Econometric Theory, **38**, 2022, 419–453.
doi:10.1017/S0266466621000189

TREND EXTRACTION FROM ECONOMIC TIME SERIES WITH MISSING OBSERVATIONS BY GENERALIZED HODRICK–PRESCOTT FILTERS

Hiroshi Yamada
Hiroshima University

The Hodrick–Prescott (HP) filter has been a popular method of trend extraction from economic time series. However, it is impractical without modification if some observations are not available. This paper improves the HP filter so that it can be applied in such situations. More precisely, this paper introduces two alternative generalized HP filters that are applicable for this purpose. We provide their properties and a way of specifying those smoothing parameters that are required for their application. In addition, we numerically examine their performance. Finally, based on our analysis, we recommend one of them for applied studies.

1. INTRODUCTION

The Hodrick–Prescott (HP) (1997) filter has been a popular method of trend extraction from economic time series such as real gross domestic product and has attracted a lot of attention among econometricians.[1] Recent studies of the filter include de Jong and Sakarya (2016); Cornea-Madeira (2017); Hamilton (2018); Phillips and Jin (2020); Phillips and Shi (2020); Sakarya and de Jong (2020); Yamada (2012, 2015, 2018a, 2018b, 2020a, 2020b); Yamada and Du (2019), and Yamada and Jahra (2019).

The HP filter is defined by the following penalized least-squares problem:

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

where $y_1,\ldots,y_T$ are T observations of a univariate economic time series, $\Delta^2 x_t = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = x_t - 2x_{t-1} - x_{t-2}$ for $t = 3,\ldots,T$, and λ is a positive

The author thanks Jin Seo Cho, Kazuhiko Hayakawa, three anonymous referees, and the editor, Peter C. B. Phillips, for their valuable suggestions and comments on an earlier version of the paper. The usual caveat applies. The Japan Society for the Promotion of Science supported this work through KAKENHI Grant Numbers 16H03606 and 20H01484. Address correspondence to Hiroshi Yamada, School of Informatics and Data Science, Hiroshima University, 1-2-1 Kagamiyama, Higashi-Hiroshima 739-8525, Japan; e-mail: yamada@hiroshima-u.ac.jp.

[1]The filter was developed by Whittaker (1923) and others. See Phillips (2005, 2010) and Weinert (2007).

smoothing parameter that controls the trade-off between goodness of fit and smoothness.

In this paper, we consider the situation such that some of $y_2, \ldots, y_{T-1}$ are not available. The HP filter is clearly impractical without suitable modification in such a situation. In this paper, we improve the HP filter so that we can apply it even though some observations are not available. More precisely, we introduce two generalized HP filters, denoted by gHP$_n$ filter and gHP$_T$ filter, that are applicable for trend extraction of available observations. We provide their properties and a way of specifying their smoothing parameters that are required for their application. In addition, we numerically examine their performance. Finally, based on our analysis, we recommend the latter filter for applied studies.

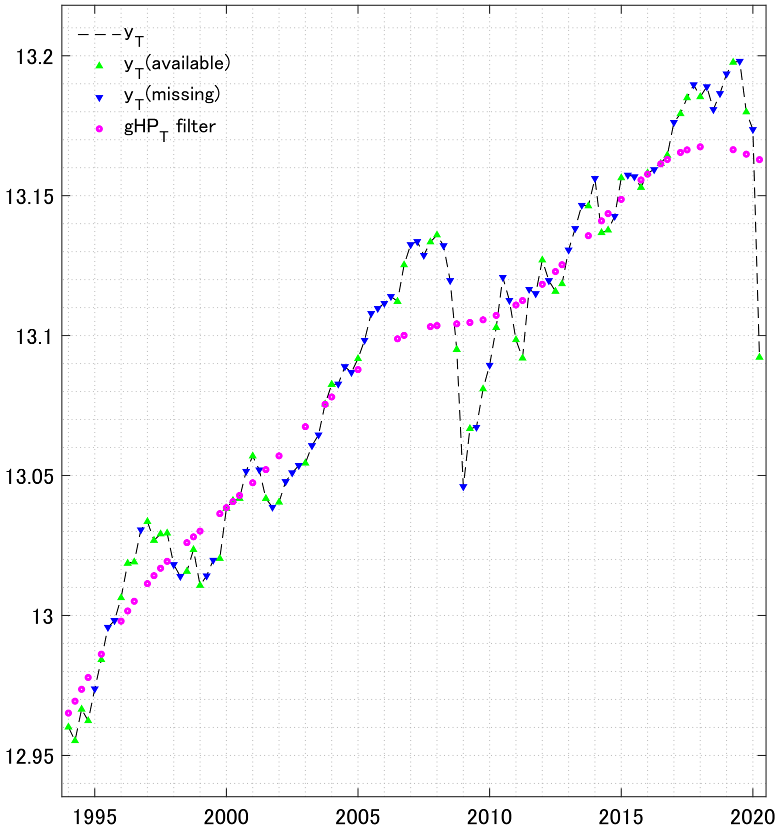

FIGURE 1. y_T denotes the log of seasonally adjusted Japanese real gross domestic product (GDP) over the sample period 1994:Q1 to 2020:Q2 (and accordingly, $T = 106$). y_T (missing) denotes $T/2 = 53$ missing observations selected randomly. gHP$_T$ filter denotes the trend estimated by the gHP$_T$ filter with $\lambda_T = 1600$.

https://doi.org/10.1017/S0266466621000189 Published online by Cambridge University Press

To illustrate the focus of this paper, see Figure 1. y_T in the figure denotes the log of seasonally adjusted Japanese real gross domestic product (GDP) over the sample period 1994:Q1 to 2020:Q2 (and accordingly, $T = 106$).[2] y_T (missing) denotes $T/2 = 53$ missing observations selected randomly from $y_2, \ldots, y_{T-1}$. gHP_T filter denotes the trend estimated by the gHP_T filter. From Figure 1, it is observable that the gHP_T filter provides plausible trend estimates for the available observations even though there are missing observations. Again, we want to emphasize that the HP filter is not applicable in such a situation. Incidentally, we superimposed the trend estimated by the HP filter onto Figure 1. We note that it is estimated from not only available observations, but also missing observations (see Figure 2). From the figure, we observe that the two estimated trend functions are very similar.

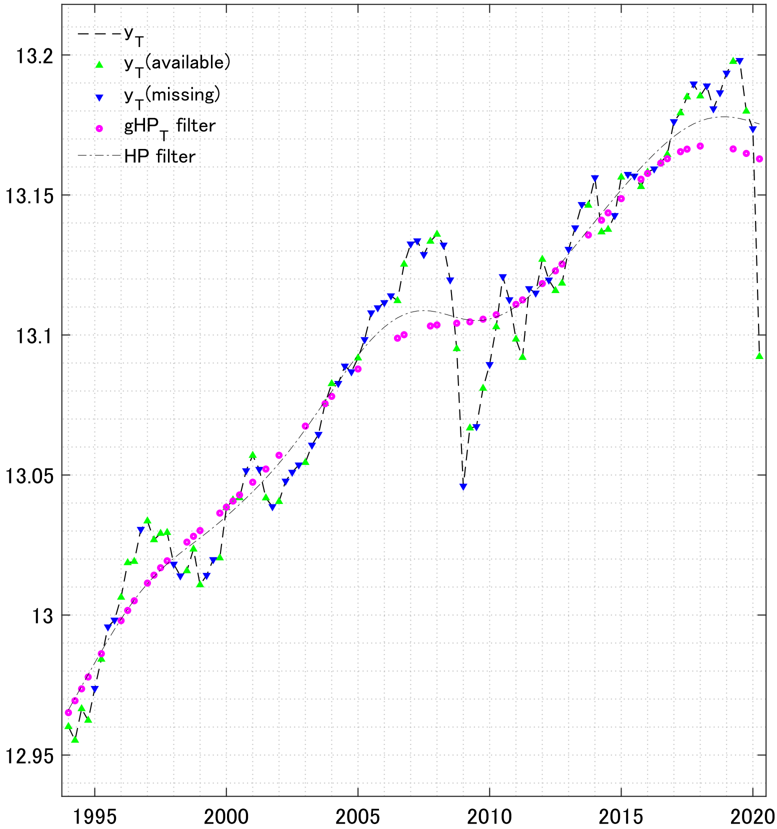

FIGURE 2. The trend estimated by the Hodrick–Prescott (HP) filter with $\lambda = 1600$, denoted by HP filter , is superimposed onto Figure 1. Note that it is estimated from not only available observations, but also missing observations.

[2]We obtained the data from the website of the Japanese Cabinet Office.

The organization of this paper is as follows. We introduce the two generalized HP filters mentioned in Section 2, following which we describe their algebraic properties in detail in Section 3. Then, in Section 4, we provide a way of specifying their smoothing parameters which are required for their application. In Section 5, we numerically examine their performance. Section 6 concludes the paper. The organization of the Appendix is as follows. In Section A.2 (resp. Sections A.3 and A.4), more details on the gHP$_n$ (resp. gHP$_T$) filter are given. In Section A.5, another minimization problem that gives the same trend estimate as the gHP$_T$ filter is presented. In Section A.6, more details on specifying the smoothing parameter for the gHP$_n$ filter are provided. In Section A.7, miscellaneous proofs are presented.

The organization of the Online Supplementary Material is as follows. In Section B.2, several Matlab user-defined functions referred to in this paper are provided. Section B.3 presents several figures referred to in Section 5 of this paper.

2. TWO GENERALIZED HP FILTERS

We denote available observations by $y_{t_1},\ldots,y_{t_n}$, where $y_{t_1}=y_1$ and $y_{t_n}=y_T$, and missing observations by $y_{s_1},\ldots,y_{s_{T-n}}$. For example, if y_3 and y_{T-1} are missing and y_2,y_4, and y_{T-2} are available, then it follows that $y_{t_1}=y_1, y_{t_2}=y_2, y_{s_1}=y_3, y_{t_3}=y_4$, $y_{t_{n-1}}=y_{T-2}$, $y_{s_{T-n}}=y_{T-1}$, and $y_{t_n}=y_T$. See Figure 3. Note that, by definition, $\{1,\ldots,T\}=\{t_1,\ldots,t_n\}\cup\{s_1,\ldots,s_{T-n}\}$ and $\{t_1,\ldots,t_n\}\cap\{s_1,\ldots,s_{T-n}\}=\emptyset$.

In the rest of the paper, we suppose that the following assumption holds:

Assumption 1. (i) y_1 and y_T are available and (ii) T and n are such that $T>n\geq 3$.

Thus, we suppose that at least one observation other than y_1 and y_T is always available and there must be at least one missing observation. We remark that a modification of the HP filter for the case where y_T is missing is proposed in Yamada and Du (2019). (We note that a similar modification for the case where y_1 is missing is possible.) The idea of it is similar to that of the gHP$_T$ filter.

