Coherent Trends, Turning Points, and Forecasts for ACS Data

Tucker McElroy
U.S. Census Bureau

Abstract

The American Community Survey (ACS) provides one-year (1y), three-year (3y), and five-year (5y) multi-year estimates (MYEs) of various demographic and economic variables for each "community," although for small communities the 1y and 3y may not be available. These survey estimates are not truly measuring the same quantities, since they each cover different time spans. We present a metric to measure the compatibility of different MYEs; for those that are deemed to be sufficiently compatible, we describe methods for generating trends, turning points, and forecasts of ACS data at 1y, 3y, and 5y intervals, in such a way that the different estimates can be compared with one another. The filters utilized are non-model-based, require only a short span of data, and are designed to preserve the appropriate linear characteristics of the time series that are relevant for trends, turning points, and forecasts respectively. The basic method, which only requires polynomial algebra, is outlined and applied on ACS data. The resulting filters are analyzed in the frequency domain.

Keywords. Filtering, Frequency Domain, Nonstationary time series.

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

1 Introduction

The American Community Survey (ACS) replaces the former Census Long Form, providing timely estimates available throughout the decade. The ACS sample size is comparable to that of the Census Long Form; variability in the sampling error component of the ACS is partially reduced through a rolling sample (Kish, 1981). The rolling sample refers to the pooling of sample respondents over time – in some cases this may be viewed as an approximate temporal moving average of single period estimates. In particular, estimates from regions with at least $65,000$ people are produced with a single year of data, whereas if the population is between $20,000$ and $65,000$ then three years of data are combined, and if the population is less than $20,000$ then five years of data are pooled.

A somewhat dated overview of the ACS can be found in Alexander (1998). More current details can be found in Census Bureau (2006).

The ACS database for the Multi-Year Estimates Study (MYES) is now publicly available at

$$www.census.gov/acs/www/AdvMeth/Multi\_Year\_Estimates/online\_data\_year.html.$$

The MYES produces one, three, and five year estimates for counties included in the 1999 – 2001 demonstration period and their constituent geographies, using data from 1999 through 2005. One selects a Multi-Year Estimate (MYE) category – either one-year (1y), three-year (3y), or five-year (5y) – then a time period, then a county, and then a geographic type within the county (e.g., school district). Then the user can select from four types of information: demographic, economic, social, and housing. Within each of these categories are dozens of variables, which for the most part are simple totals or averages, but some of which are more complicated statistics (e.g., medians and percentiles).

Because some counties have a low population, it was deemed desirable by the U.S. Census Bureau to decrease sampling error for smaller geographies and subpopulations by using a rolling sample; this methodology was confirmed by the National Academy of Sciences Panel on the Functionality and Usability of Data from the American Community Survey (Citro and Kalton, 2007). In essence, responses over a 3y or even a 5y span are gathered together into one database, and a statistic of interest is computed over the temporally enlarged sample. In many cases, this is approximately equal to computing a simple moving average of 1y estimates (see Section 3 below). This is known as a rolling sample – see Kish (1981, 1998) and Alexander (2001) for a discussion. For larger counties, the 1y MYE would be available as well. The question of whether each year should be equally weighted was addressed by Breidt in a report to NAS (Citro and Kalton, 2007); since all the responses are pooled in the 3y and 5y cases, it was judged to be impractical to use some alternative weighting scheme (such as weighting the most recent year of data more highly). Hence, the MYEs are formed from contributions over multiple years that are *equally* weighted. Although this approach is simple, one repercussion is that some lag (or time delay) is induced by the use of rolling samples.

The effect of this time delay is that the 3y estimate is roughly speaking an estimate of the previous year's value, and the 5y estimate gives the value corresponding to two years back. For example, putting a straight line through a 3-year and 5-year moving average viewed as a concurrent filter will result in a line lagged by 1 and 2 time points respectively. Because for many regions and statistics the ACS data will be trending upwards in a local linear fashion (this is dictated by macroeconomic theory and demographic principles at work), the 3y and 5y estimates will tend to be lagging behind the 1y values. Admittedly this effect is much less pronounced for characteristics reported as percentages, versus the actual levels. This has repercussions for cross-county comparisons, which is seen as follows. Suppose that county A has a small population, so we only have a 5y estimate. But county B has a large population, so we have 1y, 3y, and 5y available. Since the 1y is at hand for county B, attention may well focus on it and the 5y will be largely ignored. It will then be natural for a user to compare the 1y MYE for county B with the 5y MYE for county A. But this is comparing apples and oranges; it would be valid, roughly speaking, to compare the 1y MYE for county B *two years back* to the 5y MYE for county A (or alternatively, to take a five year moving average of the 1y MYE for county B). Alexander (2001) provides additional discussion.

These subtleties, although easily illustrated through a few pictures, cannot be assumed to be comprehensible to the average consumer of ACS data. The dangers of deliberate or unintentional abuse seem to be great. Therefore, a method that produces trends (as well as turning points and forecasts) from available data should be developed, such that the trends (respectively, turning points and forecasts) all generate the *same* value for 1y, 3y, and 5y estimates. That is, we have three separate trend filters, for each of the 1y, 3y, and 5y estimates, such that the output of all three operations are approximately identical. It will then be mathematically valid to compare a county B 1y trend with a county A 5y trend, for example. Such trends, turning points, and forecasts could be either published along with the ACS data, or could be available in a small portable program with a GUI front-end (e.g., R with JAVA interface) that the average consumer could use with a few clicks of the mouse. The first option has the advantage of pre-empting the ability of the public user to misuse the data. A similar thing is done in the arena of seasonal adjustment, where the U.S. Census Bureau strongly encourages consumers to base all analysis upon the seasonally adjusted data rather than the raw (see the discussion in Bell and Hillmer, 1984).

In the scheme of trend filter design, there are two general approaches: the "window trend" and the "on-line trend." The former takes a span of data $X_1, \cdots, X_N$ and produces a trend value $S_1, \cdots, S_N$ for each time point. The latter may utilize less data, and produces a trend only at time N. The window trend method will typically produce trends that depend on the size of the window (since the filters that are used depend on this length). This can be a practical drawback, which we will illustrate. If we obtain an additional datum X_{N+1}, then we can obtain new trend estimates, based on either the window $X_1 \cdots X_{N+1}$ or $X_2 \cdots X_{N+1}$ (depending on whether a fixed window or expanding window is desired). Either way, the new trend estimate at time N will generally be different from the previous value S_N. This is known as the problem of revisions in the seasonal adjustment literature. A publishing agency would then have the onerous task of updating the trend estimate at time N (and the updating can continue indefinitely, although they generally tend to stabilize). This is onerous, because the public invariably complains that the agency is "changing the numbers again."

