

A SMOOTHING METHOD THAT LOOKS LIKE THE HODRICK–PRESCOTT FILTER*

HIROSHI YAMADA
Hiroshima University

In recent decades, in the research community of macroeconomic time series analysis, we have observed growing interest in the smoothing method known as the Hodrick–Prescott (HP) filter. This article examines the properties of an alternative smoothing method that looks like the HP filter, but is much less well known. We show that this is actually more like the exponential smoothing filter than the HP filter although it is obtainable through a slight modification of the HP filter. In addition, we also show that it is also like the low-frequency projection of Müller and Watson (2018, *Econometrica* 86, 775–804). We point out that these results derive from the fact that all three similar smoothing methods can be regarded as a type of graph spectral filter whose graph Fourier transform is discrete cosine transform. We then theoretically reveal the relationship between the similar smoothing methods and provide a way of specifying the smoothing parameter that is necessary for its application. An empirical examination illustrates the results.

1. INTRODUCTION

In recent decades, in the research community of macroeconomic time series analysis, we have observed growing interest in the Hodrick–Prescott (HP) (1997) filter. This is now the most prominent smoothing method for economic time series and recent studies of the filter include Phillips and Jin (2015), de Jong and Sakarya (2016), Cornea-Madeira (2017), Hamilton (2018), Phillips and Shi (2019), and Sakarya and de Jong (2020). The smoothing method is defined by

$$\text{HP: } \min_{x_1, \dots, x_T \in \mathbb{R}} \sum_{t=1}^T (y_t - x_t)^2 + \lambda \sum_{t=3}^T (\Delta^2 x_t)^2,$$

*I am grateful to In Choi, Chirok Han, Eiji Kurozumi, Takashi Yamagata, Daisuke Yamazaki, three anonymous referees, and the editor, Peter C.B. Phillips, for their valuable suggestions and comments. I also thank the participants in several conferences including the York–Hiroshima Joint Symposium 2018 held at the University of York, the BK21PLUS Korean Economic Group International Conference on Econometrics held at Korea University, and the SH3 Conference on Econometrics 2019 held at Singapore Management University. In addition, I appreciate the kindness of Ulrich K. Müller and Mark W. Watson in providing their dataset and m files. The usual caveat applies. The Japan Society for the Promotion of Science supported this work through KAKENHI Grant No. 16H03606. Address correspondence to Hiroshi Yamada, 1-2-1 Kagamiyama, Higashi-Hiroshima 739-8525, Japan; e-mail: yamada@hiroshima-u.ac.jp.

where y_1, \dots, y_T denote T observations of an economic time series, such as real gross domestic product (GDP), λ is a positive smoothing parameter that controls fidelity and smoothness, and Δ denotes a difference operator such that $\Delta x_t = x_t - x_{t-1}$ and accordingly $\Delta^2 x_t = x_t - 2x_{t-1} + x_{t-2}$.

In this article, we examine the properties of an alternative smoothing method that looks like the HP filter. It was developed by O’Sullivan (1991), Buckley (1994), and Garcia (2010) and is defined by

$$\text{mHP} : \min_{x_1, \dots, x_T \in \mathbb{R}} \sum_{t=1}^T (y_t - x_t)^2 + \lambda \sum_{t=3}^T (\Delta^2 x_t)^2 + \lambda \{(\Delta x_2)^2 + (\Delta x_T)^2\},$$

where “mHP” signifies the “modified Hodrick–Prescott.” The properties of this somewhat unfamiliar filter are less known in the research community and this motivates our examination.

Because the mHP filter is obtainable through a slight modification of the HP filter, it is reasonable to expect that these two filters share common properties. This is correct to a certain extent, as shown later. However, in the article, we also show that it is more like the exponential smoothing (ES) filter, which is presented in King and Rebelo (1993), than the HP filter. Here, the ES filter is defined by

$$\text{ES} : \min_{x_1, \dots, x_T \in \mathbb{R}} \sum_{t=1}^T (y_t - x_t)^2 + \psi \sum_{t=2}^T (\Delta x_t)^2,$$

where ψ is a positive smoothing parameter.¹ In addition, and very interestingly, it is also like the low-frequency projection (LFP) of Müller and Watson (2018), which is an orthogonal projection of a time series onto the space spanned by cosine function based low-frequency periodic column vectors and originates in Phillips (2005a).²

Why and how is the mHP filter like the ES and LFP filters? In this article, we address these questions. We show that all of these similar smoothing methods, that is, the mHP, ES, and LFP filters, can be regarded as a type of graph spectral filter (GSF) whose graph Fourier transform is discrete cosine transform (DCT) and therefore exhibit similarities, where GSF is a promising smoothing method and has been developed in the field of graph signal processing.³ Let U' denote the $T \times T$ DCT matrix, which is an orthogonal matrix and explicitly expressed by (6). Then, the minimizers of the mHP and ES filters and the LFP projection are commonly represented as

$$\hat{x}_i = U H_i U' y, \quad i = \text{mHP, ES, LFP}, \tag{1}$$

¹Both the HP and ES filters are a type of Whittaker–Henderson method of graduation. This is a classic smoothing procedure developed by Bohlmann (1899), Whittaker (1923), Henderson (1924, 1925), Aitken (1927), and others and is defined by $\min_{x_1, \dots, x_T \in \mathbb{R}} \sum_{t=1}^T (y_t - x_t)^2 + \lambda_\alpha \sum_{t=\alpha+1}^T (\Delta^\alpha x_t)^2$. For a review, see Weinert (2007), Phillips (2010), and Nocon and Scott (2012).

²Furthermore, the approach explored in Phillips (2005a) relies on Phillips (1998). See also Phillips (2005b).

³For details, see Shuman, Narang, Frossard, Ortega, and Vandergheynst (2013).



FIGURE 1. A path graph whose graph Laplacian is L in (2).

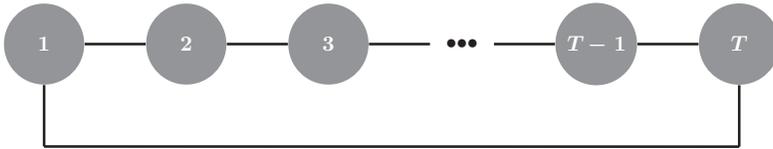


FIGURE 2. A cycle graph whose graph Laplacian is L_{cycle} in (A.2).

where $U'y$ denotes the DCT of $y = [y_1, \dots, y_T]'$ [see (5)] and H_i is a $T \times T$ diagonal matrix, as defined later. Subsequently, using (1), we examine theoretically how they are similar and differ, and provide a way of specifying the smoothing parameter that is necessary for its application. We then empirically illustrate the results.

In the article, we also remark that DCT is more appropriate for econometric time series analysis than the discrete Fourier transform (DFT). The reason is that the DCT matrix is obtainable from the graph Laplacian of the path graph (Figure 1), whereas DFT is obtainable from the cycle graph (Figure 2). As illustrated in Figures 1 and 2, the path graph is more appropriate as an underlying graph of time series than the cycle graph.

Outline of the paper. The article is organized as follows. Section 2 reveals some basic properties of the mHP, HP, and ES filters. This section also theoretically highlights how the mHP filter is like (unlike) the ES (HP) filter. Section 3 provides a review of GSF and DCT, and presents a key matrix factorization in the form of (7). Section 4 demonstrates that the mHP, ES, and LFP filters are all a type of GSF based on DCT and we present their properties as derived from this fact. In addition, we theoretically reveal the relationship between these similar filters and provide a way of specifying the mHP filter's smoothing parameter and present an empirical illustration. Lastly, we remark that DCT is more appropriate for time series econometrics than DFT. Section 5 concludes the paper.

Notations. Let $\mathbf{x} = [x_1, \dots, x_T]'$, $\mathbf{t} = [1, \dots, 1]' \in \mathbb{R}^T$, $\boldsymbol{\tau} = [1, \dots, T]'$, $\boldsymbol{\Pi} = [\mathbf{t}, \boldsymbol{\tau}]$, \mathbf{I}_q denote the $q \times q$ identity matrix, $\mathbf{Q}_t = \mathbf{I}_T - \mathbf{t}(\mathbf{t}'\mathbf{t})^{-1}\mathbf{t}'$, and, given η_1, \dots, η_T , $\bar{\eta} = \frac{1}{T} \sum_{t=1}^T \eta_t$. \mathbf{D}_1 is the $(T-1) \times T$ first-order difference matrix such that $\mathbf{D}_1\mathbf{x} = [\Delta x_2, \dots, \Delta x_T]'$. Similarly, \mathbf{D}_2 is the $(T-2) \times T$ second-order difference matrix such that $\mathbf{D}_2\mathbf{x} = [\Delta^2 x_3, \dots, \Delta^2 x_T]'$. L is the following $T \times T$ symmetric tridiagonal matrix and is described by \mathbf{D}_2 as

$$L = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{f}_1 \\ -\mathbf{D}_2 \\ \mathbf{f}_T \end{bmatrix}, \tag{2}$$

where both \mathbf{f}_1 and \mathbf{f}_T are T -dimensional row vectors and let $\mathbf{F} = [\mathbf{f}'_1, \mathbf{f}'_T] \in \mathbb{R}^{T \times 2}$. For a vector $\boldsymbol{\eta} = [\eta_1, \dots, \eta_n]'$, $\|\boldsymbol{\eta}\|^2 = \boldsymbol{\eta}'\boldsymbol{\eta} = \eta_1^2 + \dots + \eta_n^2$. Finally, for a matrix \mathbf{V} , $\mathcal{S}(\mathbf{V})$ and $\mathcal{S}^\perp(\mathbf{V})$ denote the column space of \mathbf{V} and the orthogonal complement of $\mathcal{S}(\mathbf{V})$, respectively.

A note on L in (2). First, L is a familiar matrix to econometricians and statisticians because

$$L = \mathbf{D}'_1 \mathbf{D}_1 \tag{3}$$

appears in the Durbin–Watson statistic (Durbin and Watson, 1950, 1951). Second, and more importantly with respect to this article, L is the graph Laplacian of a graph whose vertex set and edge set are, respectively, given by

$$S_1 = \{1, \dots, T\} \quad \text{and} \quad S_2 = \{\{1, 2\}, \{2, 3\}, \dots, \{T - 1, T\}\}, \tag{4}$$

which is referred to as a path graph of order T (Figure 1).⁴ In addition, \mathbf{D}'_1 is an incidence matrix of the graph. Third, \mathbf{f}_1 and \mathbf{f}_T in L correspond to the Neumann boundary conditions (O’Sullivan, 1991; Strang and MacNamara, 2014).