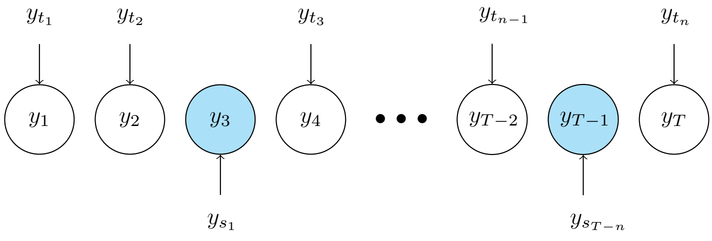

FIGURE 3. An illustration of missing observations.

https://doi.org/10.1017/S0266466621000189 Published online by Cambridge University Press

2.1. gHP$_n$ Filter

We now introduce the first of the two generalized HP filters. Given that $\sum_{t=3}^{T}(x_t - 2x_{t-1} + x_{t-2})^2$ in (1) can be represented by

$$\sum_{t=3}^{T}\left\{\frac{x_t - x_{t-1}}{t-(t-1)} - \frac{x_{t-1} - x_{t-2}}{(t-1)-(t-2)}\right\}^2,$$

it is natural to consider the following filter:

$$\min_{x_{t_1},\ldots,x_{t_n}\in\mathbb{R}} \sum_{i=1}^{n}(y_{t_i} - x_{t_i})^2 + \lambda_n \sum_{i=3}^{n}\left(\frac{x_{t_i} - x_{t_{i-1}}}{t_i - t_{i-1}} - \frac{x_{t_{i-1}} - x_{t_{i-2}}}{t_{i-1} - t_{i-2}}\right)^2, \tag{2}$$

where λ_n is a positive smoothing parameter. (2) is identical to the HP filter if $n = T$ and is thus a generalization of the HP filter. We refer to this generalized HP filter, (2), as the "gHP$_n$ filter." Under Assumption 1, the objective function has a unique global minimizer. For more details, see Section 3. Notice that the second term of the objective function of the gHP$_n$ filter is non-negative and equals 0 only when $x_{t_1},\ldots,x_{t_n}$ are on the same straight line. In addition, the first term of the objective function is also non-negative and equals 0 only when $y_{t_i} = x_{t_i}$ for $i = 1,\ldots,n$.

2.2. gHP$_T$ Filter

The other generalized HP filter that will be introduced in this paper is as follows:

$$\min_{x_1,\ldots,x_T\in\mathbb{R}} \sum_{t\in\{t_1,\ldots,t_n\}}(y_t - x_t)^2 + \lambda_T \sum_{t=3}^{T}(\Delta^2 x_t)^2, \tag{3}$$

where λ_T is a positive smoothing parameter. Again, (3) is identical to the HP filter if $n = T$ and is thus another generalization of the HP filter. We refer to this generalized HP filter, (3), as the "gHP$_T$ filter." It is noteworthy that even though $n < T$, under Assumption 1, it has a unique global minimizer. We will show this in the next section (Lemma 4) and a related discussion is given in Remark 6. Notice that the second term of the objective function of the gHP$_T$ filter is the same as that of the HP filter and thus it is non-negative and equals 0 only when $x_1,\ldots,x_T$ are on the same straight line. In addition, the first term of the objective function is nonnegative and equals 0 only when $y_{t_i} = x_{t_i}$ for $i = 1,\ldots,n$.

3. PROPERTIES OF THE TWO GHP FILTERS

As in the case of the HP filter, the objective functions of both gHP filters are quadratic. Thus, if their Hessian matrices are positive definite, each filter has a unique global minimizer. We will show that under Assumption 1, their Hessian matrices are positive definite and we present the solutions of the two gHP filters explicitly. Subsequently, we describe their properties in detail. For this purpose, let us set some notations.

3.1. Notations

Let $\boldsymbol{I}_m$ be the identity matrix of order m, $\boldsymbol{\iota}_m = [1,\ldots,1]' \in \mathbb{R}^m$, $\boldsymbol{y}_T = [y_1,\ldots,y_T]'$, $\boldsymbol{x}_T = [x_1,\ldots,x_T]'$, $\boldsymbol{\tau}_T = [1,\ldots,T]'$, $\boldsymbol{\Pi}_T = [\boldsymbol{\iota}_T, \boldsymbol{\tau}_T] \in \mathbb{R}^{T\times 2}$, $\boldsymbol{J}_m = [\boldsymbol{0}, \boldsymbol{I}_{m-2}] \in \mathbb{R}^{(m-2)\times m}$, $\boldsymbol{y}_n = [y_{t_1},\ldots,y_{t_n}]'$, $\boldsymbol{x}_n = [x_{t_1},\ldots,x_{t_n}]'$, $\boldsymbol{\tau}_n = [t_1,\ldots,t_n]'$, $\boldsymbol{\Pi}_n = [\boldsymbol{\iota}_n, \boldsymbol{\tau}_n] \in \mathbb{R}^{n\times 2}$, $\boldsymbol{H} = \mathsf{diag}(t_2-t_1,\ldots,t_n-t_{n-1}) \in \mathbb{R}^{(n-1)\times(n-1)}$, $\boldsymbol{P} = \boldsymbol{\Pi}_n(\boldsymbol{\Pi}_n'\boldsymbol{\Pi}_n)^{-1}\boldsymbol{\Pi}_n'$, $\boldsymbol{Q} = \boldsymbol{I}_n - \boldsymbol{P}$, $\widehat{\boldsymbol{\beta}} = (\boldsymbol{\Pi}_n'\boldsymbol{\Pi}_n)^{-1}\boldsymbol{\Pi}_n'\boldsymbol{y}_n$, and, for a vector $\boldsymbol{a} = [a_1,\ldots,a_m]'$, $\|\boldsymbol{a}\|^2 = \boldsymbol{a}'\boldsymbol{a} = \sum_{i=1}^m a_i^2$. In addition, let $\boldsymbol{S}_\perp \in \mathbb{R}^{(T-n)\times T}$ and $\boldsymbol{S} \in \mathbb{R}^{n\times T}$ be selection matrices such that $\boldsymbol{S}_\perp \boldsymbol{y}_T = [y_{s_1},\ldots,y_{s_{T-n}}]'$ and $\boldsymbol{S}\boldsymbol{y}_T = [y_{t_1},\ldots,y_{t_n}]'$. Furthermore, let

$$\boldsymbol{D}_{(m)} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(m-1)\times m}$$

and $\boldsymbol{D}_T$ be a matrix such that $\boldsymbol{D}_T\boldsymbol{x} = [\Delta^2 x_3,\ldots,\Delta^2 x_T]'$. Explicitly, $\boldsymbol{D}_T \in \mathbb{R}^{(T-2)\times T}$ is a tridiagonal Toeplitz matrix whose first row (resp. last row) equals $[1,-2,1,0,\ldots,0]$ (resp. $[0,\ldots,0,1,-2,1]$).

3.2. gHP$_n$ Filter

(2) can be represented in matrix notation as

$$\min_{\boldsymbol{x}_n\in\mathbb{R}^n} f_n(\boldsymbol{x}_n) = \|\boldsymbol{y}_n - \boldsymbol{x}_n\|^2 + \lambda_n\|\boldsymbol{D}_n\boldsymbol{x}_n\|^2, \tag{4}$$

where

$$\boldsymbol{D}_n\boldsymbol{x}_n = \boldsymbol{D}_{(n-1)}\boldsymbol{H}^{-1}\boldsymbol{D}_{(n)}\boldsymbol{x}_n = \begin{bmatrix} \frac{x_{t_3}-x_{t_2}}{t_3-t_2} - \frac{x_{t_2}-x_{t_1}}{t_2-t_1} \\ \vdots \\ \frac{x_{t_n}-x_{t_{n-1}}}{t_n-t_{n-1}} - \frac{x_{t_{n-1}}-x_{t_{n-2}}}{t_{n-1}-t_{n-2}} \end{bmatrix}. \tag{5}$$

Explicitly, $\boldsymbol{D}_n \in \mathbb{R}^{(n-2)\times n}$ defined above is expressed as

$$\boldsymbol{D}_n = \begin{bmatrix} d_2^{-1} & -d_2^{-1}-d_3^{-1} & d_3^{-1} & 0 & \cdots & 0 \\ 0 & d_3^{-1} & -d_3^{-1}-d_4^{-1} & d_4^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & d_{n-1}^{-1} & -d_{n-1}^{-1}-d_n^{-1} & d_n^{-1} \end{bmatrix}, \tag{6}$$

where $d_i = t_i - t_{i-1}$ for $i = 2,\ldots,n$. We remark that (i) $\boldsymbol{D}_n = \boldsymbol{D}_T$ if $n = T$ and (ii) $\boldsymbol{D}_n'$ is identical to $\boldsymbol{Q}$ in Green and Silverman (1994, pp. 12–13).

As $f_n(\boldsymbol{x}_n)$ in (4) is a quadratic function whose Hessian matrix, $2(\boldsymbol{I}_n + \lambda_n\boldsymbol{D}_n'\boldsymbol{D}_n)$, is positive definite, there exists $\widehat{\boldsymbol{x}}_n$ such that

$$f_n(\boldsymbol{x}_n) > f_n(\widehat{\boldsymbol{x}}_n) \tag{7}$$

if $\boldsymbol{x}_n \neq \widehat{\boldsymbol{x}}_n$. The unique global minimizer of $f_n(\boldsymbol{x}_n)$, denoted by $\widehat{\boldsymbol{x}}_n$, is explicitly expressed as

$$\widehat{\boldsymbol{x}}_n = (\boldsymbol{I}_n + \lambda_n \boldsymbol{D}_n' \boldsymbol{D}_n)^{-1} \boldsymbol{y}_n. \tag{8}$$

We refer to $(\boldsymbol{I}_n + \lambda_n \boldsymbol{D}_n' \boldsymbol{D}_n)^{-1}$ as the smoother matrix of the gHP$_n$ filter. A Matlab user-defined function to calculate $\widehat{\boldsymbol{x}}_n$ in (8) is provided in the online supplementary material.

Including the above property, the gHP$_n$ filter has the following properties.

