSE as $\delta \to 0$, as shown in the following result. First we require the following concept: we say that a function k has tails of order γ if $k(\lambda) \sim C|\lambda|^{-\gamma}$ as $\lambda \to \pm\infty$ for some constant $C > 0$.

Proposition 2: *Suppose that the spectral density f_w of the series $\{w(t)\}$ has tails of order $\alpha \ge 0$; if $d = 0$ we require $\alpha > 0$, otherwise $\alpha = 0$ is permitted. If the PADF is used to discretize a given continuous lag filter with frf g, then the ratio of the ODF discretization MSE to the PADF discretization MSE tends to unity as $\delta \to 0$.*

This provides some justification for the PADF technique; when δ is suitably small (relative to the behavior of g and f_w), then there is no real loss in using the PADF instead of the ODF. The advantage of the PADF is ease of computation; note that no knowledge of the data process' dynamics is required, unlike the case of the ODF. This means that once we determine the CT filter, no further

2 For any frequency $\lambda \in \mathbb{R}$, its aliases with respect to sampling frequency $1/\delta$ is the set $\{\lambda + 2\pi h/\delta\}_{h=-\infty}^{\infty}$ (see Koopmans 1974), and the primary alias corresponds to taking h such that $\lambda + 2\pi h/\delta \in [-\pi/\delta, \pi/\delta]$.

reference to the underlying CT series is needed, and the coefficients can be computed for each choice of δ by direct and simple discretization. The filter coefficients of the PADF are given by

$$\psi_j = \frac{\delta}{2\pi}\int_{-\pi/d}^{\pi/\delta} g(\lambda)e_c(\lambda)e^{i\lambda\delta j}d\lambda = \frac{1}{2\pi}\int_{-\pi}^{\pi} g(\lambda/\delta)e^{i\lambda(j+c)}d\lambda.$$

Note that $\psi_j/\delta \sim \psi(\delta(j+c))$ as $\delta \to 0$, though this is of little benefit when $\delta > 0$. In the above, we only need to know the frf g – as f_w itself does not enter into the formula, and generally, the above coefficients are best computed numerically. The reason that analytical formulas tend not to be available, is that g in many scenarios is a rational function of λ^2 (e. g., when derived as the optimal signal estimators for the broad CARIMA class of processes), and there are no convenient formulas for definite bounded integrals of mixed trigonometric and polynomial functions. Now due to the symmetry of g, we can re-express the coefficients as

$$\psi_j = \frac{1}{\pi}\int_0^{\pi} g(\lambda/\delta)\cos[\lambda(j+c)]\,d\lambda, \qquad [30]$$

which can be computed via a Riemann approximation.

For a process that is flow-sampled, the above discussion can be amended slightly. Eq. [27] has a decomposition in terms of the ODF, just as we elucidated in the stock case. The flow PADF arises by setting $g_\delta = ge_c/a_\delta 1_{[\pi/\delta,\,\pi/\delta]}$, and the resulting filter coefficients are given by

$$\psi_j = \frac{1}{\pi\delta}\int_0^{\pi} g(\lambda/\delta)\lambda\frac{\sin[\lambda(j+c+1)] - \sin[\lambda(j+c)]}{2-2\cos\lambda}\,d\lambda. \qquad [31]$$

This is readily approximated by a Riemann sum. Note that we have formulated the discretization problem in terms of a stock target signal, in that we try to approximate $\Psi(L)y(\delta t)$. MT also discusses the case of a flow target signal, which changes some of the derivations. The flow target signal, in terms of an annual rate, would resemble $\Psi(L)C(L)y(\delta t)$ with $C(L) = \delta^{-1}\int_0^\delta L^x\,dx$. This changes the mathematics a bit, but is not of primary interest and so is not pursued further here.

To complete this section, let us consider a popular example, namely the continuous-lag low-pass filters of (25), written $g(\lambda; q) = (1+\lambda^{2m}/q)^{-1}$ for $q > 0$. It is easy to see that $g(\lambda/\delta; q) = g(\lambda; \delta^{2m}q)$, which has implications for the filter coefficients. Noting that the filter coefficients really depend on δ and q (as well as c, but this will be suppressed), we can write $\psi_j(\delta, q)$ in (30). Then it follows that

$$\psi_j(\kappa\delta; q) = \psi_j(\delta; \kappa^{2m} q) \quad [32]$$

for any j and c, where κ is some positive multiplier for converting sampling frequencies. This says that the PADF at a given sampling frequency $\kappa\delta$ and signal-to-noise ratio q is exactly the same as that for sampling frequency δ but with adjusted signal-to-noise ratio of $\kappa^{2m}q$. For example, if $\delta = 1/12$ and $\kappa = 3$ then the rule tells us that the quarterly coefficients (with signal-to-noise ratio q) can be obtained from the monthly coefficients with signal-to-noise ratio of $3^{2m}q$. This result can be compared to that of Ravn and Uhlig (2002), which focused on the $m = 2$ case and obtained a quartic rule for relating Hodrick-Prescott coefficients. For $m = 1$, the rule is quadratic instead. In the case of a flow-sampled series, the coefficients written as a function of sampling frequency and signal-to-noise ratio satisfy

$$\psi_j(\kappa\delta; q) = \kappa\psi_j(\delta; \kappa^{2m} q). \quad [33]$$

This relation resembles the stock case, but note the extra multiplier of κ that is required. In summary, the PADF differs according to whether the data is stock or flow sampled: formulas (32) and (33), respectively for the stock and flow cases, show how filter coefficients depend upon κ, the ratio of differing sampling frequencies. The ODF of MT also had different filters for stock and flow series, although the relationships were more complicated because the true process was involved through its spectral density.