2. BASIC PROPERTIES OF THE MHP, HP, AND ES FILTERS

In matrix notation, the mHP, HP, and ES filters are represented as follows:

$$\text{mHP} : \min_{\mathbf{x} \in \mathbb{R}^T} f_{\text{mHP}}(\mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|^2 + \lambda \|\mathbf{L}\mathbf{x}\|^2,$$

$$\text{HP} : \min_{\mathbf{x} \in \mathbb{R}^T} f_{\text{HP}}(\mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|^2 + \lambda \|\mathbf{D}_2\mathbf{x}\|^2,$$

$$\text{ES} : \min_{\mathbf{x} \in \mathbb{R}^T} f_{\text{ES}}(\mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|^2 + \psi \|\mathbf{D}_1\mathbf{x}\|^2.$$

Now, we describe some basic properties of the mHP filter:

PROPOSITION 2.1. *Let $\widehat{\mathbf{x}}_{\text{mHP}} = (\mathbf{I}_T + \lambda \mathbf{L}'\mathbf{L})^{-1} \mathbf{y} [= (\mathbf{I}_T + \lambda \mathbf{L}^2)^{-1} \mathbf{y}]$. (i) (a) $\widehat{\mathbf{x}}_{\text{mHP}}$ is a unique global minimizer of $f_{\text{mHP}}(\mathbf{x})$ and (b) it satisfies $\|\mathbf{L}\widehat{\mathbf{x}}_{\text{mHP}}\|^2 < \|\mathbf{L}\mathbf{y}\|^2$ if $\mathbf{y} \neq \widehat{\mathbf{x}}_{\text{mHP}}$. (ii) (a) Each row of the hat matrix of the mHP filter, $(\mathbf{I}_T + \lambda \mathbf{L}\mathbf{L})^{-1}$, sums to unity and (b) the hat matrix is bisymmetric (i.e., symmetric centrosymmetric). (iii) (a) $\frac{1}{T} \mathbf{1}'\widehat{\mathbf{x}}_{\text{mHP}} = \bar{y}$ and $\mathbf{1}'(\mathbf{y} - \widehat{\mathbf{x}}_{\text{mHP}}) = 0$, (b) $\lim_{\lambda \rightarrow \infty} \widehat{\mathbf{x}}_{\text{mHP}} = \bar{y}\mathbf{1}$, (c) $\lim_{\lambda \rightarrow 0} \widehat{\mathbf{x}}_{\text{mHP}} = \mathbf{y}$, and (d) if $\mathbf{y} \in \mathcal{S}(\mathbf{L})$, then $\widehat{\mathbf{x}}_{\text{mHP}} = \mathbf{y}$.*

⁴See, for example, Nakatsukasa, Saito, and Woei (2013).

Proof. See the Appendix. ■

For comparison, we describe the corresponding basic properties of the HP and ES filters:

PROPOSITION 2.2. *Let $\widehat{\mathbf{x}}_{\text{HP}} = (\mathbf{I}_T + \lambda \mathbf{D}'_2 \mathbf{D}_2)^{-1} \mathbf{y}$. (i) (a) $\widehat{\mathbf{x}}_{\text{HP}}$ is a unique global minimizer of $f_{\text{HP}}(\mathbf{x})$ and (b) it satisfies $\|\mathbf{D}_2 \widehat{\mathbf{x}}_{\text{HP}}\|^2 < \|\mathbf{D}_2 \mathbf{y}\|^2$ if $\mathbf{y} \neq \widehat{\mathbf{x}}_{\text{HP}}$. (ii) (a) Each row of the hat matrix of the HP filter, $(\mathbf{I}_T + \lambda \mathbf{D}'_2 \mathbf{D}_2)^{-1}$, sums to unity and (b) the hat matrix is bisymmetric. (iii) (a) $\frac{1}{T} \mathbf{1}' \widehat{\mathbf{x}}_{\text{HP}} = \bar{y}$ and $\mathbf{t}'(\mathbf{y} - \widehat{\mathbf{x}}_{\text{HP}}) = 0$, (b) $\lim_{\lambda \rightarrow \infty} \widehat{\mathbf{x}}_{\text{HP}} = \mathbf{\Pi}(\mathbf{\Pi}' \mathbf{\Pi})^{-1} \mathbf{\Pi}' \mathbf{y}$, (c) $\lim_{\lambda \rightarrow 0} \widehat{\mathbf{x}}_{\text{HP}} = \mathbf{y}$, and (d) if $\mathbf{y} \in \mathcal{S}(\mathbf{\Pi})$, then $\widehat{\mathbf{x}}_{\text{HP}} = \mathbf{y}$.*

Proof. For (i)(a), see Danthine and Girardin (1989); for (ii)(b), see Cornea-Madeira (2017) and Yamada (2019); for (iii)(b), see Yamada (2015); and for (iii)(c), see Kim, Koh, Boyd, and Gorinevsky (2009). Derive other results as for Proposition 2.1. ■

Remark 2.3. We would like to refer to some papers related to this proposition. (i)(b) is stated in Weinert (2007) and (iii)(d) is mentioned by Kim et al. (2009). A related argument to (ii)(a) is given in Yamada (2018a).

PROPOSITION 2.4. *Let $\widehat{\mathbf{x}}_{\text{ES}} = (\mathbf{I}_T + \psi \mathbf{D}'_1 \mathbf{D}_1)^{-1} \mathbf{y}$. (i) (a) $\widehat{\mathbf{x}}_{\text{ES}}$ is a unique global minimizer of $f_{\text{ES}}(\mathbf{x})$ and (b) it satisfies $\|\mathbf{D}_1 \widehat{\mathbf{x}}_{\text{ES}}\|^2 < \|\mathbf{D}_1 \mathbf{y}\|^2$ if $\mathbf{y} \neq \widehat{\mathbf{x}}_{\text{ES}}$. (ii) (a) Each row of the hat matrix of the ES filter, $(\mathbf{I}_T + \psi \mathbf{D}'_1 \mathbf{D}_1)^{-1}$, sums to unity and (b) the hat matrix is bisymmetric. (iii) (a) $\frac{1}{T} \mathbf{1}' \widehat{\mathbf{x}}_{\text{ES}} = \bar{y}$ and $\mathbf{t}'(\mathbf{y} - \widehat{\mathbf{x}}_{\text{ES}}) = 0$, (b) $\lim_{\psi \rightarrow \infty} \widehat{\mathbf{x}}_{\text{ES}} = \bar{y} \mathbf{1}$, (c) $\lim_{\psi \rightarrow 0} \widehat{\mathbf{x}}_{\text{ES}} = \mathbf{y}$, and (d) if $\mathbf{y} \in \mathcal{S}(\mathbf{t})$, then $\widehat{\mathbf{x}}_{\text{ES}} = \mathbf{y}$.*

Proof. This proposition can be proved as Propositions 2.1 and 2.2. ■

COROLLARY 2.5. *Let $\widehat{\boldsymbol{\tau}} = \mathbf{\Pi}(\mathbf{\Pi}' \mathbf{\Pi})^{-1} \mathbf{\Pi}' \mathbf{y}$. $\widehat{\mathbf{x}}_{\text{mHP}}$, $\widehat{\mathbf{x}}_{\text{HP}}$, and $\widehat{\mathbf{x}}_{\text{ES}}$ can alternatively be represented as (i) $\widehat{\mathbf{x}}_{\text{mHP}} = \bar{y} \mathbf{1} + (\mathbf{I}_T + \lambda \mathbf{L}' \mathbf{L})^{-1}(\mathbf{y} - \bar{y} \mathbf{1})$, (ii) $\widehat{\mathbf{x}}_{\text{HP}} = \widehat{\boldsymbol{\tau}} + (\mathbf{I}_T + \lambda \mathbf{D}'_2 \mathbf{D}_2)^{-1}(\mathbf{y} - \widehat{\boldsymbol{\tau}})$, and (iii) $\widehat{\mathbf{x}}_{\text{ES}} = \bar{y} \mathbf{1} + (\mathbf{I}_T + \psi \mathbf{D}'_1 \mathbf{D}_1)^{-1}(\mathbf{y} - \bar{y} \mathbf{1})$, respectively.*

Proof. (i) From Proposition 2.1 (iii)(d), it follows that $(\mathbf{I}_T + \lambda \mathbf{L}' \mathbf{L})^{-1}(\bar{y} \mathbf{1}) = (\bar{y} \mathbf{1})$. Then, by using the equation, we obtain: $\bar{y} \mathbf{1} + (\mathbf{I}_T + \lambda \mathbf{L}' \mathbf{L})^{-1}(\mathbf{y} - \bar{y} \mathbf{1}) = \bar{y} \mathbf{1} + \widehat{\mathbf{x}}_{\text{mHP}} - \bar{y} \mathbf{1} = \widehat{\mathbf{x}}_{\text{mHP}}$. (ii) and (iii) can be proved as in (i). ■

Remark 2.6. (a) All of these are low-pass filters and thus $(\mathbf{I}_T + \lambda \mathbf{L}' \mathbf{L})^{-1}(\mathbf{y} - \bar{y} \mathbf{1})$, $(\mathbf{I}_T + \lambda \mathbf{D}'_2 \mathbf{D}_2)^{-1}(\mathbf{y} - \widehat{\boldsymbol{\tau}})$, and $(\mathbf{I}_T + \psi \mathbf{D}'_1 \mathbf{D}_1)^{-1}(\mathbf{y} - \bar{y} \mathbf{1})$, respectively, represent a low-frequency component of $(\mathbf{y} - \bar{y} \mathbf{1})$, $(\mathbf{y} - \widehat{\boldsymbol{\tau}})$, and $(\mathbf{y} - \bar{y} \mathbf{1})$. (b) The results highlight that the mHP filter is more like the ES filter than the HP filter. (c)(ii) is mentioned in Kim et al. (2009) and is empirically illustrated in Yamada (2018b).

Concerning the endpoints, we have the following result:

COROLLARY 2.7. *$\widehat{\mathbf{x}}_{\text{mHP}}$ and $\widehat{\mathbf{x}}_{\text{HP}}$ satisfy the following inequality: $\|\mathbf{f}_1 \widehat{\mathbf{x}}_{\text{HP}}\|^2 + \|\mathbf{f}_T \widehat{\mathbf{x}}_{\text{HP}}\|^2 > \|\mathbf{f}_1 \widehat{\mathbf{x}}_{\text{mHP}}\|^2 + \|\mathbf{f}_T \widehat{\mathbf{x}}_{\text{mHP}}\|^2$ if $\widehat{\mathbf{x}}_{\text{mHP}} \neq \widehat{\mathbf{x}}_{\text{HP}}$.*

Proof. From (2), it follows that $\|\mathbf{Lx}\|^2 = \|\mathbf{f}_1 \mathbf{x}\|^2 + \|\mathbf{D}_2 \mathbf{x}\|^2 + \|\mathbf{f}_T \mathbf{x}\|^2$. Given Propositions 2.1 (i)(a) and 2.2 (i)(a), if $\widehat{\mathbf{x}}_{\text{mHP}} \neq \widehat{\mathbf{x}}_{\text{HP}}$, then

$$\begin{aligned} & \|y - \widehat{x}_{mHP}\|^2 + \lambda (\|f_1 \widehat{x}_{HP}\|^2 + \|D_2 \widehat{x}_{mHP}\|^2 + \|f_T \widehat{x}_{HP}\|^2) \\ & > \|y - \widehat{x}_{HP}\|^2 + \lambda (\|f_1 \widehat{x}_{HP}\|^2 + \|D_2 \widehat{x}_{HP}\|^2 + \|f_T \widehat{x}_{HP}\|^2) \\ & > \|y - \widehat{x}_{mHP}\|^2 + \lambda (\|f_1 \widehat{x}_{mHP}\|^2 + \|D_2 \widehat{x}_{mHP}\|^2 + \|f_T \widehat{x}_{mHP}\|^2). \end{aligned}$$

Given $\lambda > 0$, by subtracting $\|y - \widehat{x}_{mHP}\|^2 + \lambda \|D_2 \widehat{x}_{mHP}\|^2$ from these inequalities, we obtain the result. ■

Remark 2.8. The above result is derived from the fact that $(\Delta x_2)^2 + (\Delta x_T)^2$, which equals $\|f_1 x\|^2 + \|f_T x\|^2$, is included in the objective function of the mHP filter. Note that it holds for any finite sample size, T .

Nevertheless, there remains a close relationship between the mHP and HP filters, because the mHP filter is obtainable through a slight modification of the HP filter. Given the relation between D_2 and L shown in (2), we obtain the following results.

PROPOSITION 2.9. (i) \widehat{x}_{mHP} can be represented with \widehat{x}_{HP} as $\widehat{x}_{mHP} = \widehat{x}_{HP} - A^{-1} F (\lambda^{-1} I_2 + F' A^{-1} F)^{-1} F' \widehat{x}_{HP}$, where $A = I_T + \lambda D_2' D_2$. (ii) $\widehat{x}_{mHP} = \widehat{x}_{HP}$ if and only if $f_1 \widehat{x}_{HP} = 0$ and $f_T \widehat{x}_{HP} = 0$.