In contrast, the on-line approach requires no updating, since a trend is only produced at the current time point (using one fixed concurrent filter). Although the trend values will then be consistent, there is invariable phase delay (i.e., lag in the trend estimates). Generally there is a trade-off between smoothness and lag, which represents an inherent scientific limitation; the acquisition of a zero-lagging smoother would be an incredibly powerful tool in economic and demographic analysis. Nevertheless, careful design of concurrent filters may help to strike a proper balance between smoothness and lag. Due to the greater possibilities of user confusion – and the fact that the greater lag inherent in concurrent filtering can be mitigated – we advocate the on-line approach.

From this discussion, we have gathered several criteria for ACS "signals"; by signals, we refer to trends, turning points, forecasts, and potential additional applications.

1. Coherency between signals of 1y, 3y, and 5y estimates
2. Proper treatment of linear dynamics
3. Concurrent filtering with minimal phase-delay
4. Filter length is short

Item 1 has been discussed above, and will be formulated mathematically below. As for the linear dynamics in 2, we mean that underlying local linear behavior in the time series data should be appropriately preserved. For trends, this means that the exact linear polynomial should be unaltered by the filtering; for turning points, the sign of the slope coefficient of the line should be transmitted (so that a negative slope indicates a down-turn, whereas a positive slope indicates an up-turn). In forecasting, the line should actually be advanced one step in time. Each of these three concepts can be given a precise mathematical formulation. Item 3 indicates that we adopt the on-line filtering approach, and seek to mitigate the effects of lag. This is related to item 2, since in producing trends we seek a filter that is the identity operator on linear polynomials, and thus the lag at "frequency zero" will be automatically reduced. A full discussion of this important topic requires a frequency domain analysis of the signal filters. Finally, point 4 is of practical importance; since we use on-line filtering, we must have the signals produced from a minimal span of data. If the filter consists of d weights, then d data points are required to produce the signal, i.e., data $X_{N-d+1}, \cdots, X_N$ are required for the signal S_N. Thus no signal will ever be available for $X_1, \cdots, X_{d-1}$; we therefore seek to minimize d.

The remainder of this paper is structured as follows. In Section 2 we discuss the basic approach to constructing the signal filters through using polynomial algebra; the specific cases of trends, forecasts, and turning points are considered in Sections 2.1 through 2.3. Section 3 provides a way of assessing compatibility between various MYEs, essentially providing the conditions for the validity of Section Two's methodology. Then Section 4 illustrates the methods on some actual publicly available ACS data. Finally in Section 5 the properties of the various filters are explored in the frequency domain through squared gain and phase delay plots. A discussion of the applications is given in Section 6, and some derivations are contained in the Appendix.

2 Mathematical Formulation and Solution

In this section we assume that the 3y and 5y estimates are exactly equal to simple moving averages of the 1y data. We write the k-year estimate at year t as $Y_t^{(k)}$, where

$$Y_t^{(k)} = \Theta^{(k)}(B) X_t \tag{1}$$

and $k = 1, 3, 5$ (B is the backshift operator). Here X_t is the 1y estimate (which is not directly observable for small geographical areas), and $\Theta^{(k)}(z)$ is the Simple Moving Average (SMA) polynomial of order k given by

$$\Theta^{(k)}(z) = \frac{1}{k}\left(1 + z + \cdots + z^{k-1}\right).$$

Now (1) may be only approximately true in reality, but in many cases we may assume that (1) holds with a small amount of error; for more discussion of this assumption see Section 3. Now for a signal of interest S_t, we look for three filters $\Psi^{(k)}(B)$ for $k = 1, 3, 5$ such that

$$S_t = \Psi^{(k)}(B) Y_t^{(k)} \qquad k = 1, 3, 5. \tag{2}$$

This ensures condition 1, namely that we can filter the 1y, 3y, and 5y estimates with different filters in order to obtain a common signal S_t. This signal must have the characteristics dictated by its definition, and this is where conditions 2 and 3 enter. Because we are doing concurrent filtering, we can write

$$\Psi^{(k)}(z) = \sum_{j \geq 0} \psi_j^{(k)} z^j.$$

In practice only a finite number of the coefficients $\psi_j^{(k)}$ are nonzero. The specifics of condition 2 are discussed in the following subsections. Now combining (2) and (1) yields

$$\Psi^{(1)}(z) = \Psi^{(3)}(z)\Theta^{(3)}(z) = \Psi^{(5)}(z)\Theta^{(5)}(z). \tag{3}$$

Since $\Theta^{(3)}(z)$ and $\Theta^{(5)}(z)$ are known ahead of time, we only need determine $\Psi^{(3)}(z)$ and $\Psi^{(5)}(z)$; then $\Psi^{(1)}(z)$ is determined.

Now focusing on trends, we seek concurrent filters $\Psi^{(k)}$ such that the composite with the SMA passes lines, i.e.,

$$\Psi^{(k)}(B)\Theta^{(k)}(B)\ [at+b] = at+b \tag{4}$$

for any a, b and all integers t, and $k = 1, 3, 5$. This property, together with those discussed above, define the optimal trend filter for our purposes. The following theorem provides explicit formulas for the optimal trend filters.

Theorem 1 *The minimal length concurrent filters $\Psi^{(k)}$ satisfying (3) and (4) are given by*

$$\begin{aligned}
\Psi^{(5)}(z) &= \left(4 + z + z^2 - 3z^3\right)/3 \\
\Psi^{(3)}(z) &= \left(4 + z + z^2 + z^3 + z^4 - 3z^5\right)/5 \\
\Psi^{(1)}(z) &= \left(4 + 5z + 6z^2 + 3z^3 + 3z^4 - z^5 - 2z^6 - 3z^7\right)/15.
\end{aligned}$$

Next focusing on 1-step ahead forecasting, we seek concurrent filters $\Psi^{(k)}$ such that the composite with the SMA advances lines by one time step, i.e.,

$$\Psi^{(k)}(B)\Theta^{(k)}(B)\ [at+b] = a(t+1)+b \tag{5}$$

for any a, b and all integers t, and $k = 1, 3, 5$. This property defines the optimal forecast filter, and the following theorem provides explicit formulas:

Theorem 2 *The minimal length concurrent filters $\Psi^{(k)}$ satisfying (3) and (5) are given by*

$$\begin{aligned}
\Psi^{(5)}(z) &= \left(5 + z + z^2 - 4z^3\right)/3 \\
\Psi^{(3)}(z) &= \left(5 + z + z^2 + z^3 + z^4 - 4z^5\right)/5 \\
\Psi^{(1)}(z) &= \left(5 + 6z + 7z^2 + 3z^3 + 3z^4 - 2z^5 - 3z^6 - 4z^7\right)/15.
\end{aligned}$$

Finally, consider turning point (t.p.) filters $\Psi^{(j)}$ for $j = 1, 3, 5$. A t.p. estimate should indicate a change in the overall direction of the data; a crude t.p. signal can be generated by differencing consecutive trend values. Since we wish to anticipate turning points, we take the difference between the forecast filter and the trend filter, which yields the following for the $5y$, $3y$, and $1y$ t.p.'s respectively:

$$\begin{aligned}
\Psi^{(5)}(z) &= \left(1 - z^3\right)/3 \\
\Psi^{(3)}(z) &= \left(1 - z^5\right)/5 \\
\Psi^{(1)}(z) &= \left(1 + z + z^2 - z^5 - z^6 - z^7\right)/15.
\end{aligned} \tag{6}$$

This last t.p. filter is just $(1-z)\Theta^{(3)}(z)\Theta^{(5)}(z)$. These filters satisfy conditions 1, 2, and 4.

3 A Compatibility Measure

The above development relies on (1), which in reality is only approximately true. For a 3y MYE, a sample of people is gathered over all three years and their responses are pooled together. For the 5y MYE the sample is gathered over five years. Then for whatever characteristic is being estimated from the sample, it will be rescaled by the sampling frequency. Thus in comparing a 3y MYE and 5y MYE for the same time period and area, we might expect (1) to hold true. However, for some characteristics we found large discrepancies between the different MYEs, and it was observed that (1) fails. The validity of (1) depends most upon what type of characteristic is being measured: for totals and averages (e.g., total population) the approach of Section 2 was largely valid, whereas for other estimates (e.g., median income) we saw that (1) was false. Thus it is necessary to devise a "compatibility measure" for each characteristic (in each region) that assesses the validity of our main approach. Consider the following modification of (1):

$$Y_t^{(k)} = \Theta^{(k)}(B) X_t + \epsilon_t^{(k)} \tag{7}$$

for $k = 1, 3, 5$. Here $\epsilon_t^{(k)}$ represents an error process, but $\epsilon_t^{(1)} = 0$ by assumption (there is no loss of generality in assuming this). Now we will develop the exposition under the supposition that MYE's are available for $k = 1, 3, 5$; if only 3y and 5y MYE's are available, then set $\epsilon_t^{(3)} = 0$ and ignore the $k = 1$ equation. Likewise, if only a 5y MYE is available, then there is no compatibility issue, since there is only one type of estimate. Now applying a trend, t.p., or forecasting filter $\Psi^{(k)}(B)$ yields

$$\Psi^{(k)}(B) Y_t^{(k)} = \Psi^{(1)}(B) X_t + \Psi^{(k)}(B) \epsilon_t^{(k)}. \tag{8}$$

On the left is the estimate, whereas $\Psi^{(1)}(B) X_t$ should be viewed as our target, or signal. The associated error process is $\Psi^{(k)}(B)\epsilon_t^{(k)}$. Our objective is to quantify the size of $\epsilon_t^{(k)}$ in order to understand its impact on the error process $\Psi^{(k)}(B)\epsilon_t^{(k)}$. The closer $\epsilon_t^{(k)}$ is to zero, the more valid the methodology of Section 2 will be. The basic idea is to consider a "noise-signal" ratio

$$\frac{\epsilon_t^{(k)}}{\Theta^{(k)}(B) X_t} = \frac{Y_t^{(k)}}{\Theta^{(k)}(B) X_t} - 1.$$

This is only well-defined when $\Theta^{(k)}(B) X_t$ is nonzero, and we generally suppose that it is positive at all times. Since it is more convenient we will instead work with the first-order Taylor series approximation of the "noise-signal" ratio, which is given by

$$NSR_t^{(k)} = \log Y_t^{(k)} - \log \Theta^{(k)}(B) X_t.$$

The NSR notation stands for Noise-Signal Ratio. For $k = 1$ it is trivially equal to zero, but for $k = 3, 5$ it may be positive or negative. In order to compute it, we substitute $X_t = Y_t^{(1)}$. Now this quantity shall be calculated over all times t for which it is possible to do so, and then we let the compatibility measure be defined by

$$C^{(k)} = \max_t |NSR_t^{(k)}|.$$

Clearly $C^{(k)}$ will get updated as more data is available. Now it follows from the definitions that

$$\left(e^{-C^{(k)}} - 1\right) \Theta^{(k)}(B) X_t \leq \epsilon_t^{(k)} \leq \left(e^{C^{(k)}} - 1\right) \Theta^{(k)}(B) X_t.$$

Applying this to (8) yields

$$e^{-C^{(k)}} \Psi^{(1)}(B) X_t \leq \Psi^{(k)}(B) Y_t^{(k)} \leq e^{C^{(k)}} \Psi^{(1)}(B) X_t.$$

This inequality shows us how to set tolerance thresholds for the compatibility measure, e.g., $C^{(k)} = \log(1.01)$ (approximately one percent discrepancy). When the compatibility measure (for both $k = 3$ and $k = 5$) is below the threshold, then we can apply the methods of Section 2, confident in the knowledge that the error process is very small.

4 Illustrations on ACS Data

As of the time of writing of this article, the following MYEs are available: 00–05 for 1y, 01–05 for 3y, and 03–05 for 5y. The year index here refers to the last year that entered into the sample, and so is consistent with our notation for $Y_t^{(k)}$. Letting t range between 00 and 05 (referring to the year), the available database is $Y_{00}^{(1)}, \cdots, Y_{05}^{(1)}, Y_{01}^{(3)}, \cdots, Y_{05}^{(3)}, Y_{03}^{(5)}, \cdots, Y_{05}^{(5)}$. In order to apply our methods, we need 8 years of 1y data, 6 years of 3y data, and 4 years of 5y data. So we need two more $1y$ estimates, and one more of the $3y$ and $5y$. So in order to apply our method, we need to extend our database.