6 Numerical Results for Primary Alias Discretization

In order to provide a numerical exploration of Proposition 2, we consider the PADF approximation to the low-pass (LP) filter $(1+\lambda^{2m}/q)^{-1}$ with $m = 1, 2$ and various values of q. This is applied to various underlying processes that are assumed to be stock-sampled, including a CAR(1), a CARMA(2,1), and a CARIMA (1,1,0) with various parameter values. Although we would not recommend using the LP filter on stationary data, we consider the CARMA examples in order to demonstrate the quality of the PADF approximation under a scenario that is sub-optimal (primarily because the pseudo-spectral density of non-stationary processes has tails that decay more quickly, the PADF method tends to give more accurate results relative to the stationary case). The CARMA processes considered for a simulated process $\{X(t)\}$ are as follows:

$$\begin{aligned}
&\text{Process 1}: \quad (D-.5)X(t) = \epsilon(t)\\
&\text{Process 2}: \quad (D-.1)X(t) = \epsilon(t)\\
&\text{Process 3}: \quad (D^2 - \cos(\pi/60)D + 1/4)X(t) = \epsilon(t)\\
&\text{Process 4}: \quad (D^2 - 2\cos(\pi/12)D + 1)X(t) = (1-2D)\epsilon(t)\\
&\text{Process 5}: \quad D(D-.2)X(t) = \epsilon(t)
\end{aligned}$$

Here $\{\epsilon(t)\}$ is WN with variance parameter set to one. For each model, we consider the frequencies $\delta = 1, 1/4, 1/12, 1/52$ and a range of q values. Results are summarized in Tables 1 and 2, for the cases of $m = 1, 2$ respectively. For

Table 1: Discretization MSEs.

δ					q
Model 1	**100**	**10**	**1**	**1/10**	**1/100**
1	0.88323	0.86745	0.93736	0.97305	0.97896
1/4	0.89336	0.96720	0.99420	0.99820	0.99865
1/12	0.96578	0.99489	0.99929	0.99979	0.99985
1/52	0.99722	0.99969	0.99996	0.99999	0.99999
Model 2	**100**	**10**	**1**	**1/10**	**1/100**
1	0.89274	0.88226	0.95537	0.99226	0.99837
1/4	0.89435	0.96838	0.99544	0.99944	0.99989
1/12	0.96591	0.99503	0.99943	0.99993	0.99999
1/52	0.99723	0.99969	0.99997	0.99999	0.99999
Model 3	**100**	**10**	**1**	**1/10**	**1/100**
1	0.97861	0.97390	0.99351	0.99951	0.99989
1/4	0.97842	0.99647	0.99983	0.99999	0.99999
1/12	0.99605	0.99980	0.99999	0.99999	0.99999
1/52	0.99992	0.99999	0.99999	0.99999	0.99997
Model 4	**100**	**10**	**1**	**1/10**	**1/100**
1	0.81529	0.75587	0.77713	0.76865	0.75795
1/4	0.88690	0.96797	0.98090	0.98112	0.98001
1/12	0.96498	0.99382	0.99779	0.99787	0.99774
1/52	0.99718	0.99963	0.99988	0.99989	0.99989
Model 5	**100**	**10**	**1**	**1/10**	**1/100**
1	0.97980	0.97537	0.99413	0.99966	0.99999
1/4	0.97850	0.99650	0.99983	0.99999	0.99999
1/12	0.99606	0.99980	0.99999	0.99996	0.99997
1/52	0.99991	0.99993	0.99996	0.99994	0.99944

Ratio of ODF discretization MSE to the PADF discretization MSE for the low pass filter with frequency response function $(1+\lambda^2/q)^{-1}$, as a function of q and δ. The five processes are discussed in Section 6.

Table 2: Discretization MSEs.

δ					q
Model 1	**100**	**10**	**1**	**1/10**	**1/100**
1	0.83857	0.90243	0.95309	0.97124	0.97704
1/4	0.98143	0.99322	0.99696	0.99815	0.99853
1/12	0.99792	0.99924	0.99966	0.99979	0.99984
1/52	0.99989	0.99996	0.99998	0.99999	0.99999
Model 2	**100**	**10**	**1**	**1/10**	**1/100**
1	0.85297	0.91931	0.97166	0.99040	0.99639
1/4	0.98264	0.99445	0.99820	0.99939	0.99977
1/12	0.99806	0.99938	0.99980	0.99993	0.99997
1/52	0.99989	0.99997	0.99999	0.99999	0.99999
Model 3	**100**	**10**	**1**	**1/10**	**1/100**
1	0.95929	0.98648	0.99858	0.99976	0.99988
1/4	0.99966	0.99997	0.99999	0.99999	0.99999
1/12	0.99999	0.99999	0.99999	0.99999	0.99999
1/52	0.99999	0.99999	0.99999	0.99999	0.99997
Model 4	**100**	**10**	**1**	**1/10**	**1/100**
1	0.73481	0.77050	0.78660	0.77701	0.76545
¼	0.97206	0.98180	0.98314	0.98203	0.98081
1/12	0.99685	0.99795	0.99810	0.99798	0.99783
1/52	0.99983	0.99989	0.99990	0.99989	0.99989
Model 5	**100**	**10**	**1**	**1/10**	**1/100**
1	0.96121	0.98763	0.99896	0.99993	0.99999
1/4	0.99967	0.99998	0.99999	0.99999	0.99999
1/12	0.99999	0.99999	0.99999	0.99999	0.99998
1/52	0.99999	0.99987	0.99992	0.99988	0.99914

Ratio of ODF discretization MSE to the PADF discretization MSE for the low pass filter with frequency response function $(1+\lambda^4/q)^{-1}$, as a function of q and δ. The five processes are discussed in Section 6.

spectral densities that decay more rapidly at higher frequencies, the approximation error in the PADF will be less than with slowly decaying spectra. Therefore we can expect the ratio of minimal discretization MSE to PADF discretization MSE to be higher for Process 2 than Process 1, because the latter is closer to the behavior of a Brownian Motion, and has spectrum decaying as λ^{-2} for large λ. Likewise, Process 4 should have more PADF discretization error than Process 3, because the presence of the moving average term ensures that the spectrum decays less rapidly than the pure CAR(2). Because Process 5 is integrated, the PADF discretization MSE should be close to the minimal MSE. These patterns are evident in the Tables.