Proof. See the Appendix. ■

Example 2.10. The case where $y = [1, 2, -2, 5, 1, 2]'$ and $\lambda = 1$ is an example such that $\widehat{x}_{mHP} = \widehat{x}_{HP}$. In this case, because $\widehat{x}_{HP} = [1, 1, 1, 2, 2, 2]'$, Proposition 2.9 (ii) ensures that $\widehat{x}_{mHP} = \widehat{x}_{HP}$.

3. GSF AND DCT

3.1. GSF

We now review the graph spectral filter (GSF) presented in Shuman et al. (2013). Let $\mathcal{L} \in \mathbb{R}^{n \times n}$ be a graph Laplacian. \mathcal{L} is a real symmetric matrix and thus its eigenvalues, denoted by $\kappa_1, \dots, \kappa_T$ in ascending order, are real and the associated normalized eigenvectors, denoted by v_1, \dots, v_T , are orthogonal to each other. These are, respectively, referred to as graph spectral eigenvalues and graph spectral eigenvectors of \mathcal{L} . Because \mathcal{L} is a matrix such that $\mathcal{L} \mathbf{1} = \mathbf{0}$ and $\mathcal{L} = \mathcal{D}' \mathcal{D}$, where \mathcal{D}' is an incidence matrix associated with \mathcal{L} , (i) \mathcal{L} is a positive semidefinite matrix, (ii) κ_1 , which is the smallest eigenvalue of \mathcal{L} , equals 0, (iii) the associated normalized eigenvector with κ_1 , denoted by v_1 , equals $\sqrt{\frac{1}{T}} \mathbf{1}$, (iv) the eigenvectors associated with the other eigenvalues, denoted by v_2, \dots, v_T , belong to $\mathcal{S}^\perp(\mathbf{1})$, and (v) the eigenvectors satisfy the following inequality: $0 = \|\mathcal{D} v_1\|^2 \leq \|\mathcal{D} v_2\|^2 \leq \dots \leq \|\mathcal{D} v_T\|^2$.

The GSF associated with \mathcal{L} is defined by $\mathcal{U} \mathcal{H} \mathcal{U}' y$, where $\mathcal{U} = [v_1, \dots, v_T]$ and $\mathcal{H} = \text{diag}\{h(\kappa_1), \dots, h(\kappa_T)\}$. h in \mathcal{H} is a transfer function of the filter that amplifies or attenuates the entries of $\mathcal{U}' y$. That is, the GSF consists of the following three

steps: (i) y is linearly transformed into $\mathcal{U}'y$, which is referred to as graph Fourier transform of y , (ii) $\mathcal{U}'y$ is filtered by premultiplying by \mathcal{H} , and (iii) $\mathcal{H}\mathcal{U}'y$ is linearly transformed into $\mathcal{U}\mathcal{H}\mathcal{U}'y$, which is referred to as inverse graph Fourier transform of $\mathcal{H}\mathcal{U}'y$.

3.2. DCT

Ahmed, Natarajan, and Rao (1974) developed the DCT, defined by

$$c_1 = \sqrt{\frac{1}{T}} \sum_{t=1}^T y_t, \quad c_k = \sqrt{\frac{2}{T}} \sum_{t=1}^T \cos\{(k-1)\theta_t\} y_t, \quad k = 2, \dots, T, \tag{5}$$

where $\theta_t = \frac{(t-\frac{1}{2})\pi}{T}$ for $t = 1, \dots, T$. Then, by letting $c = [c_1, \dots, c_T]'$ and

$$U = \begin{bmatrix} \sqrt{\frac{1}{T}} & \sqrt{\frac{2}{T}} \cos\{(2-1)\theta_1\} & \cdots & \sqrt{\frac{2}{T}} \cos\{(T-1)\theta_1\} \\ \sqrt{\frac{1}{T}} & \sqrt{\frac{2}{T}} \cos\{(2-1)\theta_2\} & \cdots & \sqrt{\frac{2}{T}} \cos\{(T-1)\theta_2\} \\ \vdots & \vdots & & \vdots \\ \sqrt{\frac{1}{T}} & \sqrt{\frac{2}{T}} \cos\{(2-1)\theta_T\} & \cdots & \sqrt{\frac{2}{T}} \cos\{(T-1)\theta_T\} \end{bmatrix} \in \mathbb{R}^{T \times T}, \tag{6}$$

DCT is represented in matrix notation as $c = U'y$ and for this reason U' is referred to as the DCT matrix.⁵

3.3. Chebyshev Polynomials and DCT

Let $C_k(x) = \cos(k \arccos x)$ where $|x| \leq 1$ and $k \in \mathbb{N}$, which are orthogonal polynomials and are referred to as the Chebyshev polynomials of the first kind. These are polynomials of degree k and the coefficient of x^k equals 2^{k-1} . For example, $C_4(x) = 8x^4 - 8x^2 + 1$. It is noteworthy here that U in (6) can be expressed with these (Hamming, 1973, pp. 472–473; Ahmed et al., 1974):

$$U = \begin{bmatrix} \sqrt{\frac{1}{T}} C_0(x_1) & \sqrt{\frac{2}{T}} C_1(x_1) & \cdots & \sqrt{\frac{2}{T}} C_{T-1}(x_1) \\ \sqrt{\frac{1}{T}} C_0(x_2) & \sqrt{\frac{2}{T}} C_1(x_2) & \cdots & \sqrt{\frac{2}{T}} C_{T-1}(x_2) \\ \vdots & \vdots & & \vdots \\ \sqrt{\frac{1}{T}} C_0(x_T) & \sqrt{\frac{2}{T}} C_1(x_T) & \cdots & \sqrt{\frac{2}{T}} C_{T-1}(x_T) \end{bmatrix},$$

where $x_i = \cos \theta_i$ for $i = 1, \dots, T$ are zeros of $C_T(x)$ and are referred to as Chebyshev nodes/roots.⁶ Recall that $\theta_i = \frac{(i-\frac{1}{2})\pi}{T}$ for $i = 1, \dots, T$.

⁵There are eight DCTs (DCT-1 through DCT-8) as shown in Strang (1999) and the DCT in (5) is, more precisely, DCT-2.

⁶For more details about Chebyshev polynomials of the first kind, see, for example, Mason and Handscomb (2003).

3.4. DCT as a Type of Graph Fourier Transform

Notably, from von Neumann (1941), Anderson (1963), Jain (1979), and O’Sullivan (1991), the spectral decomposition of L in (2) can be described with the DCT matrix as⁷

$$L = UGU', \tag{7}$$

where, by letting $g_k = 2 - 2 \cos \left\{ \frac{(k-1)\pi}{T} \right\} = 4 \sin^2 \left\{ \frac{(k-1)\pi}{2T} \right\}$ for $k = 1, \dots, T$, $G = \text{diag}(g_1, \dots, g_T)$. From the current point of view, the graph Fourier transform matrix associated with a graph Laplacian, L , is the DCT matrix.

Let u_k for $k = 1, \dots, T$ denote the k th column of U in (6). From (6), it is observable that (i) $u_1 = \sqrt{\frac{1}{T}}\mathbf{1}$ and (ii) u_2, \dots, u_T represent cosine curves of different periods as illustrated in Figure 3. More precisely, the period of u_{k+1} is $\frac{2T}{k}$ for $k = 1, \dots, T - 1$. For example, the period of u_5 is $\frac{T}{2}$. See again Figure 3. Because L is a graph Laplacian, u_1, \dots, u_T may be referred to as graph Laplacian eigenvectors. Recall that from $0 = g_1 < \dots < g_T < 4$ and $g_k = u_k' L u_k = u_k' D_1' D_1 u_k = \|D_1 u_k\|^2 = \sum_{t=2}^T (\Delta u_{k,t})^2$, where $u_k = [u_{k,1}, \dots, u_{k,T}]'$, for $k = 1, \dots, T$, eigenvectors, u_1, \dots, u_T , satisfy

$$0 = \sum_{t=2}^T (\Delta u_{1,t})^2 < \sum_{t=2}^T (\Delta u_{2,t})^2 < \dots < \sum_{t=2}^T (\Delta u_{T,t})^2 < 4. \tag{8}$$

These inequalities indicate that u_k is smoother than u_l if $k < l$ and we can observe that these inequalities hold from Figure 3.

Finally, we stress that among the column vectors, u_2, \dots, u_T , u_2 has a special feature. As depicted in Figure 3, this is a monotonic function of time and thus $u_1(u_1' u_1)^{-1} u_1' y + u_2(u_2' u_2)^{-1} u_2' y = \bar{y}\mathbf{1} + c_2 u_2$ can represent a trend in economic time series, y . Note that $u_2 \in \mathcal{S}^\perp(\mathbf{1})$.

4. THE THREE FILTER S AS A TYPE OF GSF

4.1. mHP Filter

From (7), it follows that $L'L = L^2 = (UGU')^2 = UG^2U'$. Then, \hat{x}_{mHP} can be expressed with the DCT matrix as

$$\hat{x}_{\text{mHP}} = U(I_T + \lambda G^2)^{-1} U' y = U H_{\text{mHP}} U' y, \tag{9}$$

where $H_{\text{mHP}} = (I_T + \lambda G^2)^{-1}$, and thus it can also be represented as $\hat{x}_{\text{mHP}} = \text{IDCT}\{H_{\text{mHP}} \text{DCT}(y)\}$, where DCT and IDCT, respectively, represent DCT and inverse DCT (Garcia, 2010). Thus, we have the following result:

PROPOSITION 4.1. *The mHP filter is a GSF based on the path graph.*

Proof. It follows from (9). ■

⁷See also Strang (1999) and Nakatsukasa et al. (2013).

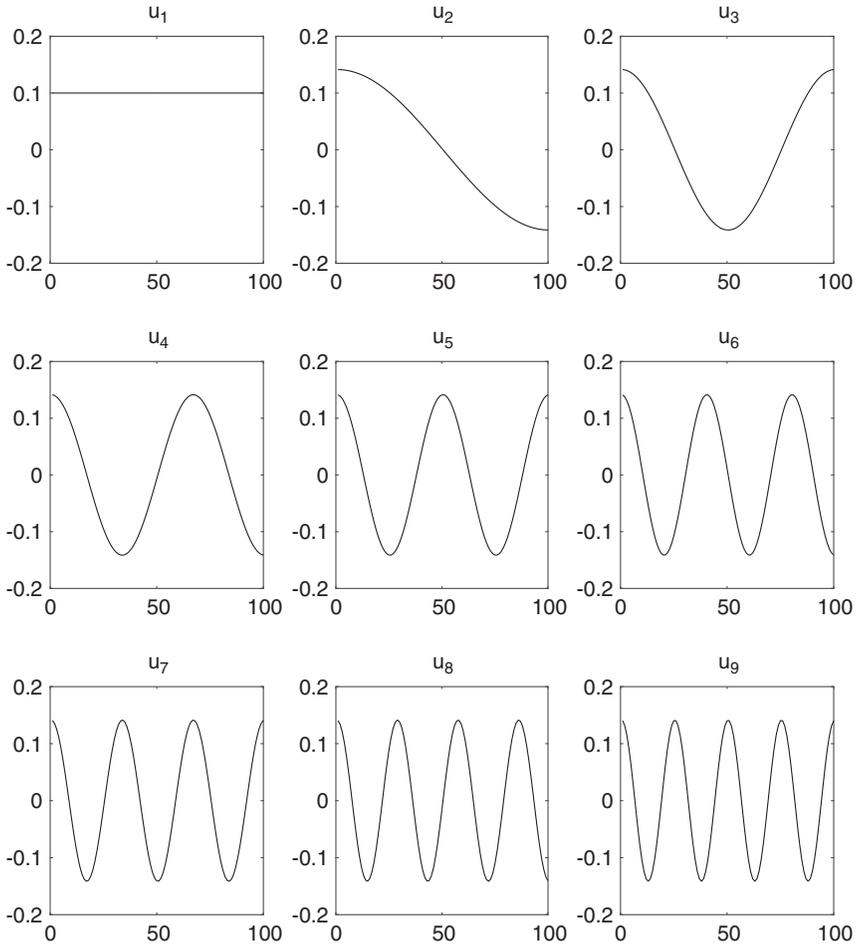


FIGURE 3. The first nine columns of $U = [u_1, \dots, u_T]$ in (6) for $T = 100$.