PROPOSITION 2. *(i) $\widehat{\boldsymbol{x}}_n$ is the unique global minimizer of $f_n(\boldsymbol{x}_n)$ and (ii) it satisfies $\|\boldsymbol{D}_n \boldsymbol{y}_n\|^2 > \|\boldsymbol{D}_n \widehat{\boldsymbol{x}}_n\|^2$ if $\boldsymbol{y}_n \neq \widehat{\boldsymbol{x}}_n$. (iii) Each row of the smoother matrix of the gHP$_n$ filter sums to unity, (iv) $\frac{1}{n}\boldsymbol{\iota}_n' \widehat{\boldsymbol{x}}_n = \frac{1}{n}\sum_{i=1}^{n} y_{t_i}$, (v) $\boldsymbol{\iota}_n'(\boldsymbol{y}_n - \widehat{\boldsymbol{x}}_n) = 0$, (vi) if $\boldsymbol{y}_n$ belongs to the column space of $\boldsymbol{\Pi}_n$, then $\widehat{\boldsymbol{x}}_n = \boldsymbol{y}_n$, and (vii) $\widehat{\boldsymbol{x}}_n = \boldsymbol{P}\boldsymbol{y}_n + (\boldsymbol{I}_n + \lambda_n \boldsymbol{D}_n' \boldsymbol{D}_n)^{-1}\left(\boldsymbol{y}_n - \boldsymbol{P}\boldsymbol{y}_n\right)$. (viii) $\lim_{\lambda_n \to \infty} \widehat{\boldsymbol{x}}_n = \boldsymbol{P}\boldsymbol{y}_n$ and (ix) $\lim_{\lambda_n \to 0} \widehat{\boldsymbol{x}}_n = \boldsymbol{y}_n$.*

Proof. (i) follows from (7). (ii) follows from (A.3). (iii) follows from (A.16). (iv) and (v) follow from (A.17). (vi) follows from (A.18). (vii) follows from (A.19). (viii) follows from (A.14). (ix) follows from (A.23). Note that these equations except for (7) are located in Section A.2 in the Appendix. ■

Remark 3. (a) Proposition 2 corresponds to Proposition 2.2 of Yamada (2020), which presents some basic properties of the HP filter. Proposition 2 shows that the gHP$_n$ filter inherits characteristics from the HP filter. (b) Proposition 2(i) is just like the property that the Ordinary Least Squares (OLS) estimation of a linear regression satisfying the full rank assumption has. (c) (8) and Proposition 2(ii) shows that the gHP$_n$ filter gives a linear smoother of $\boldsymbol{y}_n$. (d) Proposition 2(iii)–(v) is just like the properties that the OLS estimation of a linear regression with a constant term has. (e) Proposition 2(vi) implies that if all entries of $\boldsymbol{y}_n$ are on the same straight line, then $\widehat{\boldsymbol{x}}_n$ equals $\boldsymbol{y}_n$. (f) $\boldsymbol{P}\boldsymbol{y}_n$ in Proposition 2(vii)–(viii) represents the orthogonal projection of $\boldsymbol{y}_n$ onto the column space of $\boldsymbol{\Pi}_n$ and it thus represents a linear trend. (g) Given that the gHP$_n$ filter is a low-pass filter, $(\boldsymbol{I}_n + \lambda_n \boldsymbol{D}_n' \boldsymbol{D}_n)^{-1}\left(\boldsymbol{y}_n - \boldsymbol{P}\boldsymbol{y}_n\right)$ in Proposition 2(vii) represents a low-frequency component of $\left(\boldsymbol{y}_n - \boldsymbol{P}\boldsymbol{y}_n\right)$. (h) Proposition 2(viii) partly reflects that $\boldsymbol{D}_n \widehat{\boldsymbol{x}}_n$ must be zero for minimization when λ_n goes to infinity and partly reflects $\|\boldsymbol{y}_n - \boldsymbol{x}_n\|^2 \geq \|\boldsymbol{y}_n - \boldsymbol{P}\boldsymbol{y}_n\|^2$ if $\boldsymbol{x}_n$ belongs to the column space of $\boldsymbol{\Pi}_n$. (i) The result shown in Proposition 2(ix) is quite reasonable if we consider the objective function of the gHP$_n$ filter.

3.3. gHP$_T$ Filter

(3) can be represented in matrix notation as

$$\min_{\boldsymbol{x}_T \in \mathbb{R}^T} f_T(\boldsymbol{x}_T) = \|\boldsymbol{y}_n - \boldsymbol{S}\boldsymbol{x}_T\|^2 + \lambda_T \|\boldsymbol{D}_T \boldsymbol{x}_T\|^2. \tag{9}$$

Regarding $f_T(\boldsymbol{x}_T)$ in (9), we have the following result:

LEMMA 4. *Under Assumption 1, $2(\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T' \boldsymbol{D}_T)$, which is the Hessian matrix of $f_T(\boldsymbol{x}_T)$, is positive definite.*

Proof. See Section A.7.1. ■

As $f_T(\boldsymbol{x}_T)$ in (9) is a quadratic function whose Hessian matrix is positive definite, there exists $\widehat{\boldsymbol{x}}_T$ such that

$$f_T(\boldsymbol{x}_T) > f_T(\widehat{\boldsymbol{x}}_T) \tag{10}$$

if $\boldsymbol{x}_T \neq \widehat{\boldsymbol{x}}_T$. The unique global minimizer of $f_T(\boldsymbol{x}_T)$, denoted by $\widehat{\boldsymbol{x}}_T$, is explicitly expressed as

$$\widehat{\boldsymbol{x}}_T = (\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'\boldsymbol{y}_n \tag{11}$$

and accordingly we have

$$\boldsymbol{S}\widehat{\boldsymbol{x}}_T = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'\boldsymbol{y}_n, \tag{12}$$

$$\boldsymbol{S}_\perp\widehat{\boldsymbol{x}}_T = \boldsymbol{S}_\perp(\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'\boldsymbol{y}_n. \tag{13}$$

We refer to $(\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'$ and $\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'$ as the smoother matrices of the gHP$_T$ filter. A Matlab user-defined function to calculate $\widehat{\boldsymbol{x}}_T$ in (11) and $\boldsymbol{S}\widehat{\boldsymbol{x}}_T$ in (12) is provided in the Online Supplementary Material.

Remark 5. (i) It is noteworthy that the gHP$_T$ filter provides not only $\boldsymbol{S}\widehat{\boldsymbol{x}}_T \in \mathbb{R}^n$, but also $\boldsymbol{S}_\perp\widehat{\boldsymbol{x}}_T \in \mathbb{R}^{T-n}$, whereas the gHP$_n$ filter only provides $\widehat{\boldsymbol{x}}_n \in \mathbb{R}^n$. The reason why the gHP$_T$ filter can provide $\boldsymbol{S}_\perp\widehat{\boldsymbol{x}}_T$ is that its objective function includes $\sum_{t=3}^{T}(\Delta^2 x_t)^2$. Thereby, it is estimated so that $\widehat{\boldsymbol{x}}_T$ will become smooth. See also Remark 13. (ii) $\widehat{\boldsymbol{x}}_T$ in (11) can be derived from another minimization problem. See Section A.5 for details.

Remark 6. Although the gHP$_n$ filter requires $n \geq 3$, the gHP$_T$ filter can be defined even when only y_1 and y_T are available. In the extreme case, as expected, we have the following result:

PROPOSITION 7. *When only y_1 and y_T are observable, it follows that*

$$\widehat{\boldsymbol{x}}_T = \left(\frac{Ty_1 - y_T}{T-1}\right)\boldsymbol{\iota}_T + \left(\frac{y_T - y_1}{T-1}\right)\boldsymbol{\tau}_T. \tag{14}$$

Proof. See Section A.7.2. ■

We remark that (14) represents the linear interpolant between two points $(1, y_1)$ and (T, y_T).

Including (10), the gHP$_T$ filter has the following properties.

PROPOSITION 8. *(i) $\widehat{\boldsymbol{x}}_T$ is the unique global minimizer of $f_T(\boldsymbol{x}_T)$ and (ii) it satisfies $\|\boldsymbol{D}_T\boldsymbol{y}_T\|^2 > \|\boldsymbol{D}_T\widehat{\boldsymbol{x}}_T\|^2$ if $\boldsymbol{y}_T \neq \widehat{\boldsymbol{x}}_T$. (iii) Each row of the smoother matrices of the gHP$_T$ filter sums to unity, (iv) $\frac{1}{n}\boldsymbol{\iota}_n'\boldsymbol{S}\widehat{\boldsymbol{x}}_T = \frac{1}{n}\sum_{i=1}^{n} y_{t_i}$, (v) $\boldsymbol{\iota}_n'(\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T) = 0$, (vi) if $\boldsymbol{y}_n$ belongs to the column space of $\boldsymbol{\Pi}_n$, then $\boldsymbol{S}\widehat{\boldsymbol{x}}_T = \boldsymbol{y}_n$, and (vii) $\boldsymbol{S}\widehat{\boldsymbol{x}}_T = \boldsymbol{P}\boldsymbol{y}_n + \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'\left(\boldsymbol{y}_n - \boldsymbol{P}\boldsymbol{y}_n\right)$. (viii) $\lim_{\lambda_T\to\infty}\widehat{\boldsymbol{x}}_T = \boldsymbol{\Pi}_T\widehat{\boldsymbol{\beta}}$ and in particular $\lim_{\lambda_T\to\infty}\boldsymbol{S}\widehat{\boldsymbol{x}}_T = \boldsymbol{P}\boldsymbol{y}_n$. (ix) $\lim_{\lambda_T\to 0}\boldsymbol{S}\widehat{\boldsymbol{x}}_T = \boldsymbol{S}\boldsymbol{y}_T (= \boldsymbol{y}_n)$, but $\lim_{\lambda_T\to 0}\boldsymbol{S}_\perp\widehat{\boldsymbol{x}}_T$ is not necessarily equal to $\boldsymbol{S}_\perp\boldsymbol{y}_T$.*

Proof. (i) follows from (10). (ii) follows from (A.24). (iii) follows from (A.25). (iv) and (v) follow from (A.27). (vi) follows from (A.28). (vii) follows from (A.30). (viii) follows from (A.37) and (A.57). (ix) follows from (A.53) and (A.59). Note that these equations except for (10) are located in Sections A.3 and A.4. ■

Remark 9. (a) As Proposition 2 corresponds to Proposition 2.2 of Yamada (2020), Proposition 8 also corresponds to it. Proposition 8 shows that the gHP$_T$ filter also inherits characteristics from the HP filter. (b) Proposition 8(i) is just like the property that the OLS estimation of a linear regression satisfying the full rank assumption has. (c) (11) and Proposition 8(ii) shows that the gHP$_T$ filter gives a linear smoother of $\boldsymbol{y}_T$. (d) Proposition 8(iii)-(v) is just like the properties that the OLS estimation of a linear regression with a constant term has. (e) Proposition 8(vi) implies that if all entries of $\boldsymbol{y}_n$ are on the same straight line, then $\boldsymbol{S}\widehat{\boldsymbol{x}}_T$ equals $\boldsymbol{y}_n$. (f) $\boldsymbol{P}\boldsymbol{y}_n$ in Proposition 8(vii)-(viii) represents the orthogonal projection of $\boldsymbol{y}_n$ onto the column space of $\boldsymbol{\Pi}_n$ and it thus represents a linear trend. (g) Given that the gHP$_T$ filter is a low-pass filter, $\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S}+\lambda_T\boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'\left(\boldsymbol{y}_n-\boldsymbol{P}\boldsymbol{y}_n\right)$ in Proposition 8(vii) represents a low-frequency component of $\left(\boldsymbol{y}_n-\boldsymbol{P}\boldsymbol{y}_n\right)$. (h) Proposition 8(viii) partly reflects that $\boldsymbol{D}_T\widehat{\boldsymbol{x}}_T$ must be zero for minimization when λ_T goes to infinity and partly reflects $\|\boldsymbol{y}_n-\boldsymbol{S}\boldsymbol{x}_T\|^2\geq\|\boldsymbol{y}_n-\boldsymbol{P}\boldsymbol{y}_n\|^2$ if $\boldsymbol{x}_T$ belongs to the column space of $\boldsymbol{\Pi}_T$. (i) The results shown in Proposition 8(ix) is quite reasonable if we consider the objective function of the gHP$_T$ filter.

4. SPECIFYING THE VALUE OF SMOOTHING PARAMETERS

As for the HP filter, applying the two gHP filters, gHP$_n$ and gHP$_T$, requires specification of their smoothing parameters, λ_n in (2) and λ_T in (3). In this section, we discuss how to specify them.