It is also clear that decreasing δ increases the ratio, as expected. The effect of changing q is less clear, although a smaller q implies a more quickly decaying frequency response function. This is for the most part reflected in Tables 1 and 2, in that the ratio of discretization MSEs increase towards unity as q decreases, although the relationship is not monotonic.

These results demonstrate how the approximation error changes with q and δ. Even when the PADF discretizes in a nearly optimal fashion, the signal would still be measured imperfectly. In the limit, if we were able to construct the CT estimate, the discrepancy that would remain represents an inherent part of the signal extraction, after controlling for sampling conditions. As another contribution to the overall estimation error, the ODF discrepancy (from optimal CT signal) arises from the innate loss of information due to data discretization; the tables demonstrate that we will not add appreciably to the ODF error in using the simpler PADF approach.

We exhibit the ODF discretization MSE next, in comparison to the signal extraction MSE. This intrinsic MSE can be consistently compared across different data processes subject to different kinds and frequencies of measurement. So consider the application of the LLM trend filter with parameter q with $m=1$, as defined above, applied to flow-sampled series. This is the optimal filter when the true DGP consists of a Brownian Motion based on shock dispersion $q\sigma^2$ plus an independent White Noise of dispersion σ^2. Then the data process requires one differentiation to stationarity, and $\{W(t)\}$ corresponds to a CARMA(0,1) with MA polynomial $1+D/\sqrt{q}$ and variance $q\sigma^2$. The flow-sampled trend process turns out to be an ARIMA(0,1,1) with autocovariance function equal to $2q\delta^3\sigma^2/3$ at lag zero and $q\delta^3\sigma^2/6$ at lag one (see Harvey and Trimbur 2007); the discrete MA coefficient is $2-\sqrt{3}$ (obtained by spectral factorization) with an innovation variance of $q\delta^3\sigma^2/(12-6\sqrt{3})$. The flow-sampled irregular process is a discrete-time white noise with variance $\delta\sigma^2$.

The signal extraction MSE for the continuous-lag filter is the integral of $f_u f_n/f_w$, divided by 2π, according to Theorem 1; in this case it equals $\sqrt{q}/2$. As for the discretization MSE, it increases roughly according to a $\sqrt{q}$ rule. In Table 3 we show discretization MSE for the ODF, computed using the formulas in the Appendix (see the proof of Proposition 2) for various δ and q, and with $\sigma^2=1$. This can be compared to the signal extraction error $\sqrt{q}/2$, which is typically larger. Note that, whereas the other tables display relative discretization MSEs, Table 3 shows absolute MSEs; as expected, for a given q, the MSE falls as δ decreases. The results indicate that discretization error is small relative to signal extraction error; the ratio of discretization MSE to signal extraction MSE in the highest case $(\delta=1)$ is 2.06 %, 3.54 %, 4.08 %, 4.15 %, and 4.15 % respectively, for $q=100$, 10, 1, .1, .01. Both types of MSE decrease as q decreases, but signal

Table 3: Discretization MSEs.

δ					q
	100	**10**	**1**	**1/10**	**1/100**
1	0.102828	0.056003	0.020400	0.006561	0.002077
1/4	0.011665	0.004061	0.001297	0.000411	0.000130
1/12	0.001425	0.000456	0.000144	0.000046	0.000014
1/52	0.000077	0.000024	0.000008	0.000002	0.000001
CL MSE	5	1.581	0.5	0.15811	0.05

The ODF discretization MSE based on asymptotic calculations, displayed by q and δ. Also shown in the last row is the continuous-lag signal extraction MSE, denoted by CL MSE, as a function of q.

extraction MSE decreases more rapidly, so that discretization MSE has the most impact when $\delta = 1$ and $q = .01$.

7 Applications

Consider the inflation series based on the consumer price index (CPI), which were analyzed in MT as both monthly and annual flow variables based on the underlying, continuous rate of change, *dP/dt*. Using the stock- or point-target formulas for signal extraction from a flow series, the analysis of MT discretized the Local Level Model for the purpose of extracting the instantaneous trend in *dP/dt* over the sample period; these filter expressions used some technical derivations provided in that paper. The resulting monthly and annual signals had the desired relationships; indeed, the annual trend interpolated to monthly time interpolants matched quite closely the monthly trend. However, a drawback of this method is the requirement for elaborate derivations. A key motivation of the PADF method is to facilitate discretization of a given CT filter without requiring knowledge of the underlying data process' dynamics (i. e., its spectral density). So it is natural to ask: does the application of the PADF method to the CPI produce similar results as obtained from the ODF results of MT?