Remark 4.2. Equation (9) implies that \hat{x}_{mHP} can be expressed as a linear combination of orthonormal vectors, u_1, \dots, u_T . Of course, this is not so important because any vectors in T -dimensional euclidean space can be expressed as a linear combination of them. The HP filter can also be expressed as a linear combination of these as follows:

$$\hat{x}_{HP} = U(I_T + \lambda U' D_2' D_2 U)^{-1} U' y. \tag{10}$$

Equation (10) shows that the HP filter is not a GSF based on the path graph because $U' D_2' D_2 U$ is not a diagonal matrix.

The following proposition is derived from (9).

PROPOSITION 4.3. Let $\widehat{\boldsymbol{\xi}}_{\text{mHP},k} = \widehat{c}_{\text{mHP},k} \mathbf{u}_k$ for $k = 1, \dots, T$, where $\widehat{c}_{\text{mHP},k} = \frac{c_k}{1+\lambda g_k^2}$. (i) $\widehat{\mathbf{x}}_{\text{mHP}}$ can be decomposed as $\widehat{\mathbf{x}}_{\text{mHP}} = \widehat{\boldsymbol{\xi}}_{\text{mHP},1} + \dots + \widehat{\boldsymbol{\xi}}_{\text{mHP},T}$, where $\widehat{\boldsymbol{\xi}}_{\text{mHP},k} \widehat{\boldsymbol{\xi}}_{\text{mHP},l} = 0$ if $k \neq l$ and $\widehat{\boldsymbol{\xi}}_{\text{mHP},k} \widehat{\boldsymbol{\xi}}_{\text{mHP},l} = \widehat{c}_{\text{mHP},k}^2$ if $k = l$. (ii) (a) $\widehat{\boldsymbol{\xi}}_{\text{mHP},1} = \bar{y} \boldsymbol{\iota}$ and thus $\widehat{\boldsymbol{\xi}}_{\text{mHP},1} \in \mathcal{S}(\boldsymbol{\iota})$ and $\frac{1}{T} \boldsymbol{\iota}' \widehat{\boldsymbol{\xi}}_{\text{mHP},1} = \bar{y}$. (b) For $k = 2, \dots, T$, $\boldsymbol{\iota}' \widehat{\boldsymbol{\xi}}_{\text{mHP},k} = 0$ and thus $\widehat{\boldsymbol{\xi}}_{\text{mHP},k} \in \mathcal{S}^\perp(\boldsymbol{\iota})$ and $\frac{1}{T} \boldsymbol{\iota}' \widehat{\boldsymbol{\xi}}_{\text{mHP},k} = 0$. (iii) The sum of the squared deviations from the sample mean of $\widehat{\mathbf{x}}_{\text{mHP}}$ can be represented as $\|\mathbf{Q}_T \widehat{\mathbf{x}}_{\text{mHP}}\|^2 = \|\widehat{\boldsymbol{\xi}}_{\text{mHP},2}\|^2 + \dots + \|\widehat{\boldsymbol{\xi}}_{\text{mHP},T}\|^2 = \widehat{c}_{\text{mHP},2}^2 + \dots + \widehat{c}_{\text{mHP},T}^2$. (iv) By denoting the (i, j) -entry of $(\mathbf{I}_T + \lambda \mathbf{L}' \mathbf{L})^{-1}$ by $\widehat{h}_{\text{mHP},i,j}$ for $i, j = 1, \dots, T$, it follows that

$$\widehat{h}_{\text{mHP},i,j} = \frac{1}{T} + \frac{2}{T} \sum_{k=2}^T \frac{\cos\{(k-1)\theta_i\} \cos\{(k-1)\theta_j\}}{1 + \lambda g_k^2}, \tag{11}$$

$$\widehat{h}_{\text{mHP},i,j} \rightarrow 2 \int_0^1 \frac{\cos\{(i-\frac{1}{2})r\pi\} \cos\{(j-\frac{1}{2})r\pi\}}{1 + \lambda \{2 - 2 \cos(r\pi)\}^2} dr \quad (T \rightarrow \infty), \tag{12}$$

where $\theta_t = \frac{(t-\frac{1}{2})\pi}{T}$ for $t = 1, \dots, T$.

Proof. See the Appendix. ■

Remark 4.4. (a) In contrast, the (i, j) -entry of $(\mathbf{I}_T + \lambda \mathbf{D}'_2 \mathbf{D}_2)^{-1}$ is much more complicated. See de Jong and Sakarya (2016, Thm. 1), Cornea-Madeira (2017, Cor. 1), and Yamada and Jahra (2019, Cor. 3.2). (b) By recalling that $2 \cos A \cos B = \cos(A + B) + \cos(A - B)$ and letting $\zeta = r\pi$, (12) can be alternatively represented as

$$\widehat{h}_{\text{mHP},i,j} \rightarrow \frac{1}{\pi} \int_0^\pi \frac{\cos\{(i+j-1)\zeta\} + \cos\{(i-j)\zeta\}}{1 + \lambda (2 - 2 \cos \zeta)^2} d\zeta \quad (T \rightarrow \infty), \tag{13}$$

which corresponds to equation (9.4) in Strang and MacNamara (2014). In (13),

$$\frac{1}{\pi} \int_0^\pi \frac{\cos\{(i-j)\zeta\}}{1 + \lambda (2 - 2 \cos \zeta)^2} d\zeta \quad \text{and} \quad \frac{1}{\pi} \int_0^\pi \frac{\cos\{(i+j-1)\zeta\}}{1 + \lambda (2 - 2 \cos \zeta)^2} d\zeta,$$

respectively, represent “the Toeplitz part” and “the Hankel part” (Strang and MacNamara, 2014).

The following two propositions provide alternative representations of the mHP.

PROPOSITION 4.5. By letting $\widehat{\boldsymbol{\phi}}_{\text{mHP}} = (\mathbf{I}_T + \lambda \mathbf{G}^2)^{-1} \mathbf{U}' \mathbf{y}$, it follows that $\widehat{\mathbf{x}}_{\text{mHP}} = \mathbf{U} \widehat{\boldsymbol{\phi}}_{\text{mHP}}$ and $\widehat{\boldsymbol{\phi}}_{\text{mHP}}$ is a unique global minimizer of $\min_{\boldsymbol{\phi} \in \mathbb{R}^T} \|\mathbf{y} - \mathbf{U} \boldsymbol{\phi}\|^2 + \lambda \|\mathbf{G} \boldsymbol{\phi}\|^2$.

Proof. See the Appendix. ■

Remark 4.6. Let $\widehat{\boldsymbol{\phi}}_{\text{mHP}} = [\widehat{\phi}_{\text{mHP},1}, \dots, \widehat{\phi}_{\text{mHP},T}]'$. Then, given that \mathbf{U} is an orthogonal matrix, it follows that $\widehat{\phi}_{\text{mHP},k} = \arg \min_{\phi_k \in \mathbb{R}} \|\mathbf{y} - \mathbf{u}_k \phi_k\|^2 + \lambda (g_k \phi_k)^2 = \frac{c_k}{1+\lambda g_k^2} = \widehat{c}_{\text{mHP},k}$ for $k = 1, \dots, T$.

PROPOSITION 4.7. Let $\mathbf{E} = [g_2\mathbf{u}_2, \dots, g_T\mathbf{u}_T]' \in \mathbb{R}^{(T-1) \times T}$. Then, it follows that $\hat{\mathbf{x}}_{\text{mHP}} = \arg \min_{\mathbf{x} \in \mathbb{R}^T} \|\mathbf{y} - \mathbf{x}\|^2 + \lambda \|\mathbf{E}\mathbf{x}\|^2 = (\mathbf{I}_T + \lambda \mathbf{E}'\mathbf{E})^{-1} \mathbf{y}$.

Proof. See the Appendix. ■

Remark 4.8. (a) \mathbf{E} is of full row rank and is a matrix such that $\mathbf{E}\mathbf{1} = \mathbf{0}$. (b) By applying the Sherman–Morrison–Woodbury formula (Seber, 2008) to $(\mathbf{I}_T + \lambda \mathbf{E}'\mathbf{E})^{-1}$, we obtain $\hat{\mathbf{x}}_{\text{mHP}} = \mathbf{y} - \mathbf{E}'(\lambda^{-1}\mathbf{I}_{T-1} + \mathbf{E}\mathbf{E}')^{-1}\mathbf{E}\mathbf{y} = \mathbf{y} - \mathbf{E}'\hat{\boldsymbol{\delta}}$, where $\hat{\boldsymbol{\delta}} = (\lambda^{-1}\mathbf{I}_{T-1} + \mathbf{E}\mathbf{E}')^{-1}\mathbf{E}\mathbf{y}$. Thus, $\hat{\boldsymbol{\delta}}_{\text{mHP}} (= \mathbf{y} - \hat{\mathbf{x}}_{\text{mHP}}) = \mathbf{E}'\hat{\boldsymbol{\delta}}$. It is of interest that $\hat{\boldsymbol{\delta}}$ is a unique global minimizer of the following penalized least squares problem: $\min_{\boldsymbol{\delta} \in \mathbb{R}^{T-1}} \|\mathbf{y} - \mathbf{E}'\boldsymbol{\delta}\|^2 + \lambda^{-1}\|\boldsymbol{\delta}\|^2$. Likewise, we can derive several closely related least squares problems as shown in Yamada (2018a).

4.2. ES Filter

As with (9), $\hat{\mathbf{x}}_{\text{ES}}$ can be alternatively expressed as

$$\hat{\mathbf{x}}_{\text{ES}} = \mathbf{U}(\mathbf{I}_T + \psi \mathbf{G})^{-1} \mathbf{U}' \mathbf{y} = \mathbf{U} \mathbf{H}_{\text{ES}} \mathbf{U}' \mathbf{y}, \tag{14}$$

where $\mathbf{H}_{\text{ES}} = (\mathbf{I}_T + \psi \mathbf{G})^{-1}$, and thus it can also be represented as $\hat{\mathbf{x}}_{\text{ES}} = \text{IDCT}\{\mathbf{H}_{\text{ES}} \text{DCT}(\mathbf{y})\}$. Thus, we have the following result:

PROPOSITION 4.9. The ES filter is a GSF based on the path graph.

Proof. It follows from (14). ■

Accordingly, we obtain the following results.