Before the discussion, we review a way of specifying the smoothing parameter of the HP filter, λ in (1), shown in King and Rebelo (1993), Gómez (2001), and Harvey and Trimbur (2003).[3] (1) can be represented in matrix notation as

$$\min_{\boldsymbol{x}_T\in\mathbb{R}^T} f(\boldsymbol{x}_T)=\|\boldsymbol{y}_T-\boldsymbol{x}_T\|^2+\lambda\|\boldsymbol{D}_T\boldsymbol{x}_T\|^2 \tag{15}$$

and there exists $\widehat{\boldsymbol{x}}$ such that $f(\boldsymbol{x}_T)>f(\widehat{\boldsymbol{x}})$ if $\boldsymbol{x}_T\neq\widehat{\boldsymbol{x}}$. Explicitly,

$$\widehat{\boldsymbol{x}}=(\boldsymbol{I}_T+\lambda\boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{y}_T \tag{16}$$

and thus $\widehat{\boldsymbol{x}}$ satisfies

$$(\boldsymbol{I}_T+\lambda\boldsymbol{D}_T'\boldsymbol{D}_T)\widehat{\boldsymbol{x}}=\boldsymbol{y}_T. \tag{17}$$

(17) is a system of linear equations and the equations for $t=3,\ldots,T-2$ are

$$\zeta(L)\widehat{x}_t=y_t, \tag{18}$$

[3]Because of their contributions, the HP filter has been used as a *bandpass* filter. See also Pedersen (2001), Mills (2003), Iacobucci and Noullez (2005), OECD (2012), and Yamada (2012).

where $\widehat{\boldsymbol{x}} = [\widehat{x}_1, \ldots, \widehat{x}_T]'$ and

$$\zeta(L) = 1 + \lambda(L^2 - 4L + 6 - 4L^{-1} + L^{-2}) = 1 + \lambda\left\{(1-L)(1-L^{-1})\right\}^2. \quad \textbf{(19)}$$

Here, L denotes the lag operator such that $L\widehat{x}_t = \widehat{x}_{t-1}$, $L^{-1}\widehat{x}_t = \widehat{x}_{t+1}$, and $LL^{-1} = 1$. Let $W(\omega;\lambda)$ denote the gain function corresponding to the lag polynomial $1 - \zeta(L)^{-1}$. Then, given $(1-e^{-\mathrm{i}\omega})(1-e^{\mathrm{i}\omega}) = 2(1-\cos\omega) = \left\{2\sin\left(\frac{\omega}{2}\right)\right\}^2$, where i denotes the imaginary unit, it follows that

$$\begin{aligned} W(\omega;\lambda) = |1-\zeta(e^{-\mathrm{i}\omega})^{-1}| &= \frac{\lambda\{2(1-\cos\omega)\}^2}{1+\lambda\{2(1-\cos\omega)\}^2} \\ &= \frac{\lambda\left\{2\sin\left(\frac{\omega}{2}\right)\right\}^4}{1+\lambda\left\{2\sin\left(\frac{\omega}{2}\right)\right\}^4}. \end{aligned} \quad \textbf{(20)}$$

Figure 4 plots $W(\omega; 500)$, $W(\omega; 1600)$, and $W(\omega; 10^4)$, and it shows (i) how $W(\omega;\lambda)$ depends on λ and (ii) $W(\omega; 1600)$, for example, approximates the ideal high-pass filter whose cutoff frequency is 0.1538 (depicted by a thin dashed line). Note that the period that corresponds to the cutoff frequency is $\frac{2\pi}{0.1538} = 40.853$, which is about 10 years if the data frequency is quarterly.

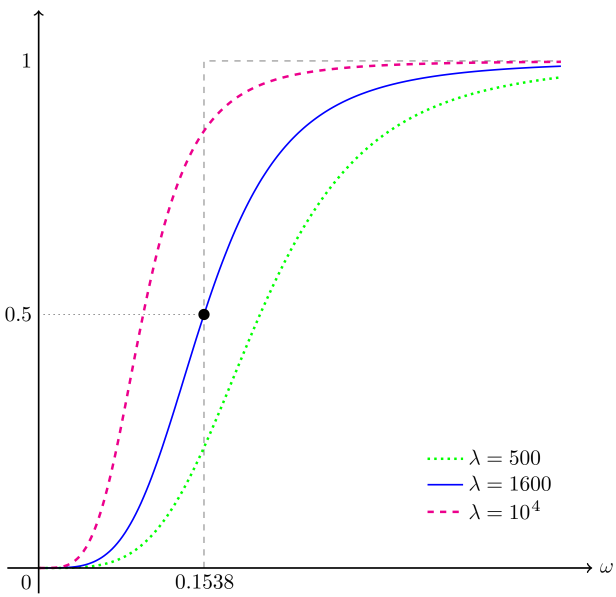

FIGURE 4. Graphs of the gain function $W(\omega;\lambda)$ in (20). The thin dashed line depicts the ideal high-pass filter whose cutoff frequency equals 0.1538.

https://doi.org/10.1017/S0266466621000189 Published online by Cambridge University Press

Thinking similarly, for a given cutoff frequency, denoted by ω_c, the corresponding value of λ, denoted by λ_c, can be specified as $W(\omega_c;\lambda_c)=\frac{1}{2}$. Explicitly, λ_c is expressed as

$$\lambda_c=\left\{2\sin\left(\frac{\omega_c}{2}\right)\right\}^{-4}. \tag{21}$$

By letting $p_c=\frac{2\pi}{\omega_c}$, which is the period corresponding to ω_c, λ_c is expressed with p_c as

$$\lambda_c=\left\{2\sin\left(\frac{\pi}{p_c}\right)\right\}^{-4}. \tag{22}$$

For example, $\lambda_c=133107.9$ (resp. $\lambda_c=13.9$) for $p_c=120$ (resp. $p_c=12$), which is used for the Organisation for Economic Co-operation and Development (OECD)'s composite leading indicators (see, OECD, 2012). We note that for monthly data, $p_c=120$ (resp. $p_c=12$) corresponds to 10 years (resp. 1 year).

Remark 10. (20) is based on an infinite sample (Phillips and Jin, 2020). As far as we know, any finite-sample correction of (21)/(22) has not been proposed.

4.1. Specifying the Value of λ_T in (3)

Given $\boldsymbol{y}_n=\boldsymbol{S}\boldsymbol{y}_T$, $\widehat{\boldsymbol{x}}_T$ in (11) satisfies

$$(\boldsymbol{S}'\boldsymbol{S}+\lambda_T\boldsymbol{D}_T'\boldsymbol{D}_T)\widehat{\boldsymbol{x}}_T=\boldsymbol{S}'\boldsymbol{S}\boldsymbol{y}_T. \tag{23}$$

Let $\widehat{\boldsymbol{x}}_T=[\widehat{x}_{T,1},\ldots,\widehat{x}_{T,T}]'$. As in the case of the HP filter, (23) is a system of linear equations, and the equations for $t=3,\ldots,T-2$, $t=t_i$, and $i=1,\ldots,n$ are

$$\zeta(L)\widehat{x}_{T,t}=y_t. \tag{24}$$

Taking the similarity between (18) and (24) into account, we consider that (21)/(22) is also valid for the gHP$_T$ filter.

4.2. Specifying λ_n in (2)

For the gHP$_n$ filter, unfortunately, we cannot use the strategy that was used for the gHP$_T$ filter, and thus, we must try an alternative strategy. Then, given $\widehat{\boldsymbol{x}}_T$, we propose a method to estimate $\widehat{\boldsymbol{x}}_n$ such that

$$\|\boldsymbol{y}_n-\widehat{\boldsymbol{x}}_n\|^2=\|\boldsymbol{y}_n-\boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2.$$

This can be accomplished by setting

$$\lambda_n=\frac{\widehat{\boldsymbol{x}}_n^{*\prime}(\boldsymbol{y}_n-\widehat{\boldsymbol{x}}_n^*)}{\|\boldsymbol{D}_n\widehat{\boldsymbol{x}}_n^*\|^2}. \tag{25}$$

https://doi.org/10.1017/S0266466621000189 Published online by Cambridge University Press

Here, $\widehat{\boldsymbol{x}}_n^*$ in (25) is the solution of the following convex problem:

$$\min_{\boldsymbol{x}_n \in \mathbb{R}^n} \|\boldsymbol{D}_n \boldsymbol{x}_n\|^2, \tag{26}$$

$$\text{s.t. } \|\boldsymbol{y}_n - \boldsymbol{x}_n\|^2 \leq \|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2, \tag{27}$$

where $\boldsymbol{y}_n$ is assumed not to belong to the column space of $\boldsymbol{\Pi}_n$.

Remark 11. (i) For more details, see Section A.6. (ii) A Matlab user-defined function to calculate $\widehat{\boldsymbol{x}}_n^*$, which requires CVX, a package for specifying and solving convex programs (CVX Research, Inc., 2011; Grant and Boyd, 2008), is provided in the Online Supplementary Material.

5. NUMERICAL EXAMINATIONS

In this section, we numerically illustrate the performance of the gHP filters, i.e., gHP_n and gHP_T. For this purpose, we generate $\boldsymbol{y}_T$ as follows:

$$\boldsymbol{y}_T = \boldsymbol{x}_T + \boldsymbol{u}, \quad \boldsymbol{D}_T \boldsymbol{x}_T = \boldsymbol{v}, \tag{28}$$

$$\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \sigma_u^2 \boldsymbol{I}_T & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_v^2 \boldsymbol{I}_{T-2} \end{bmatrix} \right), \tag{29}$$

where $\sigma_u^2 > 0$ and $\sigma_v^2 > 0$. Note that $\boldsymbol{D}_T \boldsymbol{x}_T = \boldsymbol{v}$ in (28) can be represented by

$$\Delta^2 x_t = v_t, \quad t = 3, \ldots, T, \tag{30}$$

where $\boldsymbol{v} = [v_3, \ldots, v_T]'$.