So suppose that a low-pass filter with frf $(1+\lambda^2/q)^{-1}$ is used for trend estimation, which corresponds to a random walk extracted from ambient idiosyncratic noise in CT. Taking $q = 11.4$ and order of integration equal to one (identified in MT), we can then proceed to construct the PADF monthly and annual filters (in the annual case, we consider interpolants $c = k/12$ for

$k = 0, 1, 2, \cdots, 11$ for comparison to monthly trends). Since the series are flow variables with a stock-signal target, we adopt (31) with $\delta = 1/12$ for the monthly case and $\delta = 1$ for the annual case. The upper left panel of Figure 5 shows the two filter coefficients (monthly case), and the discrepancies are very small. The resulting monthly trends have no differences visible to the naked eye, and so are not displayed. (To produce trend values at the beginning and end of the sample, we use the standard device of extending the series by forecasting and backcasting, so we can then apply the symmetric filters to the extended series; for the Local Level Model, the forecast path just amounts to repetition of the initial or terminal value of the smoothed level.)

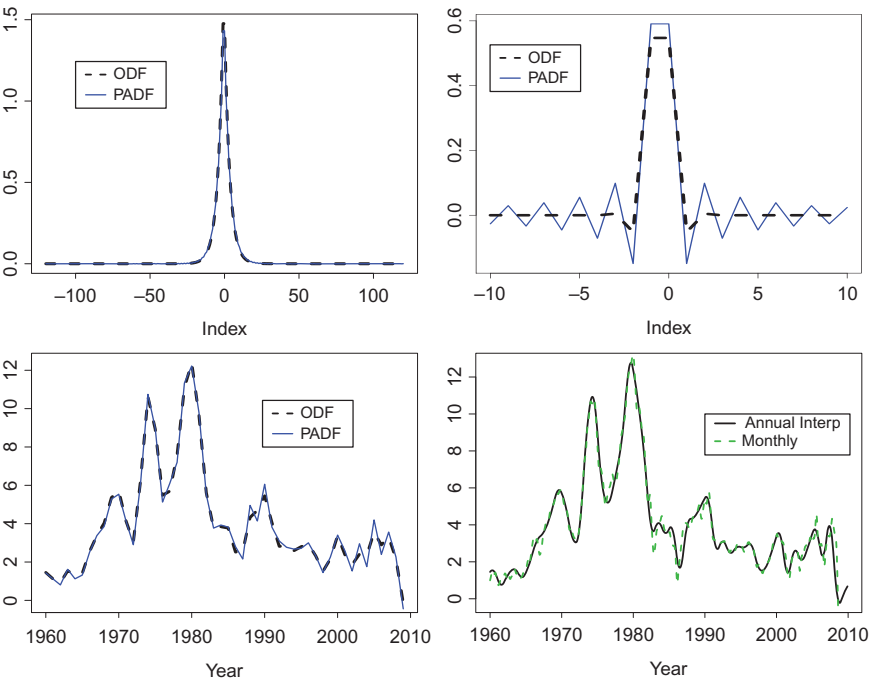

Figure 5: Analysis of the CPI monthly and annual flow series. Upper left panel: monthly filters for the ODF and PADF. Upper right panel: annual filters for the ODF and PADF. Lower left panel: annual trends arising from the ODF and PADF. Lower right panel: monthly trend from PADF, with annual trend interpolated to monthly frequency from PADF.

One can also compare the annual filters for the ODF and the PADF, displayed in the upper right panel of Figure 5. Now there are some notable discrepancies, although the overall shape is largely similar. It is not surprising that there is more contrast, as compared to the monthly case, in view of Proposition 2, because

the discrepancies vanish only as $\delta \to 0$. The resulting annual trends from both methods are plotted in the lower left panel of Figure 5, and there is little overall divergence. Finally, the lower right panel of Figure 5 plots the monthly trend together with the annual trend with monthly interpolants, computed from the PADF method (the same plot, but considering the ODF instead of PADF, is provided in Figure 3 of MT). The close agreement of the trends demonstrates that this method of filter discretization respects sampling frequency, in the sense that the estimated trends for CPI inflation at the two different frequencies are largely coherent, matching one another rather closely on average.

We now give an application to related housing time series available at different sampling frequencies. Two representative series for the housing sector are the monthly time series of US Housing Starts, that is, the total number of New Privately Owned Housing Units Started Starts (Available from the US Census Bureau); and the quarterly time series of US Residential Investment (Available from the Bureau of Economic Analysis), with both series seasonally adjusted and expressed as annual rates. Both series are flows but represent different quantity types, with Starts being a unit count and Residential Investment being a real dollar value (expressed in billions with base year 2005). We are most interested in the period from the early 1990s to present, during which the housing market has undergone a dramatic expansion followed by a virtual collapse, with the most recent years indicating signs of recovery; these major economic developments are reflected in both series. So the sample for Starts is January 1990 through June 2012, while for Residential Investment it extends from the first quarter of 1990 through second quarter of 2012. Suppose we aim to apply low-pass filters in a coherent manner and extract comparable signals across series, perhaps to help convey a consistent picture of trends in construction activity over this time period.

To investigate trends in both indicators together and efficiently use the available data requires coherency across the monthly and quarterly time intervals. Therefore, we use the Low Pass filter with $m = 1$ (Illustration 1 of Section 4), for $q = 1, 10$, and 100, and apply our approximate discretization method. This last value is suggested by the actual estimate of the Butterworth q; in a preliminary analysis of Starts (the higher-frequency series), the Local Level Model can be fitted to the data with a CT signal-to-noise ratio of $\hat{q} = 139.98$. This estimate may seem somewhat large, due partly to the level functioning as a "trend-cycle" – absorbing virtually all the systematic movements, with the remainder just noise that is relegated to the irregular – in this simple model, and partly to the intensity of trend movements over this episode; furthermore, given that q is approximately transformed by the square of δ in moving to a discrete-time filter, a value of 100 actually implies a reasonable balance between level and irregular fluctuations. In any case, moving from the

neighborhood of the fitted value to a smaller ratio, as in the $q = 10$ case allows for a perhaps more attractive trend with greater noise elimination and clearer detection of major transitions or turning points. The $q = 1$ value gives an even smoother signal, though the high degree of stability also means that the trend adapts more slowly to ongoing developments, and indeed, the fit of the associated model deteriorates appreciably for this case. The choice $q = 10$ represents a reasonable balance between fit (the trend-noise model, though optimized in the neighborhood of 100, still performs decently in this intermediate case) and signal properties of smoothness and ease of tracking main developments.