PROPOSITION 4.10. Let $\hat{\boldsymbol{\xi}}_{\text{ES},k} = \hat{c}_{\text{ES},k} \mathbf{u}_k$ for $k = 1, \dots, T$, where $\hat{c}_{\text{ES},k} = \frac{c_k}{1 + \psi g_k}$. (i) $\hat{\mathbf{x}}_{\text{ES}}$ can be decomposed as $\hat{\mathbf{x}}_{\text{ES}} = \hat{\boldsymbol{\xi}}_{\text{ES},1} + \dots + \hat{\boldsymbol{\xi}}_{\text{ES},T}$, where $\hat{\boldsymbol{\xi}}'_{\text{ES},k} \hat{\boldsymbol{\xi}}_{\text{ES},l} = 0$ if $k \neq l$ and $\hat{\boldsymbol{\xi}}'_{\text{ES},k} \hat{\boldsymbol{\xi}}_{\text{ES},l} = \hat{c}_{\text{ES},k}^2$ if $k = l$. (ii) $\hat{\boldsymbol{\xi}}_{\text{ES},1} = \bar{y} \mathbf{1}$ and thus $\frac{1}{T} \mathbf{1}' \hat{\boldsymbol{\xi}}_{\text{ES},1} = \bar{y}$, whereas $\frac{1}{T} \mathbf{1}' \hat{\boldsymbol{\xi}}_{\text{ES},k} = 0$ for $k = 2, \dots, T$. (iii) The sum of the squared deviations from the sample mean of $\hat{\mathbf{x}}_{\text{ES}}$ can be represented as $\|\mathbf{Q}_t \hat{\mathbf{x}}_{\text{ES}}\|^2 = \|\hat{\boldsymbol{\xi}}_{\text{ES},2}\|^2 + \dots + \|\hat{\boldsymbol{\xi}}_{\text{ES},T}\|^2 = \hat{c}_{\text{ES},2}^2 + \dots + \hat{c}_{\text{ES},T}^2$. (iv) By denoting the (i, j) -entry of $(\mathbf{I}_T + \psi \mathbf{D}' \mathbf{D}_1)^{-1}$ by $\hat{h}_{\text{ES},i,j}$ for $i, j = 1, \dots, T$, it follows that

$$\hat{h}_{\text{ES},i,j} = \frac{1}{T} + \frac{2}{T} \sum_{k=2}^T \frac{\cos\{(k-1)\theta_i\} \cos\{(k-1)\theta_j\}}{1 + \psi g_k},$$

$$\hat{h}_{\text{ES},i,j} \rightarrow 2 \int_0^1 \frac{\cos\{(i - \frac{1}{2})r\pi\} \cos\{(j - \frac{1}{2})r\pi\}}{1 + \psi \{2 - 2 \cos(r\pi)\}} dr \quad (T \rightarrow \infty),$$

where $\theta_t = \frac{(t - \frac{1}{2})\pi}{T}$ for $t = 1, \dots, T$.

Proof. This proposition can be proved as for Proposition 4.3. ■

Remark 4.11. (a) Yamada and Jahra (2018, Cor. 2.2) provided an alternative but more complex representation of $\hat{h}_{\text{ES},i,j}$ in Proposition 4.10(iv). In deriving it, they

applied not the spectral decomposition of $D_1'D_1(=L)$ but that of D_1D_1' , which is a tri-diagonal symmetric Toeplitz matrix. (b) As with (13), we obtain

$$\hat{h}_{ES,i,j} \rightarrow \frac{1}{\pi} \int_0^\pi \frac{\cos\{(i+j-1)\xi\} + \cos\{(i-j)\xi\}}{1 + \psi(2 - 2\cos\xi)} d\xi \quad (T \rightarrow \infty).$$

PROPOSITION 4.12. *Let $\hat{\phi}_{ES} = (I_T + \psi G)^{-1}U'y$. Then, it follows that $\hat{x}_{ES} = U\hat{\phi}_{ES}$ and $\hat{\phi}_{ES}$ is a unique global minimizer of $\min_{\phi \in \mathbb{R}^T} \|y - U\phi\|^2 + \psi\phi'G\phi$.*

Proof. This proposition can be proved as for Proposition 4.5. ■

Remark 4.13. Let $\hat{\phi}_{ES} = [\hat{\phi}_{ES,1}, \dots, \hat{\phi}_{ES,T}]'$. Then, because U is an orthogonal matrix, it follows that $\hat{\phi}_{ES,k} = \arg \min_{\phi_k \in \mathbb{R}} \|y - u_k\phi_k\|^2 + \psi g_k\phi_k^2 = \frac{c_k}{1 + \psi g_k} = \hat{c}_{ES,k}$ for $k = 1, \dots, T$.

4.3. LFP Filter

The LFP of Müller and Watson (2018) is similar to the mHP and ES filters. Let $\Psi = T^{\frac{1}{2}}[u_2, \dots, u_{q+1}] \in \mathbb{R}^{T \times q}$. Recall that u_k denotes the k th column of U in (6). By the definition of Ψ , it follows that $S(\Psi) \subseteq S^\perp(t)$ and $\Psi'\Psi = TI_q$.⁸ Then, the LFP filter can be described as

$$\text{LFP: } \hat{x}_{LFP} = \Psi(\Psi'\Psi)^{-1}\Psi'y = c_2u_2 + \dots + c_{q+1}u_{q+1} = UH_{LFP}U'y, \tag{15}$$

where $H_{LFP} = \text{diag}(0, \underbrace{1, \dots, 1}_q, 0, \dots, 0) \in \mathbb{R}^{T \times T}$, and thus it can also be represented as $\hat{x}_{LFP} = \text{IDCT}\{H_{LFP} \text{DCT}(y)\}$. Accordingly, we have the following result:

PROPOSITION 4.14. *The LFP filter is a GSF based on the path graph.*

Proof. It follows from (15). ■

Because the period of u_{q+1} , which is the last column in Ψ , is $\frac{2T}{q}$, the LFP filter is designed to isolate variation in the series with periods longer than $\frac{2T}{q}$ (Müller and Watson, 2018). See again Figure 3.

4.4. Relationship Between the Three Filters

Recall that \hat{x}_{mHP} , \hat{x}_{ES} , and \hat{x}_{LFP} are GSFs based on the path graph. These are, respectively, represented as a linear combination of T -dimensional column vectors, $c_1u_1, c_2u_2, \dots, c_Tu_T$, as follows:

$$\hat{x}_i = \sum_{k=1}^T w_{i,k}c_ku_k, \quad i = \text{mHP, ES}, \tag{16}$$

⁸The equality in $S(\Psi) \subseteq S^\perp(t)$ holds if and only if $q = T - 1$.

$$\widehat{\mathbf{x}}_{\text{LFP}} + \bar{y}t = \sum_{k=1}^T w_{\text{LFP},k} c_k \mathbf{u}_k, \tag{17}$$

where $w_{\text{mHP},k} = (1 + \lambda g_k^2)^{-1}$ and $w_{\text{ES},k} = (1 + \psi g_k)^{-1}$ for $k = 1, \dots, T$ and

$$w_{\text{LFP},k} = \begin{cases} 1, & k = 1, \dots, q + 1, \\ 0, & k = q + 2, \dots, T. \end{cases} \tag{18}$$

Recall that both λ and ψ are positive. In addition, the inequalities, $0 = g_1 < g_2 < \dots < g_T < 4$, hold. Accordingly, we obtain

$$1 = w_{\text{mHP},1} > w_{\text{mHP},2} > \dots > w_{\text{mHP},T} > 0, \tag{19}$$

$$1 = w_{\text{ES},1} > w_{\text{ES},2} > \dots > w_{\text{ES},T} > 0. \tag{20}$$

Thus, it could be expected that $\widehat{\mathbf{x}}_{\text{LFP}} + \bar{y}t$ is roughly similar to $\widehat{\mathbf{x}}_{\text{mHP}}$ and $\widehat{\mathbf{x}}_{\text{ES}}$ if λ and ψ are specified so that $w_{\text{mHP},q+1} = \frac{1}{2}$ and $w_{\text{ES},q+1} = \frac{1}{2}$. Note that these λ and ψ are explicitly expressed as

$$\lambda = \frac{1}{g_{q+1}^2} \quad \text{and} \quad \psi = \frac{1}{g_{q+1}}, \tag{21}$$

where $g_{q+1} = 2 - 2 \cos\left(\frac{q\pi}{T}\right) = 4 \sin^2\left(\frac{q\pi}{2T}\right)$. Concerning (21), we have the following results.

PROPOSITION 4.15. *Let $f = \frac{2\pi}{p}$, where $p = \frac{2T}{q}$. Then, (21) can be represented as follows:*

$$\lambda = \left\{ 2 \sin\left(\frac{f}{2}\right) \right\}^{-4} \quad \text{and} \quad \psi = \left\{ 2 \sin\left(\frac{f}{2}\right) \right\}^{-2}. \tag{22}$$

Proof. We only provide the former equation of (22). From (21), it follows that $\lambda = \left\{ 2 \sin\left(\frac{q\pi}{2T}\right) \right\}^{-4}$. Then, from $f = \frac{2\pi}{p} = 2\pi \times \frac{q}{2T} = \frac{q\pi}{T}$, we obtain the former equation of (22). ■

Remark 4.16. Very interestingly, the equations in (22) are equivalent to the ways of specifying the smoothing parameters for the HP and ES filters based on the gain functions shown by King and Rebelo (1993). See, for example, Gómez (2001) and Yamada (2012) for details.

Example 4.17. *Consider the case where $T = 100$ and $p = 40$ (quarters). Then, the corresponding values of the parameters are $\lambda = 1649.3 \approx 1600$, $\psi = 40.6$, and $q = 5$.*

On the coefficients in (16) and (17), there are the following relationships:

PROPOSITION 4.18. (i) *Let $\lambda = \frac{1}{g_{q+1}^2}$ and $\psi = \frac{1}{g_{q+1}}$. Then, it follows that $w_{\text{LFP},k} = w_{\text{mHP},k} = w_{\text{ES},k} = 1$ if $k = 1$, $1 = w_{\text{LFP},k} > w_{\text{mHP},k} > w_{\text{ES},k} > \frac{1}{2}$ if*

$k = 2, \dots, q, 1 = w_{LFP,k} > w_{mHP,k} = w_{ES,k} = \frac{1}{2}$ if $k = q + 1$, and $\frac{1}{2} > w_{mHP,k} > w_{ES,k} > w_{LFP,k} = 0$ if $k = q + 2, \dots, T$. (ii) For $k = 1, \dots, T - 1, w_{mHP,k} > w_{mHP,k+1}$ and $w_{ES,k} > w_{ES,k+1}$.

Proof. See the Appendix. ■

Consequently, when $\lambda = \frac{1}{g_{q+1}^2}$ and $\psi = \frac{1}{g_{q+1}}$, it follows that (i) $\hat{x}_{LFP} + \bar{y}t$ is roughly similar to \hat{x}_{mHP} and \hat{x}_{ES} , (ii) $\hat{x}_{LFP} + \bar{y}t$ is smoother than \hat{x}_{mHP} and \hat{x}_{ES} , and (iii) \hat{x}_{mHP} is smoother than \hat{x}_{ES} . Among them, (ii) is, for example, due to the fact that $\hat{x}_{LFP} + \bar{y}t$ does not consist of a linear combination of vectors, u_{q+2}, \dots, u_T , whereas both \hat{x}_{mHP} and \hat{x}_{ES} depend on the vectors.