It is notable that (28) and (29) has a linear mixed model representation as follows (see, e.g., Paige and Trindade, 2010):

$$\boldsymbol{y}_T = \boldsymbol{A}_T \boldsymbol{\alpha}_T + \boldsymbol{u} = \boldsymbol{\Pi}_T \boldsymbol{\beta} + \boldsymbol{U}_T \boldsymbol{\gamma}_T + \boldsymbol{u}, \tag{31}$$

$$\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\gamma}_T \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \sigma_u^2 \boldsymbol{I}_T & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_v^2 \boldsymbol{I}_{T-2} \end{bmatrix} \right), \tag{32}$$

where $\boldsymbol{A}_T \in \mathbb{R}^{T \times T}$ and $\boldsymbol{U}_T \in \mathbb{R}^{T \times (T-2)}$ are matrices such that

$$\boldsymbol{A}_T = \left[\begin{array}{cc:cccc} 1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 2 & 0 & \cdots & 0 & 0 \\ 1 & 3 & 1 & \ddots & \vdots & \vdots \\ 1 & 4 & 2 & \ddots & 0 & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 & 0 \\ 1 & T & T-2 & \cdots & 2 & 1 \end{array}\right] = \left[\begin{array}{c:c} \boldsymbol{\Pi}_T & \boldsymbol{U}_T \end{array}\right]. \tag{33}$$

Remark 12. Let $\lambda = \frac{\sigma_u^2}{\sigma_v^2}$ and recall that $\boldsymbol{J}_T = [\boldsymbol{0}, \boldsymbol{I}_{T-2}] \in \mathbb{R}^{(T-2)\times T}$. It is known that

$$\begin{aligned} \widehat{\boldsymbol{\alpha}} &= \arg \min_{\boldsymbol{\alpha}_T \in \mathbb{R}^T} \|\boldsymbol{y}_T - \boldsymbol{A}_T \boldsymbol{\alpha}_T\|^2 + \lambda \|\boldsymbol{J}_T \boldsymbol{\alpha}_T\|^2 \\ &= (\boldsymbol{A}_T' \boldsymbol{A}_T + \lambda \boldsymbol{J}_T' \boldsymbol{J}_T)^{-1} \boldsymbol{A}_T' \boldsymbol{y}_T \end{aligned} \tag{34}$$

is the best linear unbiased predictor of $\boldsymbol{\alpha}_T$ in (31). See, e.g., Robinson (1991). Likewise, $\widehat{\boldsymbol{x}} = (\boldsymbol{I}_T + \lambda \boldsymbol{D}_T' \boldsymbol{D}_T)^{-1} \boldsymbol{y}_T$, which is the solution of (15), is

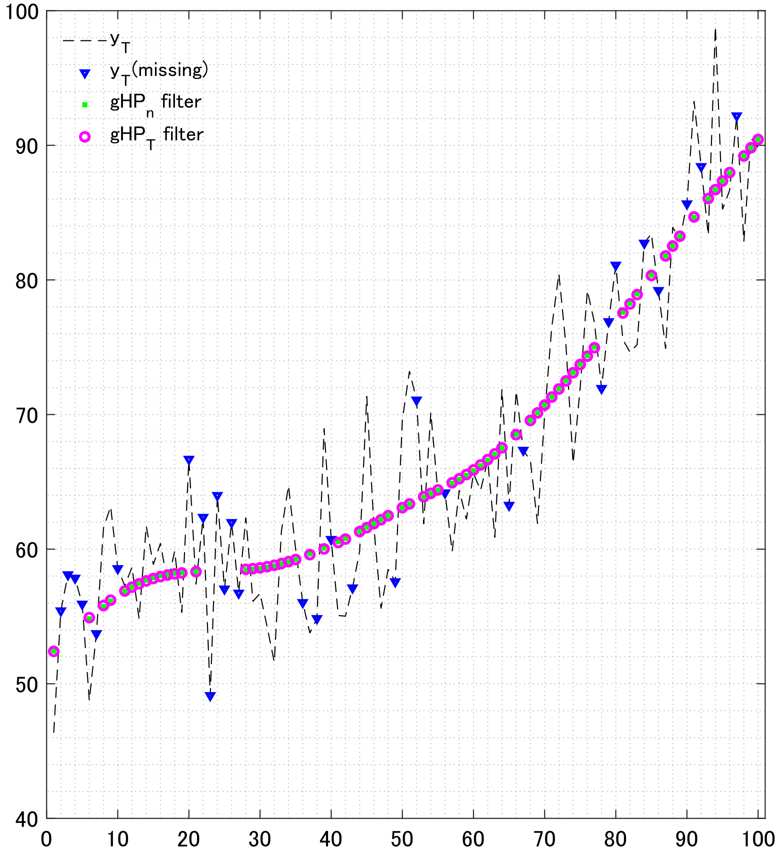

FIGURE 5. $\mathrm{y_T}$ denotes data generated by (31) and (32) by setting $T = 100$, $\boldsymbol{\beta} = [50, 0.4]'$, $\sigma_u = 5$, and $\sigma_v = \frac{1}{8}$. Accordingly, $\lambda = \frac{\sigma_u^2}{\sigma_v^2} = 1600$. Then, we selected the missing observations randomly from $\{y_2, \ldots, y_{T-1}\}$ by setting $n = 70$. Accordingly, there are $30 (= 100 - 70)$ missing observations and $\lambda = \frac{\sigma_u^2}{\sigma_v^2} = 1600$. $\mathrm{y_T}$ (missing) denotes 30 missing observations selected randomly from $\{y_2, \ldots, y_{T-1}\}$. $\mathrm{gHP_T}$ filter denotes $\boldsymbol{S}\widehat{\boldsymbol{x}}_T$ in (12) estimated with $\lambda_T = 1600$. $\mathrm{gHP_T}$ filter denotes $\widehat{\boldsymbol{x}}_n$ in (8) estimated with $\lambda_n = 1282.29$, which is specified so that $\|\boldsymbol{y}_n - \widehat{\boldsymbol{x}}_n\|^2 = \|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2$.

https://doi.org/10.1017/S0266466621000189 Published online by Cambridge University Press

the BLUP of $\boldsymbol{x}_T$ in (28). This can be shown from $\widehat{\boldsymbol{x}} - \boldsymbol{x}_T = \boldsymbol{A}_T(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_T)$ and $|\boldsymbol{A}_T| = 1 \neq 0$.

We generated $\boldsymbol{y}_T$ from (31) and (32) by setting $T = 100$, $\boldsymbol{\beta} = [50, 0.4]'$, $\sigma_u = 5$, and $\sigma_v = \frac{1}{8}$. Accordingly, $\lambda = \frac{\sigma_u^2}{\sigma_v^2} = 1600$. Then, we selected the missing observations randomly from $\{y_2, \ldots, y_{T-1}\}$ by setting $n = 90, 70, 50, 30$. In the following, we focus our attention on the case where $n = 70$ and show the corresponding figures. However, the same qualitative results are observable for the other cases. The corresponding 15 figures are provided in the Online Supplementary Material (Figures B.1–B.15).

In Figure 5, $\mathrm{y_T}$ denotes the generated data and $\mathrm{y_T}$ (missing) denotes $30 (= 100 - 70)$ missing observations. In the figure, $\mathrm{gHP_T}$ filter denotes $\boldsymbol{S}\widehat{\boldsymbol{x}}_T$ in (12) estimated with $\lambda_T = 1600$. $\mathrm{gHP_T}$ filter denotes $\widehat{\boldsymbol{x}}_n$ in (8) estimated with $\lambda_n = 1282.29$, which is specified so that $\|\boldsymbol{y}_n - \widehat{\boldsymbol{x}}_n\|^2 = \|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2$. Recall the discussions in the previous

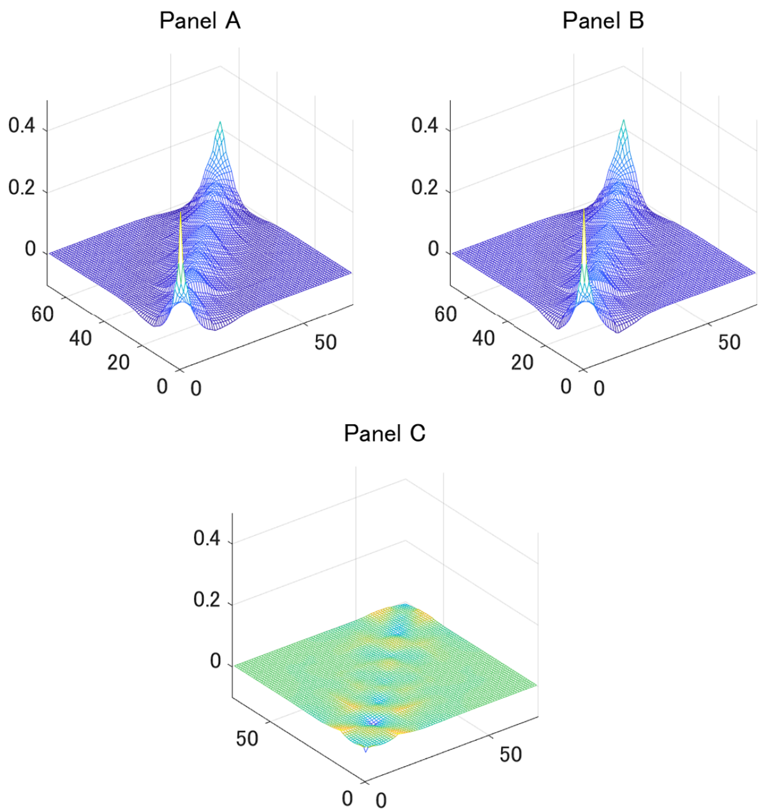

FIGURE 6. Panel A (resp. Panel B) plots the smoother matrix corresponding to $\widehat{\boldsymbol{x}}_n$ (resp. $\boldsymbol{S}\widehat{\boldsymbol{x}}_T$) in Figure 5. Panel C plots their difference.

https://doi.org/10.1017/S0266466621000189 Published online by Cambridge University Press

section. Interestingly, from the figure, we observe that $\widehat{\boldsymbol{x}}_n$ and $\boldsymbol{S}\widehat{\boldsymbol{x}}_T$ are almost the same.

The above observation motivates us to plot the corresponding smoother matrices. Panel A (resp. Panel B) in Figure 6 plots the smoother matrix corresponding to $\widehat{\boldsymbol{x}}_n$ (resp. $\boldsymbol{S}\widehat{\boldsymbol{x}}_T$) in Figure 5. Panel C in the figure plots their difference. From Panel C, we notice that the smoother matrices are almost identical. Therefore, we may consider that it leads to the above observation. Here, we remark that given Proposition 2(iii) and Proposition 8(iii), we have

$$\left\{(\boldsymbol{I}_n+\lambda_n\boldsymbol{D}_n'\boldsymbol{D}_n)^{-1}-\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S}+\lambda_T\boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'\right\}\boldsymbol{\iota}_n=\boldsymbol{\iota}_n-\boldsymbol{\iota}_n=\boldsymbol{0},$$

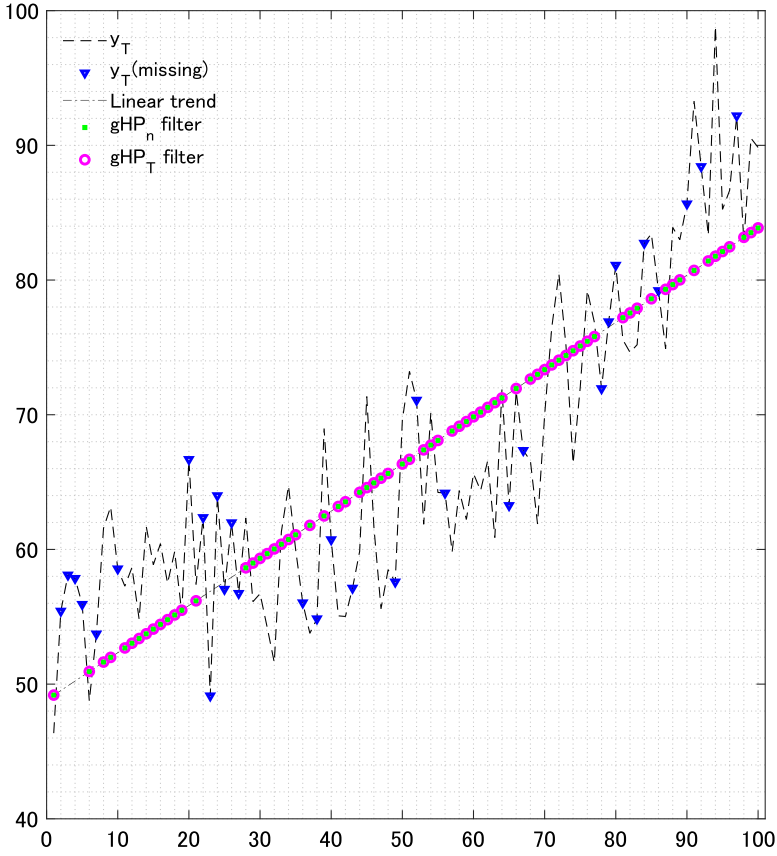

FIGURE 7. For the explanation of $\mathrm{y_T}$ and $\mathrm{y_T}$ (missing) , see Figure 5. $\mathrm{gHP_n}$ filter denotes $\widehat{\boldsymbol{x}}_n$ in (8) estimated with $\lambda_n = 10^8$ and $\mathrm{gHP_T}$ filter denotes $\boldsymbol{S}\widehat{\boldsymbol{x}}_T$ in (12) estimated with $\lambda_T = 10^8$. Linear trend denotes $\boldsymbol{P}\boldsymbol{y}_n[=\boldsymbol{\Pi}_n(\boldsymbol{\Pi}_n'\boldsymbol{\Pi}_n)^{-1}\boldsymbol{\Pi}_n'\boldsymbol{y}_n]$.

https://doi.org/10.1017/S0266466621000189 Published online by Cambridge University Press

which implies that even though there are some discrepancies between the smoother matrices, the sum of each row of their difference equals 0.