We first obtain a backcast estimate for $Y_{99}^{(1)}$ using the supposed compatibility of the MYEs. If (1) were true, then

$$Y_{99}^{(1)} = 3Y_{01}^{(3)} - \left(Y_{01}^{(1)} + Y_{00}^{(1)}\right) = 5Y_{03}^{(5)} - \left(Y_{03}^{(1)} + Y_{02}^{(1)} + Y_{01}^{(1)} + Y_{00}^{(1)}\right).$$

So these provide two estimates for $Y_{99}^{(1)}$, which can be averaged for better precision. Thus

$$\widehat{Y}_{99}^{(1)} = \frac{3}{2} Y_{01}^{(3)} + \frac{5}{2} Y_{03}^{(5)} - \frac{1}{2}\left(Y_{03}^{(1)} + Y_{02}^{(1)} + 2Y_{01}^{(1)} + 2Y_{00}^{(1)}\right).$$

Next, in order to obtain estimates for $Y_{06}^{(1)}$, $Y_{06}^{(3)}$, and $Y_{06}^{(5)}$ (we choose to forecast rather than backcast, because our projections can be verified in a year's time), we apply a very simple random walk model to obtain

$$
\begin{aligned}
\widehat{Y}_{06}^{(1)} &= Y_{05}^{(1)} + \frac{1}{5}\left(Y_{05}^{(1)} - Y_{00}^{(1)}\right) \\
\widehat{Y}_{06}^{(3)} &= Y_{05}^{(3)} + \frac{1}{4}\left(Y_{05}^{(3)} - Y_{01}^{(3)}\right) \\
\widehat{Y}_{06}^{(5)} &= Y_{05}^{(5)} + \frac{1}{2}\left(Y_{05}^{(5)} - Y_{03}^{(5)}\right).
\end{aligned}
$$

Note that we cannot use the forecast filter to get these future estimates, since that supposes that 8 1y data points are available. So with these crude forecasts, we can now apply our filters for trends, forecasts, and t.p.'s.

We consider five examples out of a plethora, which were chosen to illustrate various aspects of this research. First consider the Mean Travel Time variable for Bronx, NYC, NY, referred to as Travel. We first note the compatibility measures for this variable were $C^{(3)} = .071\%$ and $C^{(5)} = .043\%$. This indicates a very small error in (7), so that we can expect our filtering methodology to be highly valid. The extended data-set for Travel is given in Table 1; we can see the mismatch and lag between the various years just by eye-balling the chart along the various sub-diagonals (i.e., consider 1y 01, 3y 02, 5y 03, and 1y 02, 3y 03, 5y 04, and so forth).

In Table 2 all of the compatibility measures are given (in percentages) for the five examples, along with trend, forecast, and t.p. estimates for year 2006 (though the 2006 forecast should be interpreted as a trend estimate for year 2007). These examples are: Divorced Males for Lake County, IL (Divorce); Mean Travel Time for Bronx, NYC, NY (Travel); Median Income for Pima, AZ (Income); Median Age for Hampden, MA (Age); Rental Vacancy Rate for Madison, MI (Rent). The 1y, 3y, and 5y estimates of trend, forecast, and t.p. respectively are all quite similar when the compatibility measure is low, as in Divorce, Travel, and Age. The forecasts in these cases seem reasonable, but ultimately cannot be checked until next year. For Divorce, Rent, and Income the compatibility is progressively worse, and the signal estimates have corresponding discrepancies. The 3y trend, forecast, and t.p. for Income is especially noteworthy – the t.p.'s sign is even different! We observe that in this case the series is based off of a median of incomes, and thus it is understandable that the compatibility breaks down.

5 Frequency Domain Properties

The definitions of the Gain and Phase and Phase-Delay functions of a filter $\Psi(B)$ can be found in Findley and Martin (2006); this reference explains how a continuous Phase function can be constructed by allowing the Gain function to be negative. We follow this convention here. That is, we have the following decomposition:

$$\Psi(e^{-i\lambda}) = G(\lambda)\, \exp\{i\Phi(\lambda)\}.$$

Here G and Φ are continuous real functions, and are called the Gain and Phase functions, respectively. We also have the Phase-Delay function $\Upsilon(\lambda) = \Phi(\lambda)/\lambda$ when $\Phi(\lambda)/\lambda|_{\lambda=0}$ is finite; its interpretation is the amount (in time units) of delay or lag that each frequency of the data inherits from application of the filter. Note that this interpretation is valid for stationary time series, but must be adjusted somewhat when considering random walks or other nonstationary time series. Likewise, the Gain function shows how much attenuation of stochastic variance each frequency in the data inherits from application of the filter. For example, a stationary process can be written as an orthogonal increments integral

$$X_t = \int_{-\pi}^{\pi} e^{i\lambda t} d\mathbb{Z}(\lambda),$$

where $d\mathbb{Z}$ is an orthogonal increments random measure. Application of the filter $\Psi(B)$ produces

$$\Psi(B)X_t = \int_{-\pi}^{\pi} e^{i\lambda t}\Psi(e^{-i\lambda}) d\mathbb{Z}(\lambda) = \int_{-\pi}^{\pi} e^{i\lambda(t-\Upsilon(\lambda))} G(\lambda) d\mathbb{Z}(\lambda).$$

From this, we see the effect of the Phase Delay at time t (high values of $\Upsilon(\lambda)$ create a lag) and the Gain function, which scales the original random measure $d\mathbb{Z}$. For a nonstationary time series, a modified orthogonal increments representation is required and the above formulation does not hold true. Supposing that the process X_t is given by a Generalized Random Walk (GRW), we can write

$$X_t = X_0 + \int_{-\pi}^{\pi} \frac{e^{i\lambda t} - 1}{1 - e^{-i\lambda}} d\mathbb{Z}(\lambda)$$

for $t = 1, 2, \cdots$. This is a very general once-integrated stochastic process, and is plausible as a crude model for much of the ACS demographic and economic data. Application of a filter $\Psi(B)$ produces

$$\Psi(B)X_t = \Psi(1)X_0 + \int_{-\pi}^{\pi} \frac{e^{i\lambda(t-\Upsilon(\lambda))}G(\lambda) - \Psi(1)}{1 - e^{-i\lambda}} d\mathbb{Z}(\lambda).$$

So for the GRW, Υ and G have a somewhat different effect, although it is still measurable. Observe that for trend and forecast filters, $\Psi(1) = 1$. Thus the integrand in the orthogonal increments integral essentially undergoes the transformation

$$\frac{e^{i\lambda t} - 1}{1 - e^{-i\lambda}} \mapsto \frac{e^{i\lambda(t-\Upsilon(\lambda))}G(\lambda) - 1}{1 - e^{-i\lambda}}$$

under application of the filter. Thus the interpretation is that the time index t gets lagged by $\Upsilon(\lambda)$, and so in the GRW case this can still be interpreted as Phase Delay. However, G does not multiply the whole fraction $(e^{i\lambda t} - 1)/(1 - e^{-i\lambda})$, so it cannot really be interpreted as the Gain for the process. If we take first temporal differences of the above integrands, we obtain $e^{i\lambda t} \mapsto e^{i\lambda(t-\Upsilon(\lambda))}G(\lambda)$; this shows that G is interpreted as the Gain function of the differenced GRW process. Now for the t.p. filter we have $\Psi(1) = 0$, so the Υ and G functions can only really be interpreted as Phase Delay and Gain for the differenced GRW process.