4.5. Empirical Illustration

See Figures 4 and 5. In Figure 4 (Figure 5), y denotes the growth rates of US real GDP (real consumption) used in the empirical analysis in Müller and Watson (2018). LFP represents $\hat{x}_{LFP} + \bar{y}t$. It is estimated by setting $q = 12$. Thus, given that $T = 272$, in this case, the LFP filter is designed to isolate variation in the series with periods longer than $\frac{2 \times 272}{12} \approx 45$ quarters ≈ 11.3 years (Müller and Watson, 2018). We remark that LFP in Figure 4 (Figure 5) is identical to the solid (dashed) line depicted in Figure 1(c) of Müller and Watson (2018). ES (mHP) denotes

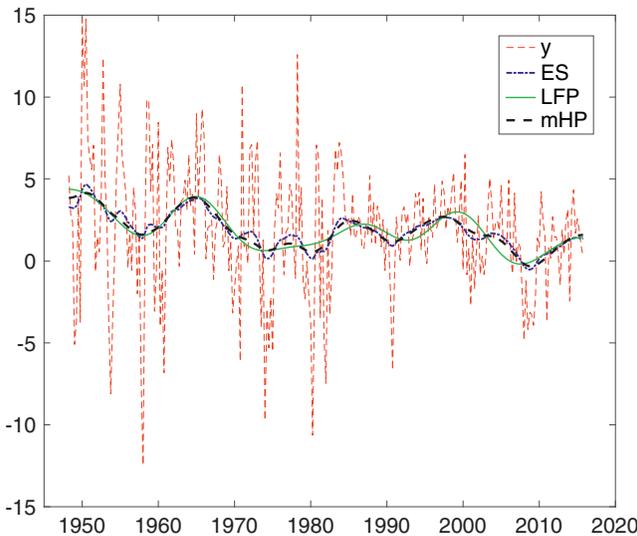


FIGURE 4. y denotes growth rates of U.S. real GDP used in the empirical analysis in Müller and Watson (2018). LFP represents $\hat{x}_{LFP} + \bar{y}t$ estimated by setting $q = 12$. We remark that it is identical to the solid line depicted in Figure 1(c) in Müller and Watson (2018). ES (mHP) represents \hat{x}_{ES} (\hat{x}_{mHP}) estimated by setting $\psi = 51.8$ ($\lambda = 2678.9$).

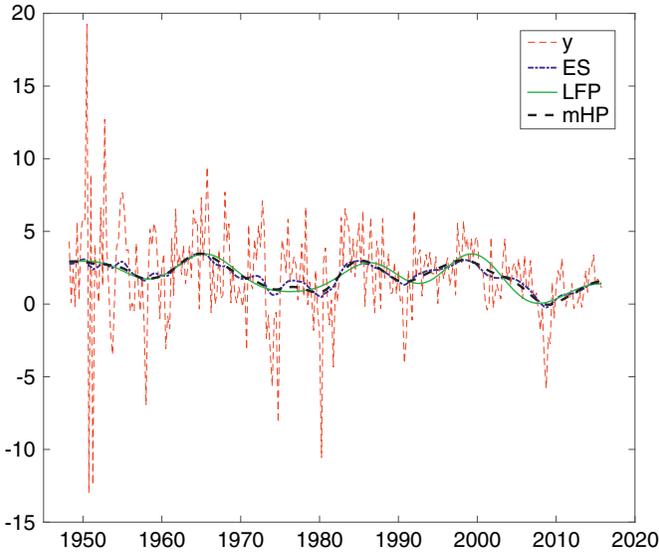


FIGURE 5. y denotes growth rates of U.S. real consumption used in the empirical analysis in Müller and Watson (2018). LFP represents $\hat{x}_{LFP} + \bar{y}t$ estimated by setting $q = 12$. We remark that it is identical to the dashed line depicted in Figure 1(c) in Müller and Watson (2018). ES (mHP) represents \hat{x}_{ES} (\hat{x}_{mHP}) estimated by setting $\psi = 51.8$ ($\lambda = 2678.9$).

\hat{x}_{ES} (\hat{x}_{mHP}) estimated by setting $\psi = 51.8$ ($\lambda = 2678.9$). Here, these values of parameters are calculated by (22). From these figures, we can empirically confirm the theoretical results shown in Section 4.4, that is, (i) $\hat{x}_{LFP} + \bar{y}t$ is roughly similar to \hat{x}_{mHP} and \hat{x}_{ES} , (ii) $\hat{x}_{LFP} + \bar{y}t$ is smoother than \hat{x}_{mHP} and \hat{x}_{ES} , and (iii) \hat{x}_{mHP} is smoother than \hat{x}_{ES} .

4.6. DCT is More Appropriate Than DFT for Econometric Time Series Analysis

As an alternative LFP to (15), we may consider the following orthogonal projection:

$$\hat{x}_{LFP2} = \Phi(\Phi^* \Phi)^{-1} \Phi^* y = d_2 v_2 + \dots + d_{q+1} v_{q+1} = W H_{LFP} W^* y, \tag{23}$$

where $W = [v_1, \dots, v_T] \in \mathbb{C}^{T \times T}$ is defined in (A.1), $\Phi = T^{\frac{1}{2}} [v_2, \dots, v_{q+1}] \in \mathbb{C}^{T \times q}$, $W^* y = [d_1, \dots, d_T] \in \mathbb{C}^T$, which is the DFT of y , and Φ^* denotes the conjugate transpose of Φ . DFT has a much longer history than DCT and the form of LFP given by (23) and its variant have appeared in econometrics, for example, in Engle (1974), Harvey (1978), and Corbae, Ouliaris, and Phillips (2002).

As DCT is a graph Fourier transform, DFT is also a graph Fourier transform. The former is based on the path graph (Figure 1), whereas the latter is based on

the cycle graph (Figure 2).⁹ We believe that the path graph is more appropriate than the cycle graph as an underlying graph of economic time series, and for this reason, we consider the DCT more appropriate than the DFT for econometric time series analysis.

5. CONCLUDING REMARKS

This article originated in the discovery of a somewhat familiar smoothing method that looks like the popular HP filter, denoted as the mHP filter in this article, and a remarkable matrix factorization of a graph Laplacian corresponding to the path graph shown in (7). Further investigation led to the realization that the path graph is more appropriate than the cycle graph as an underlying graph of economic time series. These discoveries provided a strong motivation to pursue this project.

The theoretical results of our examinations are summarized in Propositions 2.1, 2.9, 4.1, 4.3, 4.5, 4.7, 4.9, 4.10, 4.12, 4.14, 4.15, and 4.18 and Corollaries 2.5 and 2.7 and the empirical results are illustrated in Figures 4 and 5. In short, our findings revealed that the mHP filter is a GSF based on the DCT as for the ES and LFP filters and that it is more like the ES filter even though it appears like the HP filter. In addition, we provided a way of specifying the mHP filter's smoothing parameter, λ , by (22), which is necessary for its application.

REFERENCES

- Ahmed, N., T. Natarajan, & K.R. Rao (1974) Discrete cosine transform. *IEEE Transactions on Computers* C-23, 90–93.
- Aitken, A.C. (1927) On the theory of graduation. *Proceedings of the Royal Society of Edinburgh* 46, 36–45.
- Anderson, T.W. (1963) Determination of the order of dependence in normally distributed time series. In M. Rosenblatt (ed.), *Proceedings of the Symposium on Time Series Analysis*. Wiley, New York, NY.
- Bohlmann, G. (1899) Ein Ausgleichungsproblem. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1899, 260–271.
- Buckley, M.J. (1994) Fast computation of a discretized thin-plate smoothing spline for image data. *Biometrika* 81, 247–258.
- Corbae, D., S. Ouliaris, & P.C.B. Phillips (2002) Band spectral regression with trending data. *Econometrica* 70, 1067–1109.
- Cornea-Madeira, A. (2017) The explicit formula for the Hodrick–Prescott filter in finite sample. *Review of Economics and Statistics* 99, 314–318.
- Danthine, J.-P. & M. Girardin (1989) Business cycles in Switzerland: A comparative study. *European Economic Review* 33, 31–50.
- de Jong, R.M. & N. Sakarya (2016) The econometrics of the Hodrick–Prescott filter. *Review of Economics and Statistics* 98, 310–317.
- Durbin, J. & G.S. Watson (1950) Testing for serial correlation in least squares regression. I. *Biometrika* 37, 409–428.

⁹See the Appendix for details.

- Durbin, J. & G.S. Watson (1951) Testing for serial correlation in least squares regression. II. *Biometrika* 38, 159–179.
- Engle, R.F. (1974) Band spectrum regression. *International Economic Review* 15, 1–11.
- García, D. (2010) Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational Statistics and Data Analysis* 54, 1167–1178.
- Gómez, V. (2001) The use of Butterworth filters for trend and cycle estimation in economic time series. *Journal of Business and Economic Statistics* 19, 365–373.
- Hamilton, J.D. (2018) Why you should never use the Hodrick–Prescott filter. *Review of Economics and Statistics* 100, 831–843.
- Hamming, R.W. (1973) *Numerical Methods for Scientists and Engineers*. 2nd ed., McGraw-Hill, New York, NY.
- Harvey, A.C. (1978) Linear regression in the frequency domain. *International Economic Review* 19, 507–512.
- Henderson, R. (1924) A new method of graduation. *Transactions of the Actuarial Society of America* 25, 29–40.
- Henderson, R. (1925) Further remarks on graduation. *Transactions of the Actuarial Society of America* 26, 52–57.
- Hodrick, R.J. & E.C. Prescott (1997) Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking* 29, 1–16.
- Jain, A.K. (1979) A sinusoidal family of unitary transforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1* 4, 356–365.
- Kim, S., K. Koh, S. Boyd, & D. Gorinevsky (2009) ℓ_1 trend filtering. *SIAM Review* 51, 339–360.
- King, R.G. & S.T. Rebelo (1993) Low frequency filtering and real business cycles. *Journal of Economic Dynamics and Control* 17, 207–231.
- Mason, J.C. & D. Handscomb (2003) *Chebyshev Polynomials*. CRC Press, Boca Raton, FL.
- Müller, U. & M. Watson (2018) Long-run covariability. *Econometrica* 86, 775–804.
- Nakatsukasa, Y., N. Saito, & E. Woei (2013) Mysteries around the graph Laplacian eigenvalue 4. *Linear Algebra and its Applications* 438, 3231–3246.
- Nocon, A.S. & W.F. Scott. (2012) An extension of the Whittaker–Henderson method of graduation. *Scandinavian Actuarial Journal* 2012, 70–79.
- O’Sullivan, F. (1991) Discretized Laplacian smoothing by Fourier methods. *Journal of the American Statistical Association* 86, 634–642.
- Phillips, P.C.B. (1998) New tools for understanding spurious regressions. *Econometrica* 66, 1299–1325.
- Phillips, P.C.B. (2005a) HAC estimation by automated regression. *Econometric Theory* 21, 116–142.
- Phillips, P.C.B. (2005b) Challenges of trending time series econometrics. *Mathematics and Computers in Simulation* 68, 401–416.
- Phillips, P.C.B. (2010) Two New Zealand pioneer econometricians. *New Zealand Economic Papers* 44, 1–26.
- Phillips, P.C.B. & S. Jin (2015) Business cycles, trend elimination, and the HP filter, Cowles Foundation Discussion Paper No. 2005, Yale University.
- Phillips, P.C.B. & Z. Shi (2019) Boosting the Hodrick–Prescott filter, Cowles Foundation Discussion Paper No. 2192, Yale University.
- Sakarya, N. & R.M. de Jong (2020) A property of the Hodrick–Prescott filter and its application. *Econometric Theory*. doi:10.1017/S0266466619000331.
- Seber, G.A.F. (2008) *A Matrix Handbook for Statisticians*. Wiley, Hoboken, NJ.
- Shuman, D.I., S.K. Narang, P. Frossard, A. Ortega, & P. Vandergheynst (2013) The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine* 30, 83–98.
- Strang, G. (1999) The discrete cosine transform. *SIAM Review* 41, 135–147.
- Strang, G. & S. MacNamara (2014) Functions of difference matrices are Toeplitz plus Hankel. *SIAM Review* 56, 525–546.

von Neumann, J. (1941) Distribution of the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics* 12, 367–395.

Weinert, H.L. (2007) Efficient computation for Whittaker–Henderson smoothing. *Computational Statistics and Data Analysis* 52, 959–974.

Whittaker, E.T. (1923) On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society* 41, 63–75.

Yamada, H. (2012) A note on bandpass filters based on the Hodrick–Prescott filter and the OECD system of composite leading indicators. *OECD Journal: Journal of Business Cycle Measurement and Analysis* 2011/2, 105–109.