In Figure 7, $\mathrm{gHP_n}$ filter denotes $\widehat{\boldsymbol{x}}_n$ in (8) estimated with $\lambda_n = 10^8$ and $\mathrm{gHP_T}$ filter denotes $\boldsymbol{S}\widehat{\boldsymbol{x}}_T$ in (12) estimated with $\lambda_T = 10^8$. Linear trend denotes $\boldsymbol{P}\boldsymbol{y}_n[= \boldsymbol{\Pi}_n(\boldsymbol{\Pi}_n'\boldsymbol{\Pi}_n)^{-1}\boldsymbol{\Pi}_n'\boldsymbol{y}_n]$. From Propositions 2(viii) and 8(viii), we have the following theoretical results:

$$\lim_{\lambda_n\to\infty} \widehat{\boldsymbol{x}}_n = \lim_{\lambda_T\to\infty} \boldsymbol{S}\widehat{\boldsymbol{x}}_T = \boldsymbol{P}\boldsymbol{y}_n.$$

From the figure, these theoretical results are certainly confirmed.

In Figure 8, $\mathrm{gHP_n}$ filter denotes $\widehat{\boldsymbol{x}}_n$ in (8) estimated with $\lambda_n = 10^{-4}$ and $\mathrm{gHP_T}$ filter denotes $\boldsymbol{S}\widehat{\boldsymbol{x}}_T$ in (12) estimated with $\lambda_T = 10^{-4}$. Again, from Propositions 2(ix)

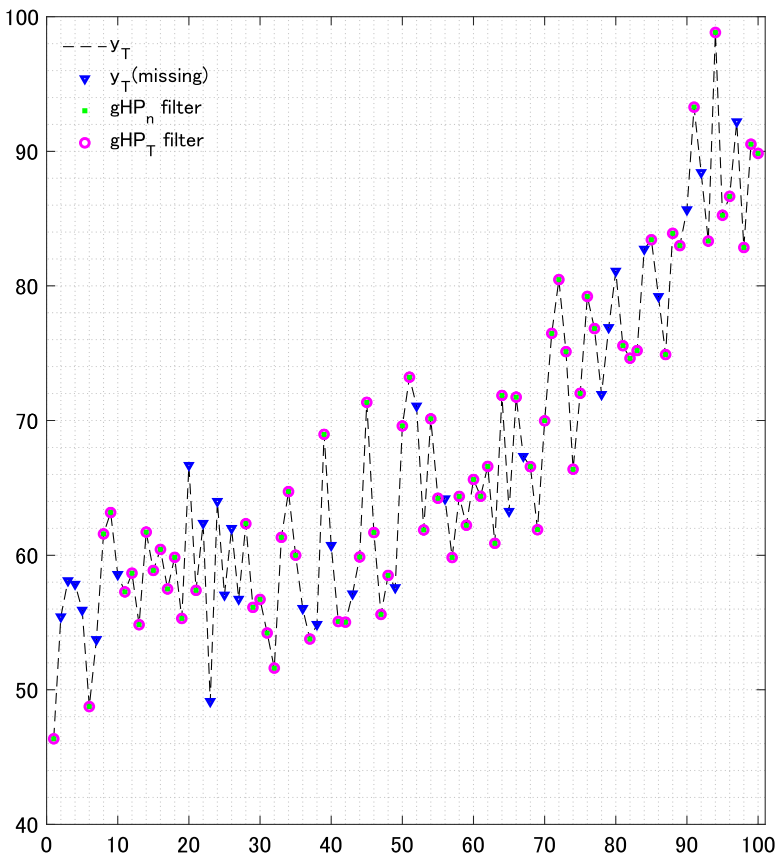

FIGURE 8. For the explanation of $\mathrm{y_T}$ and $\mathrm{y_T}$ (missing) , see Figure 5. $\mathrm{gHP_n}$ filter denotes $\widehat{\boldsymbol{x}}_n$ in (8) estimated with $\lambda_n = 10^{-4}$ and $\mathrm{gHP_T}$ filter denotes $\boldsymbol{S}\widehat{\boldsymbol{x}}_T$ in (12) estimated with $\lambda_T = 10^{-4}$.

https://doi.org/10.1017/S0266466621000189 Published online by Cambridge University Press

and 8(ix), we have the following theoretical results:

$$\lim_{\lambda_n \to 0} \widehat{\boldsymbol{x}}_n = \lim_{\lambda_T \to 0} \boldsymbol{S}\widehat{\boldsymbol{x}}_T = \boldsymbol{y}_n.$$

The figure shows that the theoretically expected results are observable.

In Figure 9, HP filter denotes $\widehat{\boldsymbol{x}}$ in (16), which is estimated with $\lambda = 1600$ from not only available observations, but also missing observations. $\mathrm{gHP_T}$ filter denotes $\boldsymbol{S}\widehat{\boldsymbol{x}}_T$ in (12) estimated with $\lambda_T = 1600$, which is identical to $\mathrm{gHP_T}$ filter shown in Figure 5. $\mathrm{gHP_T}$ filter(missing) denotes $\boldsymbol{S}_\perp\widehat{\boldsymbol{x}}_T$ in (13) estimated with $\lambda_T = 1600$. From the figure, it is observable that $\boldsymbol{S}_\perp\widehat{\boldsymbol{x}}_T$ is estimated so that $\widehat{\boldsymbol{x}}_T$ is smooth.

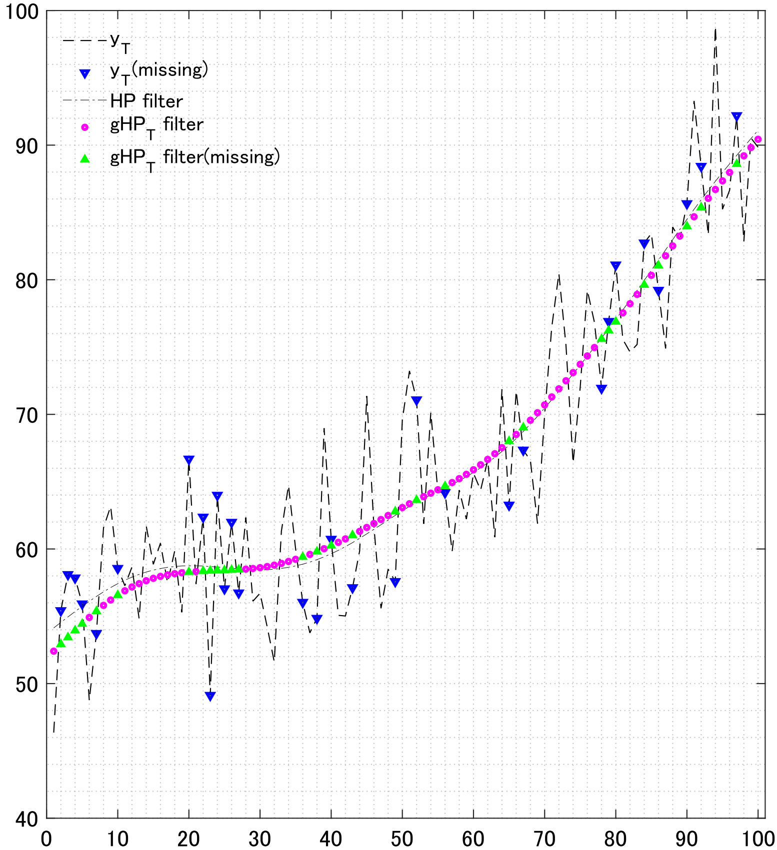

FIGURE 9. For the explanation of $\mathrm{y_T}$ and $\mathrm{y_T}$ (missing) , see Figure 5. Hodrick–Prescott (HP) filter denotes $\widehat{\boldsymbol{x}}$ in (16), which is estimated with $\lambda = 1600$ from not only available observations but also missing observations. $\mathrm{gHP_T}$ filter denotes $\boldsymbol{S}\widehat{\boldsymbol{x}}_T$ in (12) estimated with $\lambda_T = 1600$ and $\mathrm{gHP_T}$ filter(missing) denotes $\boldsymbol{S}_\perp\widehat{\boldsymbol{x}}_T$ in (13) estimated with $\lambda_T = 1600$.

TABLE 1. Simulation results on the two gHP filters

		n/T			
		0.9	0.7	0.5	0.3
$T = 100$	gHP_n filter	0.2803	0.5411	0.8129	1.1988
	gHP_T filter	0.2787	0.5373	0.8071	1.1880
$T = 200$	gHP_n filter	0.2743	0.5350	0.8003	1.1644
	gHP_T filter	0.2726	0.5312	0.7927	1.1485
$T = 400$	gHP_n filter	0.2755	0.5286	0.7808	1.1590
	gHP_T filter	0.2730	0.5238	0.7719	1.1423
$T = 800$	gHP_n filter	0.2711	0.5262	0.7865	1.1471
	gHP_T filter	0.2687	0.5209	0.7770	1.1302

Note: The root-mean-square deviations defined by (36) are tabulated. Each value in the table is calculated by generating 1000 $\boldsymbol{y}_T$'s by (31) and (32).

These results are expected to some extent. This is because $\|\boldsymbol{D}_T\boldsymbol{x}_T\|^2$, which equals $\sum_{t=3}^{T}(\Delta^2 x_t)^2$, is a part of $f_T(\boldsymbol{x}_T)$ in (9).

Remark 13. The entries of $\boldsymbol{S}_\perp\widehat{\boldsymbol{x}}_T$ are not naive linear interpolations. We give an example. Consider the case where y_3 is missing among $y_1,\ldots,y_T$. Let $\widehat{\boldsymbol{x}}_T = [\widehat{x}_{T,1},\ldots,\widehat{x}_{T,T}]'$. Then, it follows that

$$\widehat{x}_{T,3} = \frac{-\widehat{x}_{T,1} + 4\widehat{x}_{T,2} + 4\widehat{x}_{T,4} - \widehat{x}_{T,5}}{6} \neq \frac{\widehat{x}_{T,2} + \widehat{x}_{T,4}}{2}. \tag{35}$$

A proof of (35) is given in Section A.7.4.