With these observations in mind, we now proceed to the Gain, Phase, and Phase Delay plots for trend, t.p.'s, and forecast filters. Figure 1 shows the Gain function for the trend filter $\Psi^{(1)}(B)$. We see it passes frequency zero, accentuates some low frequencies, and tends to attenuate higher frequencies, albeit in a nonuniform (monotonic) fashion. This behavior is due to the enforced constraints, ensuring coherency across 1y, 3y, and 5y estimates. The zeroes are at $\lambda = 2\pi/5$, $4\pi/5$, and $2\pi/3$, which can also be deduced directly from the formulas for the filter. The Phase Delay plot in Figure 2 shows there is no lag at frequency zero (this is ensured by the condition that a line be preserved), and at the first zero $\lambda = 2\pi/5$ the delay is roughly 2.5 time units. The delay does not increase beyond 3 time units, which is entirely reasonable for a concurrent trend filter.

The t.p. filter is nonzero at frequency zero, which prohibits us from formulating a well-defined Phase Delay function. The Gain function (Figure 3) shows this complete attenuation at frequency zero; as designed, the action of this filter on a line is that only the constant slope is reproduced. Basically all frequencies are attenuated, since the Gain is always less than one in magnitude. In terms of the stochastic nature of the data, this filter greatly reduces the variance; this is fine, because a t.p. filter only reveals changes in directionality, and is not concerned with preserving the overall scale of the original data. The other zeroes of the t.p. are the same as the trend filter. The Phase function (Figure 4) gives a constant increase, with a slope of 3. Note that because the Phase is negative for $\lambda \leq \pi/2$, there is actually a "phase advance" of these frequencies. Dividing by λ to obtain the Phase Delay function, the phase advance is actually quite large at these low frequencies (the function tends to $-\infty$ at $\lambda = 0$). But Υ tends asymptotically to 3 for larger λ, indicating a time lag of three units at higher frequencies. Because of the unbounded behavior at $\lambda = 0$, we do not show the Phase Delay plot.

Finally, the forecast filter is displayed in Figures 5 and 6. The Gain function is almost identical to that of the trend filter, which is not surprising given their similar derivation. The forecast Gain has a bit more accentuation of low frequencies. The main difference lies in the Phase Delay plots; although the shape of this function is similar to that of the trend, note that Phase Delay is negative for frequencies up to roughly .1. Thus the low frequency range has a phase advance property, or in other words it is projected forward in time. So the forecast filter does indeed forecast these low frequencies. At higher frequencies, phase delay takes over and gets as high as three time units.

6 Discussion and Conclusion

We began with a discussion about key properties for any signal filter, and arrived at conditions 1 through 4. Guided by these principles, we were able to construct filters for trends, turning points, and forecasts in a straightforward fashion. One underlying assumption that should be emphasized is that much of the analysis depends upon a local linear structure in the data. The more formal analysis of Section 5 considers the data to be given (at least locally) by a GRW or stationary process. However, the methods presented in Section 2 can, in principle, be generalized to twice-integrated stochastic processes, where interest focuses on quadratic trends. The needed calculations would be much more involved. But macro-economic and demographic theory tend to support the idea of a single order of integration in the stochastic dynamics (at least locally, i.e., for a short span of data).

What about other signals? The typical econometric analysis of an economic time series considers seasonal and cyclical dynamics, as well as trading day and holiday effects (Findley, Monsell, Bell, Otto, and Chen, 1998). Since the data is measured annually, the only relevant dynamic is the cycle. The design of a cycle filter is far beyond the scope of this work, since it is unclear about how to relate such a filter to linear dynamics (the cycle is concerned with truly periodic effects somewhat removed from frequency zero). But other signals of interest can arise from the desire to forecast components. We note that the forecast filter we consider essentially forecasts the trend, since the phase advance is concentrated in the low frequency band. One possibility is to consider a high-frequency forecast filter, though in general the high-frequency components contain less interest for typical users.

If we view the evolution of an economic or demographic process from a kinetic perspective, it is often interesting to consider the "velocity" and "acceleration" of such a series, in addition to other kinematic characteristics (e.g., curvature, zero-crossings of the velocity, etc.). The turning point filter can be viewed as a low-frequency velocity filter, since it has the effect of the first derivative on the line $mx + b$. Change of sign of the turning point filter indicates a change in the overall trending behavior of the data. The ability to forecast such a zero-crossing amounts to prediction of change in an economic or demographic trend. An acceleration filter would give information about the convexity of the low-frequency portions of the series. The Newtonian characterization of local peaks and troughs through velocity and acceleration seems to be a useful concept for analysis of trending data. To this end, filters to measure velocity and acceleration, and forecasts of such, would be of value and interest, and could be designed along the lines of the methodology of Section 2.

Whatever the choice of signal and filter, the main consideration is that the $1y$, $3y$, and $5y$ results be made compatible. At the heart of the issue is the non-invertibility of $\Theta^{(3)}(B)$ and $\Theta^{(5)}(B)$; these filters essentially become components of any signal filter for the $1y$ estimate that we design (hence, we expect that only low-frequency behavior can be successfully gauged through the methods of this work). For the average user, this means that a few basic signals (such as the trend and t.p.) should be provided and published so that inapt comparisons are avoided. For the professional user (i.e., internal analyst), they should first define what characteristic or signals they are interested in, and develop the corresponding filters along the lines that have been delineated here. Presumably, interest will focus on cross-county comparisons of characteristics such as trends, forecasts, and so forth. It is essential that condition 1 be satisfied *before* any analysis is conducted. Thus, this work will be of value to two diverse groups: the statistically unsophisticated public user (for which the internal details of this methodology must remain opaque) and the professionally trained statistician (who is encouraged to learn and expand this methodology so that their analyses will obtain a heightened validity).