Using $\delta = 1/12, 1/4$, we generate the monthly trend for both Starts and Residential Investment, which includes values interpolated to monthly frequencies for Residential Investment; as with the CPI example, trends are produced at the boundaries of the sample via forecast and backcast extension. Figure 6

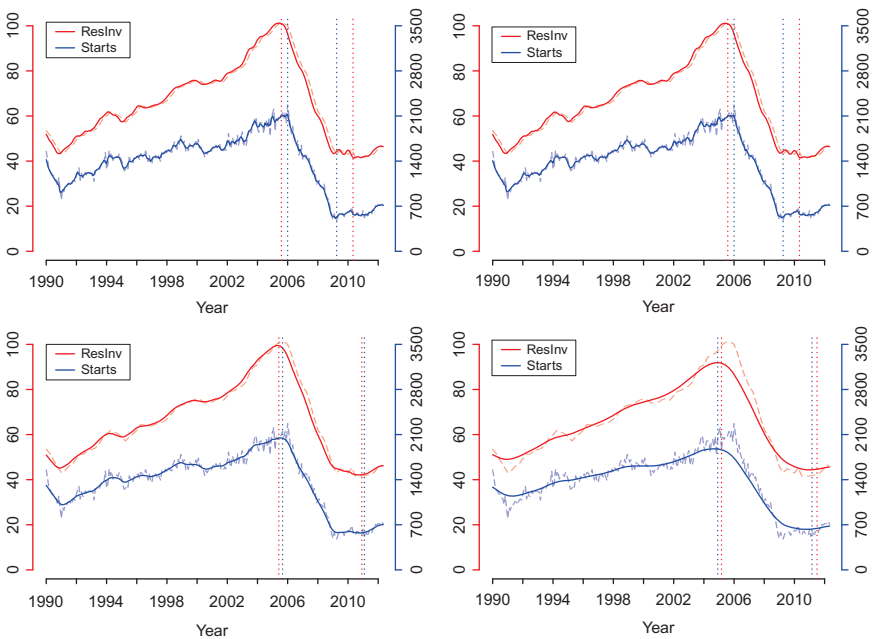

Figure 6: Analysis of monthly Housing Starts (Starts, in blue) and quarterly Residential Investment (ResInv, in red) in diverse scales (left axis for Residential Investment, right axis for Housing Starts), together with PADF trends. Signal-to-noise ratios are 139.98 (top left panel), 100 (top right panel), 10 (bottom left panel) and 1 (bottom right panel), resulting in more smoothing from as the signal-to-noise ratio decreases. Original series are in dashed lines, with trends super-imposed as solid lines. Vertical dotted lines indicate turning points for the begin and end of the recession, as determined by the estimated trends.

displays both series' trends for comparison (note that the series are plotted with different scales, with that of Starts on the right in blue and that of Residential Investment on the left in red), with various panels corresponding to differing choices of q. While both trends show similar movements generally, measured from the same base, their own specific patterns also emerge in the figure. The results for the higher and lower ratios are displayed in the right hand panels of Figure 6, whereas the left hand panels contain the fitted trend and the trend for the intermediate value $q = 10$. While the amount of smoothing varies considerably across the three ratios, the filters automatically adapt via δ to each series, so that the resulting trends maintain their coherency. The reason is that the PADF method automatically targets the same portion of the underlying spectral density in the low-pass formulation of trend extraction – for any choice of the sampling frequency.

8 Conclusion

While most macroeconomic time series represent stock or flow measurement of some underlying process conceived of as occurring in CT, the measurement conditions for the discrete observations affects the extraction of signals that ultimately describe the underlying dynamics. To extract signals of interest from such series, the usual approach in the econometrics literature is to first discretize the underlying CT process – Σ in Figure 1 – followed by application of discrete-time filtering, the mapping Ψ (Harvey and Stock 1993). This paper presents a different approach that instead discretizes the continuous-lag filters – Δ in Figure 1 – allowing for a natural treatment of interpolants. Moreover, our method offers a flexible and coherent strategy for analyzing series sampled in different ways.

We derive explicit formulas that solve the CT signal extraction problem, extending previous results to the case of a nonstationary signal, which is a key assumption for most economic data. The continuous-lag estimators map the signal-to-noise relationship of the underlying process into a relatively compact expression. We present several filters designed for use in real-world applications, including new low-pass continuous-lag filters, as well as derivative and turning point filters for monitoring transitions in signals of interest; optimal filters that target other functions of the signal may be derived in the same way. Our examples demonstrate how the core continuous-lag filters tend to have simpler forms than their discrete-time counterparts, making it convenient to investigate their properties.

Another major contribution of the paper is the Primary Alias method of filter discretization. The formulas for the discretization of a given continuous-lag filter are easy to implement and do not require explicit knowledge of the data process' dynamics, in contrast to the mean square optimal discretization. Here, in contrast, the PADF technique only requires knowledge of the original filter's kernel, and implementation is straightforward. Numerical results show the accuracy of this discretization, and its application to CPI inflation and housing data illustrates the method's coherency and efficacy.