In addition, from Figure 9, it is also observable that the gHP_T filter captures the HP filter well even though 30% of the observations are missing. Regarding this point, however, it is noteworthy that the deviations between the two filters increase as n/T decreases. Compare Figure 9 with Figures B.5, B.10, and B.15 in the Online Supplementary Material.

As stated, when $0 < \lambda = 1600 < \infty$, it is observable that (i) $\widehat{\boldsymbol{x}}_n$ and $\boldsymbol{S}\widehat{\boldsymbol{x}}_T$ are almost the same and (ii) the deviations between the gHP_T filter and the HP filter increase as n/T decreases. To check the robustness of these two results, we conducted additional examinations by generating 1000$\boldsymbol{y}_T$'s by (31) and (32) with the same parameter setting as before other than T and n. T and n used for the experiments are such that $n/T = 0.9, 0.7, 0.5, 0.3$ with $T = 100, 200, 400, 800$. Then, for each $\mathrm{y_T}$, we calculated the following root-mean-square deviation:

$$\sqrt{\frac{1}{n}\|\boldsymbol{S}\widehat{\boldsymbol{x}} - z\|^2}, \quad z = \widehat{\boldsymbol{x}}_n, \boldsymbol{S}\widehat{\boldsymbol{x}}_T, \tag{36}$$

where $\lambda = \lambda_T = 1600$ and λ_n is specified so that $\|\boldsymbol{y}_n - \widehat{\boldsymbol{x}}_n\|^2 = \|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2$. Table 1 tabulates the results. From the table, it is clearly observable that the two results above are not specific for $\mathrm{y_T}$ shown, e.g., in Figure 5.

6. CONCLUDING REMARKS

Even though the HP filter has been a popular method of trend extraction from economic time series, it is impractical without suitable modification if some observations are missing. In this paper, we introduced two generalized HP filters that are applicable in such situations. We provided their algebraic properties in detail and a way of specifying the smoothing parameters that are required for applications. In addition, we conducted numerical experiments.

Based on our analysis, among the two alternative generalized HP filters, the gHP$_n$ filter and gHP$_T$ filter, we recommend the latter for applied studies. This is because (i) if the smoothing parameters are specified as shown in Section 4, both filters give almost the same trend estimates, (ii) specifying the smoothing parameter of the gHP$_T$ filter is easier than that of the gHP$_n$ filter, and (iii) the gHP$_T$ filter also provides trend estimates for missing observations, whereas the gHP$_n$ filter does not.

We add a final remark. Numerical analysis reveals that the two alternative generalized HP filters give almost the same trend estimates when the smoothing parameter of the gHP$_n$ filter, λ_n, is specified as described in Section 4.2. As stated, this observation may reflect the fact that the corresponding smoother matrices are almost the same. Clarifying their relationship more deeply is an issue for future research.

APPENDIX

A.1. Introduction

The Appendix is organized as follows. In Section A.2 (resp. Sections A.3 and A.4), more details on the gHP$_n$ (resp. gHP$_T$) filter are given. In Section A.5, another minimization problem that gives the same trend estimate as the gHP$_T$ filter is presented. In Section A.6, more details on specifying λ_n in (2) are provided. In Section A.7, miscellaneous proofs are presented.

A.1.1. *Notations.* We fix some additional notations. Let $\boldsymbol{\Psi} = [\boldsymbol{S}'_{\perp}, \boldsymbol{S}']' \in \mathbb{R}^{T\times T}$,

$$\boldsymbol{E}_{(m)} = \begin{bmatrix} 0 & \cdots & 0 \\ 1 & \ddots & \vdots \\ \vdots & \ddots & 0 \\ 1 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{m\times(m-1)},$$

and

$$\begin{bmatrix} \boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} \\ \boldsymbol{\Phi}_{21} & \boldsymbol{\Phi}_{22} \end{bmatrix} = \boldsymbol{\Psi}\boldsymbol{D}'_T\boldsymbol{D}_T\boldsymbol{\Psi}' \left(= \begin{bmatrix} \boldsymbol{S}_{\perp}\boldsymbol{D}'_T\boldsymbol{D}_T\boldsymbol{S}'_{\perp} & \boldsymbol{S}_{\perp}\boldsymbol{D}'_T\boldsymbol{D}_T\boldsymbol{S}' \\ \boldsymbol{S}\boldsymbol{D}'_T\boldsymbol{D}_T\boldsymbol{S}'_{\perp} & \boldsymbol{S}\boldsymbol{D}'_T\boldsymbol{D}_T\boldsymbol{S}' \end{bmatrix} \right). \tag{A.1}$$

https://doi.org/10.1017/S0266466621000189 Published online by Cambridge University Press

A.1.2. *Some Preliminary Results.* We list some preliminary results as follows. (i) $\boldsymbol{E}_{(m)}$ is a right inverse of $\boldsymbol{D}_{(m)}$, i.e., $\boldsymbol{D}_{(m)}\boldsymbol{E}_{(m)} = \boldsymbol{I}_{m-1}$. (ii) By definition of $\boldsymbol{D}_T$, it follows that $\boldsymbol{D}_T = \boldsymbol{D}_{(T-1)}\boldsymbol{D}_{(T)}$. (iii) Given that $\boldsymbol{\Psi} = [\boldsymbol{S}_\perp', \boldsymbol{S}']'$ is a permutation matrix, $\boldsymbol{\Psi}$ is an orthogonal matrix and we thus have $\boldsymbol{\Psi}'\boldsymbol{\Psi} = \boldsymbol{S}_\perp'\boldsymbol{S}_\perp + \boldsymbol{S}'\boldsymbol{S} = \boldsymbol{I}_T$ and

$$\boldsymbol{\Psi}\boldsymbol{\Psi}' = \begin{bmatrix} \boldsymbol{S}_\perp \\ \boldsymbol{S} \end{bmatrix} [\boldsymbol{S}_\perp', \boldsymbol{S}'] = \begin{bmatrix} \boldsymbol{S}_\perp\boldsymbol{S}_\perp' & \boldsymbol{S}_\perp\boldsymbol{S}' \\ \boldsymbol{S}\boldsymbol{S}_\perp' & \boldsymbol{S}\boldsymbol{S}' \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_{T-n} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_n \end{bmatrix}. \tag{A.2}$$

(iv) Given that $\boldsymbol{\iota}_n = \boldsymbol{S}\boldsymbol{\iota}_T$ and $\boldsymbol{\tau}_n = \boldsymbol{S}\boldsymbol{\tau}_T$, it follows that $\boldsymbol{\Pi}_n = \boldsymbol{S}\boldsymbol{\Pi}_T$.

A.2. More Details on the gHP$_n$ Filter

In this subsection, we give more details on the gHP$_n$ filter.

Given (7), if $\boldsymbol{y}_n \neq \widehat{\boldsymbol{x}}_n$, it follows that

$$\|\boldsymbol{D}_n\boldsymbol{y}_n\|^2 = \lambda_n^{-1} f_n(\boldsymbol{y}_n) > \lambda_n^{-1} f_n(\widehat{\boldsymbol{x}}_n) > \|\boldsymbol{D}_n\widehat{\boldsymbol{x}}_n\|^2. \tag{A.3}$$

Then, given (8) and (A.3), $\widehat{\boldsymbol{x}}_n$ is a linear smoother of $\boldsymbol{y}_n$.

Let

$$\boldsymbol{A}_n = [\ \boldsymbol{\Pi}_n \ \vdots\ \boldsymbol{U}_n\ ] = \left[\begin{array}{cc:cccc} 1 & t_1 & 0 & \cdots & 0 \\ 1 & t_2 & 0 & \cdots & 0 \\ 1 & t_3 & t_3 - t_2 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 1 & t_n & t_n - t_2 & \cdots & t_n - t_{n-1} \end{array}\right] \in \mathbb{R}^{n\times n}. \tag{A.4}$$

Then, we have the following result:

LEMMA A.1. *(i) The null space of $\boldsymbol{D}_n$ equals the column space of $\boldsymbol{\Pi}_n$. (ii) $\boldsymbol{U}_n$ in (A.4) is a right inverse matrix of $\boldsymbol{D}_n$, i.e., $\boldsymbol{D}_n\boldsymbol{U}_n = \boldsymbol{I}_{n-2}$.*

Proof. Recall that $\boldsymbol{H} = \mathsf{diag}(t_2 - t_1, \ldots, t_n - t_{n-1}) \in \mathbb{R}^{(n-1)\times(n-1)}$. (i) Given $\boldsymbol{D}_{(m)}\boldsymbol{\iota}_m = \boldsymbol{0}$ and $\boldsymbol{D}_{(n)}\boldsymbol{\tau}_n = \boldsymbol{H}\boldsymbol{\iota}_{n-1}$, it immediately follows that

$$\boldsymbol{D}_n\boldsymbol{\iota}_n = \boldsymbol{D}_{(n-1)}\boldsymbol{H}^{-1}\boldsymbol{D}_{(n)}\boldsymbol{\iota}_n = \boldsymbol{0}, \tag{A.5}$$

$$\boldsymbol{D}_n\boldsymbol{\tau}_n = \boldsymbol{D}_{(n-1)}\boldsymbol{H}^{-1}\boldsymbol{D}_{(n)}\boldsymbol{\tau}_n = \boldsymbol{D}_{(n-1)}\boldsymbol{H}^{-1}\boldsymbol{H}\boldsymbol{\iota}_{n-1} = \boldsymbol{0}. \tag{A.6}$$

In addition, given that $t_i - t_{i-1} > 0$ for $i = 2, \ldots, n$, we see that $\mathsf{rank}(\boldsymbol{D}_n) = n-2$ and $\mathsf{rank}(\boldsymbol{\Pi}_n) = 2$. Combining these results completes the proof. (ii) Given that $\boldsymbol{D}_n = \boldsymbol{D}_{(n-1)}\boldsymbol{H}^{-1}\boldsymbol{D}_{(n)}$, $\boldsymbol{U}_n = \boldsymbol{E}_{(n)}\boldsymbol{H}\boldsymbol{E}_{(n-1)}$, and $\boldsymbol{E}_{(m)}$ is a right inverse of $\boldsymbol{D}_{(m)}$, we obtain

$$\boldsymbol{D}_n\boldsymbol{U}_n = \boldsymbol{D}_{(n-1)}\boldsymbol{H}^{-1}\boldsymbol{D}_{(n)}\boldsymbol{E}_{(n)}\boldsymbol{H}\boldsymbol{E}_{(n-1)} = \boldsymbol{I}_{n-2}.$$

■

By using Lemma A.1, we can show that the gHP$_n$ filter has the following alternate representation.