Appendix

A.1 Action of Filters on Polynomials

Here we discuss the action of filters on polynomials; for a related reference, see Brockwell and Davis (1996). Consider a generic filter $G(z)$ acting on a line $at + b$:

$$G(B)\left[at + b\right] = aG(B)t + G(1)b.$$

For a trend filter, we wish $G(z)$ to pass the line, which requires that $G(B)t = t$ and $G(1) = 1$. The first condition is equivalent to $(1 - G(B))\, t = 0$, or

$$1 - G(z) = H(z)(1 - z)^2$$

for some polynomial H. This is true, because $1 - B$ reduces t to unity, and another difference is needed to produce zero. It follows from this condition that $G(1) = 1$. Hence, it is necessary and sufficient for $G(z)$ to pass lines that $1 - G(z)$ is divisible by $(1 - z)^2$. If we wish for a line to be passed with a lag of j time units (where if j is negative there is an advance), we have $G(B)[t] = t - j$, or

$$(1 - G(B))[t] = j.$$

This implies that $1 - G(z)$ must accomplish differencing with a change in level, or

$$1 - G(z) = H(z)(1 - z),$$

where $H(1) = j$. For the trend filter $j = 0$, whereas $j = -1$ for the forecast filter.

A.2 Proofs of Theorems 1 and 2.

We seek trend/forecast filters that will "pass" lines, and Appendix A.1 gives conditions that guarantee this. Now for $k = 3, 5$ we have

$$1 - \Psi^{(k)}(z)\Theta^{(k)}(z) = 1 - \Psi^{(k)}(z) + \Psi^{(k)}(z)(1 - z)\left(z^{k-2} + 2z^{k-3} + \cdots + (k-2)z + (k-1)\right)/k,$$

which uses $(1-\Theta^{(k)}(z))/(1-z) = (z^{k-2}+2z^{k-3}+\cdots+(k-2)z+(k-1))/k$. Choosing $\Psi^{(k)}(z)$ such that $\Psi^{(k)}(1)=1$, we know that $1-z$ divides $1-\Psi^{(k)}(z)$. Letting $\Phi^{(k)}(z)$ denote the polynomial quotient, we have

$$\frac{1-\Psi^{(k)}(z)\Theta^{(k)}(z)}{1-z} = \Phi^{(k)}(z) + \Psi^{(k)}(z)\left(z^{k-2}+2z^{k-3}+\cdots+kz+(k-1)\right)/k.$$

We wish to construct $\Phi^{(k)}(z)$ such that the right hand side evaluated at $z=1$ is j, where $j=0$ for trends and $j=-1$ for forecasts (see Appendix A.1). Thus we obtain the condition that

$$\Phi^{(k)}(1) = j - \frac{k-1}{2}. \tag{A.1}$$

So, given a choice of $\Phi^{(k)}(z)$ such that (A.1) holds, we simply let

$$\Psi^{(k)}(z) = 1-(1-z)\Phi^{(k)}(z).$$

This will ensure that $1-\Psi^{(k)}(z)\Theta^{(k)}(z)$ is divisible by $(1-z)^2$, so that it passes lines. The other constraints of our solution depend on (3); if $\Phi^{(5)}(z)$ is degree d, then there are $d+8$ constraints and $2d+4$ degrees of freedom offered by the coefficients of $\Phi^{(3)}(z)$ and $\Phi^{(5)}(z)$. Although this implies that the minimal degree polynomial solution is $d=4$, there are redundancies among the constraints that can be exploited, such that we can take $d=2$. We next derive the solution.

If $\Phi^{(5)}(z)$ has degree 2, it follows from (3) that $\Phi^{(3)}(z)$ has degree 4. We write

$$\Phi^{(5)}(z) = a_0 + a_1 z + a_2 z^2$$

from which it follows that

$$\begin{aligned}
\Psi^{(5)}(z) &= (1-a_0) + (a_0-a_1)z + (a_1-a_2)z^2 + a_2 z^3 \\
\Psi^{(5)}(z)\Theta^{(5)}(z) &= \left((1-a_0)+(1-a_1)z+(1-a_2)z^2+z^3+z^4+a_0z^5+a_1z^6+a_2z^7\right)/5.
\end{aligned}$$

Similarly we have

$$\Phi^{(3)}(z) = b_0 + b_1 z + b_2 z^2 + b_3 z^3 + b_4 z^4$$

from which it follows that

$$\begin{aligned}
\Psi^{(3)}(z) &= (1-b_0)+(b_0-b_1)z+(b_1-b_2)z^2+(b_2-b_3)z^3+(b_3-b_4)z^4+b_4z^5 \\
\Psi^{(3)}(z)\Theta^{(3)}(z) &= \left((1-b_0)+(1-b_1)z+(1-b_2)z^2+(b_0-b_3)z^3+(b_1-b_4)z^4+b_2z^5+b_3z^6+b_4z^7\right)/3.
\end{aligned}$$

At this point we use (3) and match coefficients. Consider the trend ($j=0$) and forecast ($j=-1$) filters together. From matching coefficients, we obtain the following equations for b_k and a_k:

$$b_0 = \frac{2}{5}+\frac{3}{5}a_0 \qquad b_1 = \frac{2}{5}+\frac{3}{5}a_1 \qquad b_2 = \frac{3}{5}a_0 \qquad b_3 = \frac{3}{5}a_1 \qquad b_4 = \frac{3}{5}a_2 \tag{A.2}$$

We also obtain several constraints on the a_k's:

$$\frac{3}{5} = \frac{2}{5} + \frac{3}{5}(a_1 - a_2) \qquad \frac{3}{5} = \frac{2}{5} + \frac{3}{5}(a_0 - a_1) \qquad \frac{2}{5} = \frac{3}{5}(a_0 - a_2) \tag{A.3}$$

The third equation is implied by the other two. Finally, applying (8) gives the additional constraints

$$j - 2 = a_0 + a_1 + a_2$$
$$j - 1 = b_0 + b_1 + b_2 + b_3 + b_4.$$

The second equation, together with (A.2) yields

$$j - 1 = \frac{4}{5} + \frac{3}{5}\left(2a_0 + 2a_1 + a_2\right).$$

This results in the following matrix system

$$\begin{bmatrix} 1 \\ 1 \\ 5j-9 \end{bmatrix} = \begin{bmatrix} 0 & 3 & -3 \\ 3 & -3 & 0 \\ 6 & 6 & 3 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}. \tag{A.4}$$