Yamada, H. (2015) Ridge regression representations of the generalized Hodrick–Prescott filter. *Journal of the Japan Statistical Society* 45, 121–128.

Yamada, H. (2018a) Several least squares problems related to the Hodrick–Prescott filtering. *Communications in Statistics—Theory and Methods* 47, 1022–1027.

Yamada, H. (2018b) Why does the trend extracted by the Hodrick–Prescott filtering seem to be more plausible than the linear trend? *Applied Economics Letters* 25, 102–105.

Yamada, H. (2019) A note on Whittaker–Henderson graduation: Bisymmetry of the smoother matrix. *Communications in Statistics—Theory and Methods*, 1–6 doi:10.1080/03610926.2018.1563183.

Yamada, H. & F.T. Jahra (2018) Explicit formulas for the smoother weights of the Whittaker–Henderson graduation of order 1. *Communications in Statistics—Theory and Methods* 48, 3153–3161.

Yamada, H. & F.T. Jahra (2019) An explicit formula for the smoother weights of the Hodrick–Prescott filter. *Studies in Nonlinear Dynamics and Econometrics* 23, 1–10.

APPENDIX A

A.1. Proof of Propositions 2.1, 2.9, 4.3, 4.5, 4.7, and 4.18

A.1.1. Proof of Proposition 2.1.

- (i) (a): $f_{\text{mHP}}(\mathbf{x})$ can be represented as $f_{\text{mHP}}(\mathbf{x}) = \mathbf{x}'(\mathbf{I}_T + \lambda\mathbf{L}^2)\mathbf{x} - 2\mathbf{y}'\mathbf{x} + \mathbf{y}'\mathbf{y}$. Because $f_{\text{mHP}}(\mathbf{x})$ is a quadratic function whose Hessian matrix, $2(\mathbf{I}_T + \lambda\mathbf{L}^2)$, is positive definite, it has a unique global minimizer. More directly, the following inequality holds: $f_{\text{mHP}}(\mathbf{x}) - f_{\text{mHP}}(\widehat{\mathbf{x}}_{\text{mHP}}) = (\mathbf{x} - \widehat{\mathbf{x}}_{\text{mHP}})'(\mathbf{I}_T + \lambda\mathbf{L}^2)(\mathbf{x} - \widehat{\mathbf{x}}_{\text{mHP}}) > 0$ if $\mathbf{x} \neq \widehat{\mathbf{x}}_{\text{mHP}}$. (b): Given that $\widehat{\mathbf{x}}_{\text{mHP}}$ is a unique global minimizer of $f_{\text{mHP}}(\mathbf{x})$ and $\mathbf{y} \neq \widehat{\mathbf{x}}_{\text{mHP}}$, it follows that $\lambda\|\mathbf{L}\mathbf{y}\|^2 = f_{\text{mHP}}(\mathbf{y}) - f_{\text{mHP}}(\widehat{\mathbf{x}}_{\text{mHP}}) > \lambda\|\mathbf{L}\widehat{\mathbf{x}}_{\text{mHP}}\|^2$. Given $\lambda > 0$, we obtain the result.
- (ii) (a): From $\mathbf{L}\boldsymbol{\iota} = \mathbf{D}'_1\mathbf{D}_1\boldsymbol{\iota} = \mathbf{0}$, we have $(\mathbf{I}_T + \lambda\mathbf{L}^2)\boldsymbol{\iota} = \boldsymbol{\iota}$. Premultiplying it by $(\mathbf{I}_T + \lambda\mathbf{L}^2)^{-1}$ yields $(\mathbf{I}_T + \lambda\mathbf{L}^2)^{-1}\boldsymbol{\iota} = \boldsymbol{\iota}$. (b): Since $(\mathbf{I}_T + \lambda\mathbf{L}^2)^{-1}$ is obviously symmetric, we only show that $(\mathbf{I}_T + \lambda\mathbf{L}^2)^{-1}$ is centrosymmetric. Let \mathbf{J} be an exchange matrix such that $[\mathbf{e}_T, \mathbf{e}_{T-1}, \dots, \mathbf{e}_1]$, where \mathbf{e}_k for $k = 1, \dots, T$ denote the k th column of \mathbf{I}_T . From (2), it is clear that \mathbf{L} is centrosymmetric, that is, $\mathbf{L} = \mathbf{J}\mathbf{L}\mathbf{J}$. Then, given $\mathbf{J}^{-1} = \mathbf{J}$, we obtain $\mathbf{L}^2 = (\mathbf{J}\mathbf{L}\mathbf{J})^2 = \mathbf{J}\mathbf{L}^2\mathbf{J}$ and thus $(\mathbf{I}_T + \lambda\mathbf{L}^2) = \mathbf{J}(\mathbf{I}_T + \lambda\mathbf{L}^2)\mathbf{J}$. Accordingly, it follows that $\mathbf{J}(\mathbf{I}_T + \lambda\mathbf{L}^2)^{-1}\mathbf{J} = \{\mathbf{J}(\mathbf{I}_T + \lambda\mathbf{L}^2)\mathbf{J}\}^{-1} = (\mathbf{I}_T + \lambda\mathbf{L}^2)^{-1}$, which indicates that $(\mathbf{I}_T + \lambda\mathbf{L}^2)^{-1}$ is centrosymmetric. Therefore, $(\mathbf{I}_T + \lambda\mathbf{L}^2)^{-1}$ is bisymmetric (i.e., symmetric centrosymmetric).
- (iii) (a): Given $\boldsymbol{\iota}'\mathbf{L} = \mathbf{0}$, premultiplying $(\mathbf{I}_T + \lambda\mathbf{L}^2)\widehat{\mathbf{x}}_{\text{mHP}} = \mathbf{y}$ by $\boldsymbol{\iota}'$ yields $\boldsymbol{\iota}'\widehat{\mathbf{x}}_{\text{mHP}} = \boldsymbol{\iota}'(\mathbf{I}_T + \lambda\mathbf{L}^2)\widehat{\mathbf{x}}_{\text{mHP}} = \boldsymbol{\iota}'\mathbf{y}$ and from which we obtain $\boldsymbol{\iota}'(\mathbf{y} - \widehat{\mathbf{x}}_{\text{mHP}}) = 0$ and $T^{-1}\boldsymbol{\iota}'\widehat{\mathbf{x}}_{\text{mHP}} = \bar{y}$. (b): From (9), we obtain $\widehat{\mathbf{x}}_{\text{mHP}} = \bar{y}\boldsymbol{\iota} + \mathbf{U}_2(\mathbf{I}_{T-1} + \lambda\mathbf{G}_2^2)^{-1}\mathbf{U}_2'\mathbf{y}$,

where $U_2 = [u_2, \dots, u_T]$ and $G_2 = \text{diag}(g_2, \dots, g_T)$. Given $U_2' U_2 = I_{T-1}$, it follows that $\|\widehat{x}_{\text{mHP}} - \bar{y}u\|^2 = y' U_2 \{ (I_{T-1} + \lambda G_2^2)^{-1} \}^2 U_2' y$. Thus, if $y = 0$, $\|\widehat{x}_{\text{mHP}} - \bar{y}u\| = 0$. Otherwise, we have the following inequalities: $0 \leq \|\widehat{x}_{\text{mHP}} - \bar{y}u\|^2 / \|y\|^2 \leq \{ (1 + \lambda g_2^2)^{-1} \}^2$. Here, $\{ (1 + \lambda g_2^2)^{-1} \}^2 \rightarrow 0$ as $\lambda \rightarrow \infty$ and therefore $\|\widehat{x}_{\text{mHP}} - \bar{y}u\| \rightarrow 0$ as $\lambda \rightarrow \infty$. (c): From $y - \widehat{x}_{\text{mHP}} = \lambda L^2 \widehat{x}_{\text{mHP}} = \lambda L^2 (I_T + \lambda L^2)^{-1} y$, we obtain $\|\widehat{x}_{\text{mHP}} - y\|^2 = \lambda^2 y' (I_T + \lambda L^2)^{-1} L^4 (I_T + \lambda L^2)^{-1} y = \lambda^2 y' U (I_T + \lambda G^2)^{-1} G^4 (I_T + \lambda G^2)^{-1} U' y$. Thus, if $y = 0$, $\|\widehat{x}_{\text{mHP}} - y\| = 0$. Otherwise, we have the following inequalities: $0 \leq \|\widehat{x}_{\text{mHP}} - y\|^2 / \|y\|^2 \leq \lambda^2 g_T^4 / (1 + \lambda g_T^2)^2$. Given $\lambda^2 g_T^4 / (1 + \lambda g_T^2)^2 \rightarrow 0$ as $\lambda \rightarrow 0$, it thus follows that $\|\widehat{x}_{\text{mHP}} - y\| \rightarrow 0$ as $\lambda \rightarrow 0$. (d): It is evident from $(I_T + \lambda L^2)^{-1} (ku) = (ku)$ for any $k \in \mathbb{R}$.

A.1.2. Proof of Proposition 2.9.

- (i) From $L'L = D_2' D_2 + f_1' f_1 + f_T' f_T = D_2' D_2 + F F'$, it follows that $\widehat{x}_{\text{mHP}} = (I_T + \lambda L'L)^{-1} y = (I_T + \lambda D_2' D_2 + \lambda F F')^{-1} y = (A + \lambda F F')^{-1} y$. Applying the Sherman–Morrison–Woodbury formula (Seber, 2008) to $(A + \lambda F F')^{-1}$ yields $\widehat{x}_{\text{mHP}} = (A + \lambda F F')^{-1} y = A^{-1} y - A^{-1} F (\lambda^{-1} I_2 + F' A^{-1} F)^{-1} F' A^{-1} y = \widehat{x}_{\text{HP}} - A^{-1} F (\lambda^{-1} I_2 + F' A^{-1} F)^{-1} F' \widehat{x}_{\text{HP}}$.
- (ii) Let $\Theta = A^{-1} F (\lambda^{-1} I_2 + F' A^{-1} F)^{-1} \in \mathbb{R}^{T \times 2}$. Then, from $\widehat{x}_{\text{HP}} - \widehat{x}_{\text{mHP}} = A^{-1} F (\lambda^{-1} I_2 + F' A^{-1} F)^{-1} F' \widehat{x}_{\text{HP}} = \Theta F' \widehat{x}_{\text{HP}}$, we obtain $\|\widehat{x}_{\text{HP}} - \widehat{x}_{\text{mHP}}\|^2 = (F' \widehat{x}_{\text{HP}})' \Theta' \Theta (F' \widehat{x}_{\text{HP}})$. F has full column rank and thus Θ also has full column rank, which indicates that $\Theta' \Theta$ is positive definite. Hence, $\widehat{x}_{\text{HP}} = \widehat{x}_{\text{mHP}}$ if and only if $F' \widehat{x}_{\text{HP}} = [\hat{x}_1 - \hat{x}_2, \hat{x}_T - \hat{x}_{T-1}]' = 0$.

A.1.3. Proof of Proposition 4.3.