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 (19), it suffices to show that $\{e(t)\}$ is orthogonal to $\{w(t)\}$ and the initial values $\mathbf{y}^*\mathbf{(0)}$. 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 $d \le k < 2d$, 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 $j < d$. We first note that representation (19) 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 Δ_0 is the Dirac delta function. Note that any filter that does not pass polynomials cannot be MSE optimal, since the variance of the error process will grow unboundedly with time. So we have

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

which is orthogonal to $\mathbf{y}^*\mathbf{(0)}$ by Assumption A. Due to the representation (19), it is sufficient to show that the error process is uncorrelated with $[I^d w](t)$. For any real h

$$\mathbb{E}\left[\epsilon(t)\left[I^d w\right](t+h)\right] = \int (\psi(x)-\Delta_0(x))\mathbb{E}\left(\left[I^d u\right](t-x)\left[I^d u\right](t+h)\right)dx + \int \psi(x)\mathbb{E}\left[n(t-x)\left(n(t-h)-\sum_{j=0}^{d-1}\frac{(t+h)^j}{j!}n^{(j)}(0)\right)\right]dx$$

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

$$\mathbb{E}\left[[I^d u](t-x)[I^d u](t+h)\right] = \int_0^{t-x}\int_0^{t+h}\frac{(t-x-r)^{d-1}(t+h-z)^{d-1}}{(d-1)!^2}R_u(r-z)\,dz\,dr. \tag{35}$$

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 (35), 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 (35) 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 ψ to obtain

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

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 (34), 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,

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

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:

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

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 (34) 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 (34), which shows that $\Psi(L)$ is MSE optimal. Using similar techniques, the error spectral density is obtained as well. □

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

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

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. □

Derivation of the Low-Pass Weighting Kernel. We consider extending the frf to the complex plane, written as $f(z) = e^{iz|x|}(1+z^{2m})^{-1}$. The same strategy is used as in the $m=2$ case above, observing that the m poles of $1+z^{2m}$ are of the form $e^{ij\pi/m}$ with j an odd integer between 1 and $4m-1$ Half of these poles occur in the upper half plane, and half in the lower half complex plane. Moreover,

$$\begin{aligned} 1+z^{2m} &= \prod_{k=1}^{m} \left(z^2 - 2\cos[\pi(2k-1)/2m]\, z + 1\right) \\ &= \prod_{k=1}^{m} \left(z - e^{i(2k-1)\pi/2m}\right)\left(z - e^{i(4m-2k+1)\pi/2m}\right), \end{aligned}$$

so that the residue of f at a pole in the upper plane, say $e^{i(2k-1)\pi/2m}$, is

$$\frac{e^{iz|x|}}{\prod_{\ell \neq k} (z^2 - 2\cos[\pi(2\ell-1)/2m]z + 1)(z - e^{i(4m-2k+1)\pi/2m})} \Big|_{z = e^{i(4m-2k+1)\pi/2m}}.$$

Simplifying, and summing over the relevant residues yields

$$\psi(x) = \sum_{k=1}^{m} \frac{\exp\{(-\sin[(2k-1)\pi/2m] + i\cos[(2k-1)\pi/2m])|x|\}}{\prod_{\ell \neq k} (e^{i(2k-1)\pi/m} - 2\cos[\pi(2\ell-1)/2m]\, e^{i(2k-1)\pi/2m} + 1) 2\cos[\pi(2k-1)/m]}.$$

Proof of Proposition 1: First we show that the difference between the two filters has no bias. Letting $a(x) = x^p$ for integer p and $a_\tau = a(\delta\tau)$,

$$\begin{aligned} \Psi(L)a(\delta\tau + \delta c) &= \sum_{j=0}^{p} \binom{p}{j} (\delta\tau)^{p-j} \int_{-\infty}^{\infty} (\delta c - x)^j \psi(x)\, dx \\ \Psi_\delta(B)a_\tau &= \sum_{j=0}^{p} \binom{p}{j} (\delta\tau)^{p-j} \sum_{k=-\infty}^{\infty} \psi_k (-\delta k)^j \end{aligned}$$

follows from binomial expansion. Matching coefficients, we see that a necessary and sufficient condition for similar polynomial treatment is

$$\int_{-\infty}^{\infty} (x - \delta c)^j \psi(x)\, dx = \delta^j \sum_{k=-\infty}^{\infty} \psi_k k^j \qquad [36]$$

for $j = 0, 1, 2, \cdots, p$. In the case of handling $I(d)$ processes, we would impose this condition with $p = d$. But (36) can be compactly expressed in frequency domain as (26) using Fourier Transforms. Using the representation in eq. [5] of MT, any discrete filter that satisfies this condition ensures that the discretization error $\Psi(L)y(\delta\tau + \delta c) - \Psi_\delta(B)y_\tau$ only involves stochastic portions. Next, using

induction and results of Hannan (1970), we can represent an $I(d)$ CT process (with square integrable differentiated process w) as

$$y(t) = \sum_{j=0}^{d-1} \frac{t^j}{j!} y^{(j)}(0) + \int_{-\infty}^{\infty} (i\lambda)^{-d} \left(e^{i\lambda t} - \sum_{j=0}^{d-1} \frac{(i\lambda t)^j}{j!} \right) dZ(\lambda).$$