The inverse matrix is

$$\frac{1}{15}\begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 2 \\ -4 & -2 & 1 \end{bmatrix}. \tag{A.5}$$

Now for $j = 0$, this yields the solution

$$\Phi^{(5)}(z) = -\left(1 + 2z + 3z^2\right)/3$$
$$\Phi^{(3)}(z) = \left(1 - z^2 - 2z^3 - 3z^4\right)/5.$$

Then the 1y, 3y and 5y trend filters are

$$\Psi^{(5)}(z) = \left(4 + z + z^2 - 3z^3\right)/3$$
$$\Psi^{(3)}(z) = \left(4 + z + z^2 + z^3 + z^4 - 3z^5\right)/5$$
$$\Psi^{(1)}(z) = \left(4 + 5z + 6z^2 + 3z^3 + 3z^4 - z^5 - 2z^6 - 3z^7\right)/15.$$

For $j = -1$ we obtain

$$\Phi^{(5)}(z) = -\left(2 + 3z + 4z^2\right)/3$$
$$\Phi^{(3)}(z) = -\left(z + 2z^2 + 3z^3 + 4z^4\right)/5.$$

Then the 1y, 3y and 5y forecast filters are

$$\Psi^{(5)}(z) = \left(5 + z + z^2 - 4z^3\right)/3$$
$$\Psi^{(3)}(z) = \left(5 + z + z^2 + z^3 + z^4 - 4z^5\right)/5$$
$$\Psi^{(1)}(z) = \left(5 + 6z + 7z^2 + 3z^3 + 3z^4 - 2z^5 - 3z^6 - 4z^7\right)/15.$$

It was demonstrated above that these filters have the shortest length possible with the stated properties. This completes the proof. $\square$

References

[1] Alexander, C. (1998) Recent developments in the American Community Survey. *Proceedings of the Survey Research Methods Section, ASA*, 92–100.

[2] Alexander, C. (2001) Still rolling: Leslie Kish's "rolling samples" and the American Community Survey. *Proceedings of Statistics Canada Symposium 2001, Achieving Data Quality in a Statistical Agency: A Methodological Perspective.*

[3] Bell, W. and Hillmer, S. (1984) Issues involved with the Seasonal Adjustment of Economic Time Series. *Journal of Business and Economic Statistics* **2**, 291–320.

[4] Brockwell, P. and Davis, R. (1996) *Introduction to Time Series and Forecasting.* New York: Springer.

[5] Census Bureau, U.S. (2006) Technical Paper 67. *Design and Methodology, American Community Survey.*
http://www.census.gov/acs/www/Downloads/tp67.pdf

[6] Citro, C. and Kalton, G. (2007) *Using the American Community Survey: benefits and challenges.* The National Academies Press: Washington, D.C.

[7] Findley, D. and Martin, D. (2006) Frequency Domain Analyses of SEATS and X-12-ARIMA Seasonal Adjustment Filters for Short and Moderate-Length Time Series. *Journal of Official Statistics* **22**, 1–34.

[8] Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C. and Chen, B. C. (1998) New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program. *Journal of Business and Economic Statistics* **16**, 127–177 (with discussion).

[9] Kish, L. (1981) "Using cumulated rolling samples to integrate census and survey operations of the Census Bureau," Washington, D.C., U.S. Government Printing Office.

[10] Kish, L. (1998) Space/Time variations and rolling samples. *Journal of Official Statistics* **14**, 31–46.

	MYEs		
Time	1y	3y	5y
99	**40.05**		
00	40.00		
01	41.00	40.00	
02	41.80	41.00	
03	40.80	41.20	40.70
04	40.60	41.00	40.80
05	41.70	41.10	41.20
06	**42.04**	**41.28**	**41.45**

Table 1: MYEs for Mean Travel Time of Bronx, NYC, New York in minutes. Estimates have been backcast and forecast extended to the years 99 and 06, written in bold.

	Characteristics				
	Divorce	Travel	Income	Age	Rent
$C^{(3)}$	.59	.07	10.10	.16	5.65
$C^{(5)}$	.11	.04	9.4	.12	2.83
1y Trend	16404	41.79	36125	37.61	4.38
3y Trend	17410	41.64	53003	37.50	5.12
5y Trend	16718	41.90	32370	37.93	3.40
1y Forecast	16552	42.01	35240	37.77	4.13
3y Forecast	17775	41.82	56229	37.63	5.02
5y Forecast	16954	42.15	29173	38.18	2.85
1y T.p.	149	.22	-885	.15	-.25
3y T.p.	365	.18	3226	.13	-.10
5y T.p.	236	.25	-3197	.25	-.55

Table 2: Compatibility measures for seven series (in percentages), as well as trends, forecasts and t.p.s for 2006.