- (i) From (9), it follows that $\widehat{x}_{\text{mHP}} = \frac{u_1 u_1' y}{1 + \lambda g_1^2} + \dots + \frac{u_T u_T' y}{1 + \lambda g_T^2} = \frac{c_1}{1 + \lambda g_1^2} u_1 + \dots + \frac{c_T}{1 + \lambda g_T^2} u_T = \widehat{c}_{\text{mHP},1} u_1 + \dots + \widehat{c}_{\text{mHP},T} u_T = \widehat{\xi}_{\text{mHP},1} + \dots + \widehat{\xi}_{\text{mHP},T}$. Then, given that $u_k' u_l = 0$ if $k \neq l$ and $u_k' u_l = 1$ if $k = l$, we obtain that $\widehat{\xi}_{\text{mHP},k} \widehat{\xi}_{\text{mHP},l} = \widehat{c}_{\text{mHP},k} \widehat{c}_{\text{mHP},l} u_k' u_l = 0$ if $k \neq l$ and $\widehat{\xi}_{\text{mHP},k} \widehat{\xi}_{\text{mHP},l} = \widehat{c}_{\text{mHP},k} \widehat{c}_{\text{mHP},l} u_k' u_l = \widehat{c}_{\text{mHP},k}^2$ if $k = l$.
- (ii) (a): Given that $g_1 = 0$ and $u_1 = T^{-\frac{1}{2}} u$, it follows that $\widehat{\xi}_{\text{mHP},1} = \frac{c_1}{1 + \lambda g_1^2} u_1 = u_1 u_1' y = T^{-1} u' y = \bar{y}u$, which indicates that $\widehat{\xi}_{\text{mHP},1} \in \mathcal{S}(u)$. In addition, from $\widehat{\xi}_{\text{mHP},1} = \bar{y}u$, we obtain $T^{-1} u' \widehat{\xi}_{\text{mHP},1} = \bar{y} T^{-1} u' u = \bar{y}$. (b): For $k = 2, \dots, T$, given $u_1' u_k = 0$, we obtain $u' \widehat{\xi}_{\text{mHP},k} = T^{\frac{1}{2}} u_1' \frac{c_k}{1 + \lambda g_k^2} u_k = 0$, which indicates that $\widehat{\xi}_{\text{mHP},k} \in \mathcal{S}^\perp(u)$ and $T^{-1} u' \widehat{\xi}_{\text{mHP},k} = 0$.
- (iii) Given that $\widehat{\xi}_{\text{mHP},1} \in \mathcal{S}(u)$ and $\widehat{\xi}_{\text{mHP},k} \in \mathcal{S}^\perp(u)$ for $k = 2, \dots, T$, it follows that $Q_1 \widehat{\xi}_{\text{mHP},1} = 0$ and $Q_i \widehat{\xi}_{\text{mHP},k} = \widehat{\xi}_{\text{mHP},k}$ for $k = 2, \dots, T$. Thus, we obtain $Q_i \widehat{x}_{\text{mHP}} = Q_i (\widehat{\xi}_{\text{mHP},1} + \widehat{\xi}_{\text{mHP},2} + \dots + \widehat{\xi}_{\text{mHP},T}) = \widehat{\xi}_{\text{mHP},2} + \dots + \widehat{\xi}_{\text{mHP},T}$, which leads to the following result: $\|Q_i \widehat{x}_{\text{mHP}}\|^2 = \|\widehat{\xi}_{\text{mHP},2}\|^2 + \dots + \|\widehat{\xi}_{\text{mHP},T}\|^2 = \widehat{c}_{\text{mHP},2}^2 + \dots + \widehat{c}_{\text{mHP},T}^2$.
- (iv) By denoting the i th row of U by $[u_{i,1}, \dots, u_{i,T}]$ for $i = 1, \dots, T$, it follows that $\widehat{h}_{\text{mHP},i,j} = (1 + \lambda g_1^2)^{-1} u_{i,1} u_{j,1} + \dots + (1 + \lambda g_T^2)^{-1} u_{i,T} u_{j,T} = u_{i,1} u_{j,1} +$

$\sum_{k=2}^T \frac{u_{i,k}u_{j,k}}{1+\lambda g_k^2}$, where the second equality follows from $g_1 = 0$. From (6), because $u_{i,1}u_{j,1} = \frac{1}{T}$ and $u_{i,k}u_{j,k} = \frac{2}{T} \cos\{(k-1)\theta_i\} \cos\{(k-1)\theta_j\}$ for $k = 2, \dots, T$, we obtain (11). Recalling $\theta_i = \frac{(i-\frac{1}{2})\pi}{T}$ for $i = 1, \dots, T$ and $g_k = 2 - 2\cos\left\{\frac{(k-1)\pi}{T}\right\}$ for $k = 1, \dots, T$, $\hat{h}_{\text{mHP},i,j}$ can be rewritten as

$$\begin{aligned} \hat{h}_{\text{mHP},i,j} &= \frac{1}{T} + \frac{2}{T} \sum_{k=2}^T \frac{\cos\left\{(k-1)\frac{(i-\frac{1}{2})\pi}{T}\right\} \cos\left\{(k-1)\frac{(j-\frac{1}{2})\pi}{T}\right\}}{1 + \lambda \left[2 - 2\cos\left\{\frac{(k-1)\pi}{T}\right\}\right]^2} \\ &= \frac{1}{T} + 2 \left(\frac{1}{T} \sum_{l=1}^{T-1} \frac{\cos\left\{\frac{l}{T}(i-\frac{1}{2})\pi\right\} \cos\left\{\frac{l}{T}(j-\frac{1}{2})\pi\right\}}{1 + \lambda \left[2 - 2\cos\left(\frac{l}{T}\pi\right)\right]^2} \right), \end{aligned}$$

where $l = k - 1$, from which we obtain (12).

A.1.4. Proof of Proposition 4.5.

Given that $L^2 = UG^2U'$ and U is an orthogonal matrix, it follows that $\hat{x}_{\text{mHP}} = (I_T + \lambda UG^2U')^{-1}y = U(I_T + \lambda G^2)^{-1}U'y = U\hat{\phi}_{\text{mHP}}$, which leads to $\hat{x}'_{\text{mHP}}L^2\hat{x}_{\text{mHP}} = \hat{x}'_{\text{mHP}}UG^2U'\hat{x}_{\text{mHP}} = \hat{\phi}'_{\text{mHP}}G^2\hat{\phi}_{\text{mHP}}$. Consequently, we obtain $\|y - \hat{x}_{\text{mHP}}\|^2 + \lambda \hat{x}'_{\text{mHP}}L^2\hat{x}_{\text{mHP}} = \|y - U\hat{\phi}_{\text{mHP}}\|^2 + \lambda \hat{\phi}'_{\text{mHP}}G^2\hat{\phi}_{\text{mHP}}$. Likewise, by letting ϕ be such that $x = U\phi$, it follows that $x'L^2x = x'UG^2U'x = \phi'G^2\phi$. Thus, we obtain $\|y - x\|^2 + \lambda x'L^2x = \|y - U\phi\|^2 + \lambda \phi'G^2\phi$. Because U is nonsingular, $x = \hat{x}_{\text{mHP}}$ if and only if $\phi = \hat{\phi}_{\text{mHP}}$. Then, combining these results yields the following inequality: $\|y - U\phi\|^2 + \lambda \phi'G^2\phi = \|y - x\|^2 + \lambda x'L^2x > \|y - \hat{x}_{\text{mHP}}\|^2 + \lambda \hat{x}'_{\text{mHP}}L^2\hat{x}_{\text{mHP}} = \|y - U\hat{\phi}_{\text{mHP}}\|^2 + \lambda \hat{\phi}'_{\text{mHP}}G^2\hat{\phi}_{\text{mHP}}$ if $\phi \neq \hat{\phi}_{\text{mHP}}$.

A.1.5. Proof of Proposition 4.7.

Given that $g_1 = 0$, it follows that $UG^2U' = \sum_{k=2}^T g_k^2 u_k u_k' = E'E$, which leads to $\hat{x}_{\text{mHP}} = (I_T + \lambda UG^2U')^{-1}y = (I_T + \lambda E'E)^{-1}y = \arg \min_{x \in \mathbb{R}^T} \|y - x\|^2 + \lambda \|Ex\|^2$.

A.1.6 Proof of Proposition 4.18.

- (i) Let $r_k = \frac{g_k}{g_{q+1}}$ for $k = 1, \dots, T$. Then, it follows that $w_{\text{mHP},k} = (1 + \lambda g_k^2)^{-1} = (1 + r_k^2)^{-1}$ and $w_{\text{ES},k} = (1 + \psi g_k)^{-1} = (1 + r_k)^{-1}$. Given the inequalities, $0 = g_1 < g_2 < \dots < g_T < 4$, it follows that if $k = 1$, then $r_k = 0$; if $k = 2, \dots, q$, then $0 < r_k < 1$; if $k = q + 1$, then $r_k = 1$; and if $k = q + 2, \dots, T$, then $r_k > 1$. Accordingly, we obtain that if $k = 1$, then $(1 + r_k^2) = (1 + r_k) = 1$; if $k = 2, \dots, q$, then $1 < (1 + r_k^2) < (1 + r_k) < 2$; if $k = q + 1$, then $(1 + r_k^2) = (1 + r_k) = 2$; and if $k = q + 2, \dots, T$, then $(1 + r_k^2) > (1 + r_k) > 2$. Inverting these (in) equalities leads to the result.
- (ii) It follows from (19) and (20).

A.2. DFT as a Type of Graph Fourier Transform

DFT is defined by

$$d_1 = \sqrt{\frac{1}{T}} \sum_{t=1}^T y_t, \quad d_k = \sqrt{\frac{1}{T}} \sum_{t=1}^T e^{-i\left(\frac{2\pi}{T}\right)(t-1)\cdot(k-1)} y_t, \quad k = 2, \dots, T,$$

where i denotes the imaginary unit. By letting $\mathbf{d} = [d_1, \dots, d_T]'$ and

$$\mathbf{W} = \sqrt{\frac{1}{T}} \begin{bmatrix} \omega^{0 \cdot 0} & \omega^{0 \cdot 1} & \dots & \omega^{0 \cdot (T-1)} \\ \omega^{1 \cdot 0} & \omega^{1 \cdot 1} & \dots & \omega^{1 \cdot (T-1)} \\ \vdots & \vdots & & \vdots \\ \omega^{(T-1) \cdot 0} & \omega^{(T-1) \cdot 1} & \dots & \omega^{(T-1) \cdot (T-1)} \end{bmatrix} \in \mathbb{C}^{T \times T}, \tag{A.1}$$

where $\omega = e^{i(\frac{2\pi}{T})}$, DFT is represented in matrix notation as $\mathbf{d} = \mathbf{W}^* \mathbf{y} \in \mathbb{C}^T$. Here, \mathbf{W}^* denotes the conjugate transpose of \mathbf{W} .

Let L_{cycle} be the following circulant graph Laplacian:

$$\mathbf{L}_{\text{cycle}} = \begin{bmatrix} 2 & -1 & 0 & \dots & 0 & -1 \\ -1 & 2 & -1 & \ddots & & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & & \ddots & -1 & 2 & -1 \\ -1 & 0 & \dots & 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} f_{\text{cycle},1} \\ -\mathbf{D}_2 \\ f_{\text{cycle},T} \end{bmatrix} \in \mathbb{R}^{T \times T}. \tag{A.2}$$

More precisely, L_{cycle} is the graph Laplacian of a cycle graph of order T (Figure 2). The vertex set of the graph equals S_1 in (4) and the edge set is defined by adding an edge $\{T, 1\}$ to S_2 : $\{\{1, 2\}, \{2, 3\}, \dots, \{T - 1, T\}, \{T, 1\}\}$. For this reason, L_{cycle} and L in (2) are the same except for the first and the last rows. Let $\Gamma = \text{diag}(\gamma_1, \dots, \gamma_T)$, where $\gamma_k = 2 - 2 \cos \left\{ (k - 1) \frac{2\pi}{T} \right\} = 4 \sin^2 \left\{ (k - 1) \frac{\pi}{T} \right\}$ for $k = 1, \dots, T$. As $L_{\text{cycle}} = \mathbf{W} \Gamma \mathbf{W}^*$ (Strang, 1999), DFT is a type of graph Fourier transform.¹⁰

¹⁰Given $\mathbf{W} \in \mathbb{C}^{T \times T}$, the definition of graph Fourier transform defined in Section 3.1 should be modified accordingly.