Typically the initialization values $y^{(j)}(0)$ are random variables assumed to be independent of the differentiated process w, to which the orthogonal increments process $d\mathbf{Z}$ pertains. Stock-sampling the above representation is fairly clear, but note that flow-sampling will result in the factor $(1-e^{-i\lambda\delta})/(i\lambda)$ multiplying the complex exponential. In either the stock or flow case we can apply the discrete-lag and continuous-lag filters and cancel out the deterministic terms, leaving the stated expressions for the discretization error. The expression (27) for the discretization MSE then follows at once. This integral (27) is not guaranteed to be finite, unless there is a suitable degree of decay in f_w or the other integrands (clearly λ^{-2d} assists integrability). Also note that (26) ensures that a suitable number of zeroes occur in the integrand at frequency zero, to offset the explosive behavior of λ^{-2d} at $\lambda = 0$. □

Proof of Proposition 2: We first establish the decomposition of the discretization MSE. Using the notation $[f]_\delta(\lambda) = \delta^{-1} \sum_{h=-\infty}^{\infty} f(\lambda + 2\pi h/\delta)$ for the fold of the function f (cf., MT), the stock ODF is given by the formula $u_\delta = [g e_c f_w m_{2d}]_\delta / [f_w m_{2d}]_\delta$, where $m_j(\lambda) = \lambda^{-j}$. Then (28) follows from $[u_\delta f_w m_{2d}]_\delta = [g e_c f_w m_{2d}]_\delta$, which holds due to a property of folds, implying that the cross-terms are zero. Also the total PADF discretization MSE can be rewritten as $\frac{1}{\pi\delta}\int_0^\pi K(\lambda/\delta)d\lambda$, where

$$K(x) = \sum_{h \neq 0} |g(x + 2\pi h/\delta) e_c(x + 2\pi h/\delta) - g(x) e_c(x)|^2 f_w(x + 2\pi h/\delta)(x + 2\pi h/\delta)^{-2d},$$

which is convenient for numerical computation. The integral expression for (27) is easily approximated via a Riemann sum. The ODF discretization MSE can be computed using the formula (also discussed in MT)

$$\frac{\delta}{2\pi} \int_{-\pi/\delta}^{\pi/\delta} \left([g^2 f_w m_{2d}]_\delta - |[g e_c f_w m_{2d}]_\delta|^2 / [f_w m_{2d}]_\delta \right)(\lambda)\, d\lambda.$$

This too can be rewritten as $\frac{1}{\pi\delta}\int_0^\pi H(\lambda/\delta)d\lambda$, where

$$H(x) = \delta \left([g^2 f_w m_{2d}]_\delta(x) - |[g e_c f_w m_{2d}]_\delta(x)|^2 / [f_w m_{2d}]_\delta(x) \right).$$

This can also be computed via a Riemann approximation; of course, computation of these quantities require a knowledge of the true spectrum f_w, and thus is a theoretical exercise. Hence the PADF discretization MSE equals

$$\frac{1}{\pi\delta}\int_0^{\pi} K(\lambda/\delta)d\lambda = \frac{1}{\pi\delta}\int_0^{\pi} H(\lambda/\delta)d\lambda + \frac{1}{\pi\delta}\int_0^{\pi} [K-H](\lambda/\delta)d\lambda,$$

which decomposes the error in terms of the ODF discretization MSE and the extra MSE due to using a sub-optimal discretization.

Now for the main assertion of the proposition, it suffices to show that the ratio of $\int_0^{\pi/\delta}[K-H](\lambda)d\lambda$ to $\int_0^{\pi/\delta} H(\lambda)d\lambda$ tends to zero as $\delta \to 0$. Let us write that $g(\lambda) = O(|\lambda|^{-\beta})$ for some $\beta \geq 0$, which is always possible because g is a bounded function. Also note that m_{2d} has tails of order $2d$. The integrand of the denominator equals

$$\sum_{h \neq l} (gf_w m_{2d})(\lambda + 2\pi h/\delta) \cdot (f_w m_{2d})(\lambda + 2\pi \ell/\delta) \cdot [g(\lambda + 2\pi h/\delta) - g(\lambda + 2\pi \ell/\delta)],$$

divided by $\sum_h (f_w m_{2d})(\lambda + 2\pi h/\delta)$, which tends to $f_w(\lambda)\lambda^{-2d}$ as $\delta \to 0$. Now the terms that decay to zero slowest in the above double summation occur when either h or ℓ is zero; if both h and ℓ are nonzero (and they don't equal each other), the corresponding summands will decay more rapidly in δ due to our tail assumptions on f_w and g. Therefore we can focus on

$$(f_w m_{2d})(\lambda) \cdot \sum_{h \neq 0} (gf_w m_{2d})(\lambda + 2\pi h/\delta) \cdot [g(\lambda + 2\pi h/\delta) - g(\lambda)]$$

$$+ (gf_w m_{2d})(\lambda) \cdot \sum_{\ell \neq 0} (f_w m_{2d})(\lambda + 2\pi \ell/\delta) \cdot [g(\lambda) - g(\lambda + 2\pi \ell/\delta)].$$

In the first term, each summand is $O(\delta^{2\beta+\alpha+2d})$, whereas in the second term each summand is $O(\delta^{\beta+\alpha+2d})$; overall, the highest order term is of order $\delta^{\beta+\alpha+2d}$. Now consider the function $[K-H](\lambda)$; this is the square of

$$\sum_{h \neq 0} [g(\lambda + 2\pi h/\delta) - g(\lambda)] \cdot (f_w m_{2d})(\lambda + 2\pi h/\delta),$$

again divided by $\sum_h (f_w m_{2d})(\lambda + 2\pi h/\delta)$. When the square is expanded, every summand in the double sum is $O(\delta^{2\beta+2\alpha+4d})$. Now we know the order of growth of both integrands, which is all that matters by the Dominated Convergence Theorem. So long as $\beta + \alpha + 2d \geq 0$, the PADF MSE is asymptotic to the minimal MSE.