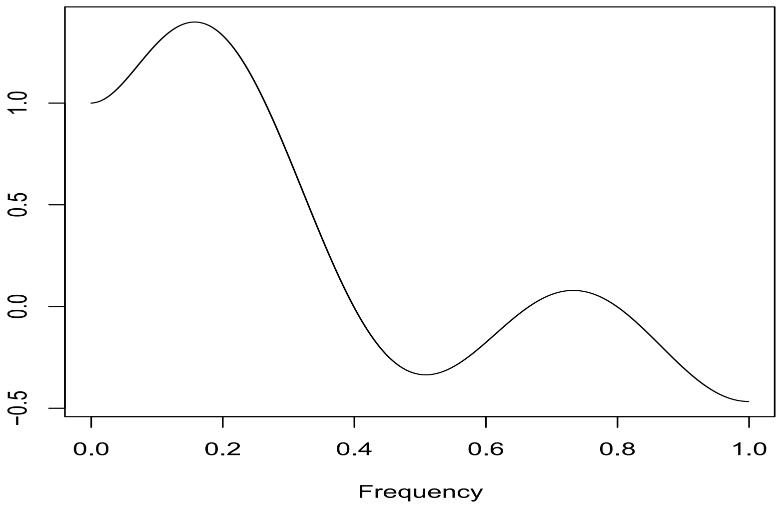

Figure 1: Gain Function for the Trend Filter

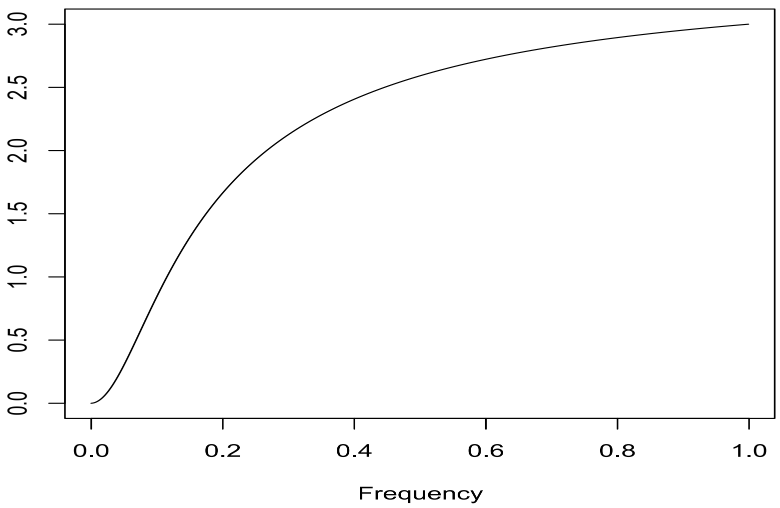

Figure 2: Phase Delay Function for the Trend Filter

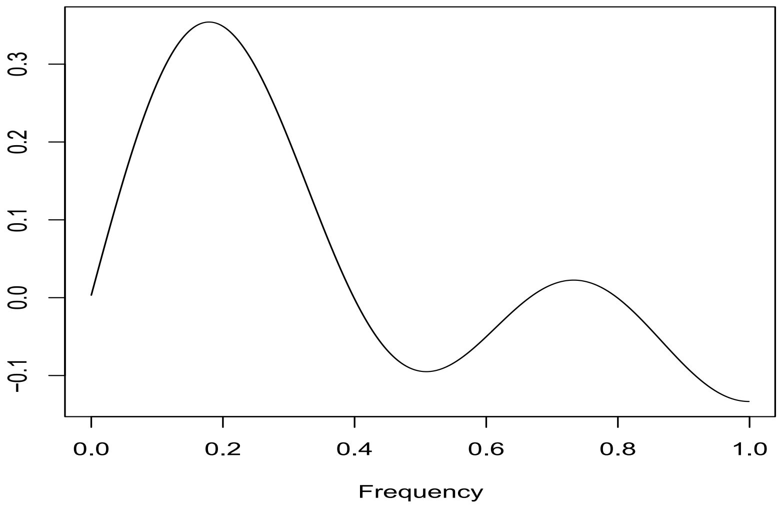

Figure 3: Gain Function for the Turning Point Filter

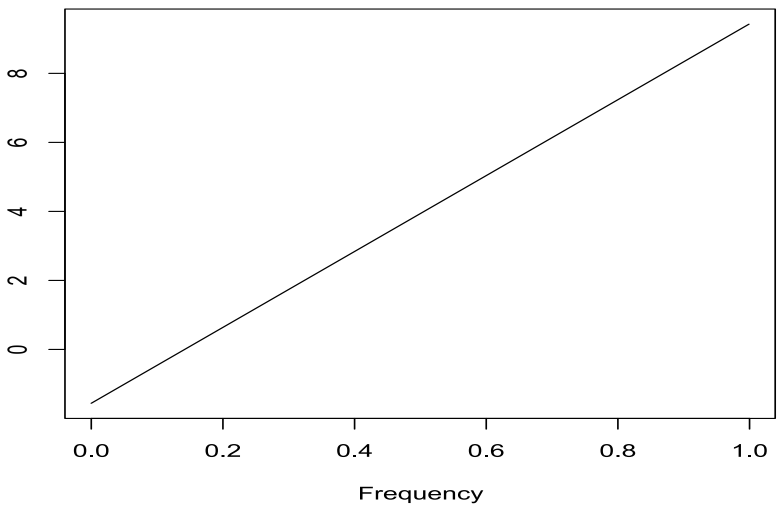

Figure 4: Phase Function for the Turning Point Filter

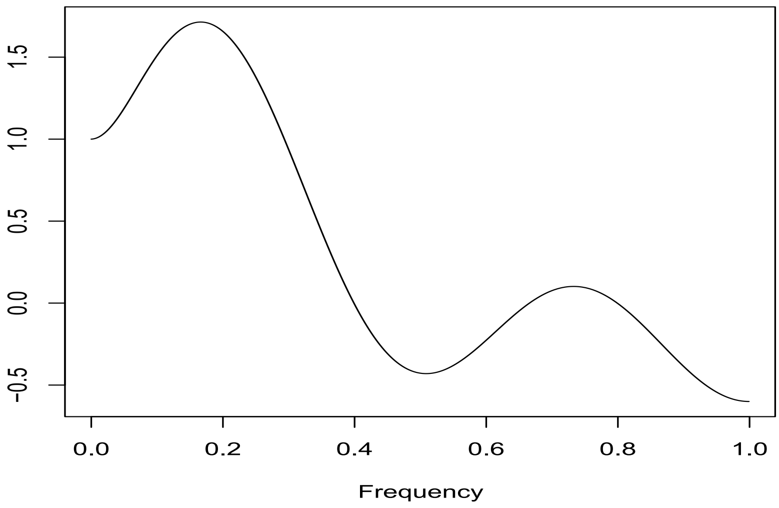

Figure 5: Gain Function for the Forecast Filter

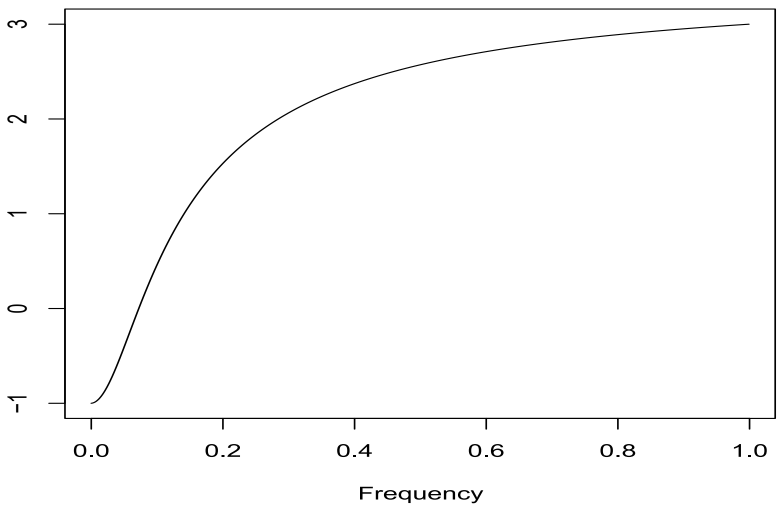

Figure 6: Phase Delay Function for the Forecast Filter