This completes the stock case. For the flow case, the ODF has frf

$$u_\delta = i[ge_c f_w m_{2d+1}]_\delta \left(1 - e^{-i\lambda\delta}\right)^{-1} / [f_w m_{2d+2}]_\delta,$$

so that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} |g(\lambda)e_c(\lambda)/a_\delta(\lambda) - \bar{u}_\delta(\lambda)|^2 |1 - e^{-i\lambda\delta}|^2 f_w(\lambda)\lambda^{-2d-2} d\lambda$$

$$+ \frac{1}{2\pi}\int_{-\infty}^{\infty} |\bar{u}_\delta(\lambda) - \bar{g}_\delta(\lambda)|^2 |1 - e^{-i\lambda\delta}|^2 f_w(\lambda)\lambda^{-2d-2} d\lambda.$$

The first term is the ODF discretization MSE, whereas the second term is the extra MSE due to using a suboptimal discretization. As in the stock case, the total error is the integral of a function K, whereas the lower bound on the error is given by the integral of a function H. In contrast to the stock case, H is given by

$$H(x) = \delta\Big([g^2 f_w m_{2d}]_\delta(x) - |[g e_c f_w m_{2d+1}]_\delta(x)|^2 / [f_w m_{2d+2}]_\delta(x) \Big).$$

in the flow case. Also, the total error in the flow case is the integral of

$$K(x) = \sum_{h \neq 0} |g(x + 2\pi h/\delta) e_c(x + 2\pi h/\delta)(x + 2\pi h/\delta)$$
$$- g(x) e_c(x) x|^2 f_w(x + 2\pi h/\delta)(x + 2\pi h/\delta)^{-2d-2}.$$

As in the case of a stock-sampled series, the MSE depends explicitly on c. With these derivations, the analysis of the ratio of PADF to ODF discretization MSE follows along the same lines as for the stock case. □

References

Ahlfors, L. 1979. *Complex Analysis*. New York: McGraw-Hill.

Alexandrov, T., S. Bianconcini, E. Dagum, P. Maass, and T. McElroy. 2012. "The Review of Some Modern Approaches to the Problem of Trend Extraction." *Econometric Reviews* 31:593–624.

Bell, W. 1984. "Signal Extraction for Nonstationary Time Series." *Annals of Statistics* 12: 646–64.

Bell, W., and S. Hillmer. 1984. "Issues Involved with the Seasonal Adjustment of Economic Time Series." *Journal of Business and Economics Statistics* 2:291–320.

Bergstrom, A. R. 1988. "The History of Continuous-Time Econometric Models." *Econometric Theory* 4:365–83.

Brockwell, P. 1995. "A Note on the Embedding of Discrete-Time ARMA Processes." *Journal of Time Series Analysis* 16:451–60.

Brockwell, P. 2001. "Lévy-Driven CARMA Processes." *Annals of the Institute of Statistical Mathematics* 53:113–24.

Brockwell, P. 2004. "Representations of Continuous-Time ARMA Processes." *Journal of Applied Probability* 41:375–82.

Brockwell, P., and T. Marquardt. 2005. "Lévy-Driven and Fractionally Integrated ARMA Processes with Continuous Time Parameter." *Statistica Sinica* 15:477–94.

Clark, P. 1987. "The Cyclical Component of U.S. Economic Activity." *The Quarterly Journal of Economics* 102:797–814.

Folland, G. 1995. *Introduction to Partial Differential Equations*. Princeton: Princeton University Press.

Hannan, E. 1970. *Multiple Time Series*. New York: Wiley.

Harvey, A. 1985. "Trends and Cycles in Macroeconomic Time Series." *Journal of Business Economic Statistics* 3:216–27.

Harvey, A. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, A., and J. Stock. 1993. "Estimation, Smoothing, Interpolation, and Distribution for Structural Time-Series Models in Continuous Time." In *Models, Methods and Applications of Econometrics*, edited by P. C. B. Phillips, 55–70. Oxford: Blackwell.

Harvey, A., and T. Trimbur. 2003. "General Model-Based Filters for Extracting Cycles and Trends in Economic Time Series." *Review of Economics and Statistics* 85:244–55.

Harvey, A., and T. Trimbur. 2007. "Trend Estimation, Signal-Noise Rations and the Frequency of Observations." In *Growth and Cycle in the Eurozone*, edited by G. L. Mazzi and G. Savio, 60–75. Basingstoke: Palgrave MacMillan.

Hodrick, R., and E. Prescott. 1997. "Postwar U.S. Business Cycles: An Empirical Investigation." *Journal of Money, Credit, and Banking* 29:1–16.

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

Koopmans, L. 1974. *The Spectral Analysis of Time Series*. New York: Academic Press.

McElroy, T. 2013. "Forecasting Continuous-Time Processes with Applications to Signal Extraction." *Annals of the Institute of Statistical Mathematics* 65:439–56.

McElroy, T., and T. Trimbur. 2011. "On the Discretization of Continuous-Time Filters for Nonstationary Stock and Flow Time Series." *Econometric Reviews* 30:475–513.

Priestley, M. 1981. *Spectral Analysis and Time Series*. London: Academic Press.

Ravn, M., and H. Uhlig. 2002. "On Adjusting the HP Filter for the Frequency of Observation." *Review of Economics and Statistics* 84:371–80.

Watson, M. 1986. "Univariate Detrending Methods with Stochastic Trends." *Journal of Monetary Economics* 18:49–75.

Whittle, P. 1983. *Prediction and Regulation*. Oxford: Blackwell Publishers.