PROPOSITION A.2. *$\widehat{\boldsymbol{x}}_n$ in (8) equals $\boldsymbol{A}_n\widehat{\boldsymbol{\alpha}}_n$, where $\boldsymbol{A}_n$ is defined in (A.4) and $\widehat{\boldsymbol{\alpha}}_n$ is defined as follows:*

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_n &= \arg\min_{\boldsymbol{\alpha}_n\in\mathbb{R}^n} \|\boldsymbol{y}_n - \boldsymbol{A}_n\boldsymbol{\alpha}_n\|^2 + \lambda_n\|\boldsymbol{J}_n\boldsymbol{\alpha}_n\|^2 \\ &= \left(\boldsymbol{A}_n'\boldsymbol{A}_n + \lambda_n\boldsymbol{J}_n'\boldsymbol{J}_n\right)^{-1}\boldsymbol{A}_n'\boldsymbol{y}_n. \end{aligned} \tag{A.7}$$

Proof. From Lemma A.1, we obtain

$$\boldsymbol{D}_n\boldsymbol{A}_n = \boldsymbol{D}_n[\boldsymbol{\Pi}_n, \boldsymbol{U}_n] = [\boldsymbol{0}, \boldsymbol{I}_{n-2}] = \boldsymbol{J}_n \in \mathbb{R}^{(n-2)\times n}. \tag{A.8}$$

In addition, given $t_1 < \cdots < t_n$, it follows that $|\boldsymbol{A}_n| = \prod_{i=2}^n (t_i - t_{i-1}) > 0$, which indicates that $\boldsymbol{A}_n$ is nonsingular. Combining these results yields

$$\boldsymbol{J}_n\boldsymbol{A}_n^{-1} = \boldsymbol{D}_n. \tag{A.9}$$

Given (A.9), it follows that

$$\begin{aligned}\boldsymbol{A}_n\widehat{\boldsymbol{\alpha}}_n &= \boldsymbol{A}_n\left(\boldsymbol{A}_n'\boldsymbol{A}_n + \lambda_n\boldsymbol{J}_n'\boldsymbol{J}_n\right)^{-1}\boldsymbol{A}_n'\boldsymbol{y}_n \\ &= \left\{\boldsymbol{I}_n + \lambda_n\left(\boldsymbol{J}_n\boldsymbol{A}_n^{-1}\right)'\left(\boldsymbol{J}_n\boldsymbol{A}_n^{-1}\right)\right\}^{-1}\boldsymbol{y}_n \\ &= (\boldsymbol{I}_n + \lambda_n\boldsymbol{D}_n'\boldsymbol{D}_n)^{-1}\boldsymbol{y}_n = \widehat{\boldsymbol{x}}_n. \end{aligned}$$

■

Remark A.3. (i) The truncated power basis functions of degree 1 with knots located at $t_2, \ldots, t_{n-1}$, denoted by $g_1(x), \ldots, g_n(x)$, are continuous piecewise linear functions defined by $g_1(x) = 1$, $g_2(x) = x$, and for $i = 3, \ldots, n$,

$$g_i(x) = \begin{cases} x - t_{i-1} & (x > t_{i-1}), \\ 0 & (x \le t_{i-1}). \end{cases} \tag{A.10}$$

Let

$$\boldsymbol{B} = \left[\begin{array}{cc:cccc} g_1(t_1) & g_2(t_1) & g_3(t_1) & \cdots & g_n(t_1) \\ g_1(t_2) & g_2(t_2) & g_3(t_2) & \cdots & g_n(t_2) \\ g_1(t_3) & g_2(t_3) & g_3(t_3) & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & g_n(t_{n-1}) \\ g_1(t_n) & g_2(t_n) & g_3(t_n) & \cdots & g_n(t_n) \end{array}\right] \in \mathbb{R}^{n\times n}.$$

Then, it is observable that $\boldsymbol{A}_n = \boldsymbol{B}$, and thus Proposition A.2 implies that the gHP$_n$ filter can be regarded as a penalized truncated power basis spline of degree 1.[4] (ii) Proposition A.2 is a generalization of Paige and Trindade (2010, Theorem 4.1) because Proposition A.2 reduces to it if $n = T$.

Recall that $\boldsymbol{Q} = \boldsymbol{I}_n - \boldsymbol{P}$, where $\boldsymbol{P} = \boldsymbol{\Pi}_n(\boldsymbol{\Pi}_n'\boldsymbol{\Pi}_n)^{-1}\boldsymbol{\Pi}_n'$. Given Proposition A.2, from, e.g., Yamada (2017, Theorem 3.1), we have

$$\widehat{\boldsymbol{x}}_n = \boldsymbol{A}_n\widehat{\boldsymbol{\alpha}}_n = \boldsymbol{\Pi}_n\widehat{\boldsymbol{\beta}} + \boldsymbol{Q}\boldsymbol{U}_n\widehat{\boldsymbol{\gamma}}_n, \tag{A.11}$$

[4]See, e.g., Ruppert, Wand, and Carroll (2003, pp. 69–70) for more details about penalized truncated power basis splines.

where $\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}_n \in \mathbb{R}^2} \|\boldsymbol{y}_n - \boldsymbol{\Pi}_n \boldsymbol{\beta}_n\|^2 = (\boldsymbol{\Pi}_n'\boldsymbol{\Pi}_n)^{-1}\boldsymbol{\Pi}_n'\boldsymbol{y}_n$ and

$$\widehat{\boldsymbol{\gamma}}_n = \arg\min_{\boldsymbol{\gamma}_n \in \mathbb{R}^{n-2}} \|\boldsymbol{Q}\boldsymbol{y}_n - \boldsymbol{Q}\boldsymbol{U}_n\boldsymbol{\gamma}_n\|^2 + \lambda_n\|\boldsymbol{\gamma}_n\|^2$$
$$= (\boldsymbol{U}_n'\boldsymbol{Q}\boldsymbol{U}_n + \lambda_n\boldsymbol{I}_{n-2})^{-1}\boldsymbol{U}_n'\boldsymbol{Q}\boldsymbol{y}_n, \quad \textbf{(A.12)}$$

which is a ridge regression estimate as in Hoerl and Kennard (1970).

Remark A.4. As stated, $\boldsymbol{U}_n$ is a right inverse matrix of $\boldsymbol{D}_n$ such that $\boldsymbol{D}_n\boldsymbol{U}_n = \boldsymbol{I}_{n-2}$. Instead of $\boldsymbol{U}_n$, we may consider another right inverse matrix of $\boldsymbol{D}_n$. Now, let us use $\boldsymbol{D}_{n,r}^{-1} = \boldsymbol{D}_n'(\boldsymbol{D}_n\boldsymbol{D}_n')^{-1} \in \mathbb{R}^{n\times(n-2)}$ as a right inverse matrix of $\boldsymbol{D}_n$. $\boldsymbol{D}_{n,r}^{-1}$ is attractive because $(\boldsymbol{D}_{n,r}^{-1})'\boldsymbol{\Pi}_n = (\boldsymbol{D}_n\boldsymbol{D}_n')^{-1}\boldsymbol{D}_n\boldsymbol{\Pi}_n = \boldsymbol{0}$. In this case, we have

$$\widehat{\boldsymbol{x}}_n = \boldsymbol{P}\boldsymbol{y}_n + (\boldsymbol{D}_{n,r}^{-1})\left\{(\boldsymbol{D}_{n,r}^{-1})'(\boldsymbol{D}_{n,r}^{-1}) + \lambda_n\boldsymbol{I}_{n-2}\right\}^{-1}(\boldsymbol{D}_{n,r}^{-1})'\boldsymbol{y}_n$$
$$= \boldsymbol{P}\boldsymbol{y}_n + \boldsymbol{D}_n'(\boldsymbol{D}_n\boldsymbol{D}_n')^{-1/2}(\boldsymbol{I}_{n-2} + \lambda_n\boldsymbol{D}_n\boldsymbol{D}_n')^{-1}(\boldsymbol{D}_n\boldsymbol{D}_n')^{-1/2}\boldsymbol{D}_n\boldsymbol{y}_n, \quad \textbf{(A.13)}$$

which corresponds to Theorem 1 of Phillips and Jin (2020).[5]

Proof of (A.13). Let $\boldsymbol{C} = [\boldsymbol{\Pi}_n, \boldsymbol{D}_{n,r}^{-1}] \in \mathbb{R}^{n\times n}$. Given $(\boldsymbol{D}_{n,r}^{-1})'\boldsymbol{\Pi}_n = \boldsymbol{0}$, it follows that

$$(\boldsymbol{C}'\boldsymbol{C} + \lambda_n\boldsymbol{C}'\boldsymbol{D}_n'\boldsymbol{D}_n\boldsymbol{C})^{-1} = \left(\begin{bmatrix} \boldsymbol{\Pi}_n'\boldsymbol{\Pi}_n & \boldsymbol{0} \\ \boldsymbol{0} & (\boldsymbol{D}_{n,r}^{-1})'(\boldsymbol{D}_{n,r}^{-1}) \end{bmatrix} + \lambda_n\boldsymbol{J}_n'\boldsymbol{J}_n\right)^{-1}$$
$$= \begin{bmatrix} (\boldsymbol{\Pi}_n'\boldsymbol{\Pi}_n)^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \left\{(\boldsymbol{D}_{n,r}^{-1})'(\boldsymbol{D}_{n,r}^{-1}) + \lambda_n\boldsymbol{I}_{n-2}\right\}^{-1} \end{bmatrix}.$$

Thus, given that $\boldsymbol{C}$ is nonsingular, we have

$$\widehat{\boldsymbol{x}}_n = (\boldsymbol{I}_n + \lambda_n\boldsymbol{D}_n'\boldsymbol{D}_n)^{-1}\boldsymbol{y}_n = \boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C} + \lambda_n\boldsymbol{C}'\boldsymbol{D}_n'\boldsymbol{D}_n\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{y}_n$$
$$= \boldsymbol{P}\boldsymbol{y}_n + (\boldsymbol{D}_{n,r}^{-1})\left\{(\boldsymbol{D}_{n,r}^{-1})'(\boldsymbol{D}_{n,r}^{-1}) + \lambda_n\boldsymbol{I}_{n-2}\right\}^{-1}(\boldsymbol{D}_{n,r}^{-1})'\boldsymbol{y}_n$$
$$= \boldsymbol{P}\boldsymbol{y}_n + \boldsymbol{D}_n'(\boldsymbol{D}_n\boldsymbol{D}_n')^{-1}\left\{(\boldsymbol{D}_n\boldsymbol{D}_n')^{-1} + \lambda_n\boldsymbol{I}_{n-2}\right\}^{-1}(\boldsymbol{D}_n\boldsymbol{D}_n')^{-1}\boldsymbol{D}_n\boldsymbol{y}_n$$
$$= \boldsymbol{P}\boldsymbol{y}_n + \boldsymbol{D}_n'(\boldsymbol{D}_n\boldsymbol{D}_n')^{-1/2}(\boldsymbol{I}_{n-2} + \lambda_n\boldsymbol{D}_n\boldsymbol{D}_n')^{-1}(\boldsymbol{D}_n\boldsymbol{D}_n')^{-1/2}\boldsymbol{D}_n\boldsymbol{y}_n.$$

■

It follows from (A.12) that $\lim_{\lambda_n\to\infty}\widehat{\boldsymbol{\gamma}}_n = \boldsymbol{0}$. Thus, from (A.11) and $\boldsymbol{\Pi}_n\widehat{\boldsymbol{\beta}} = \boldsymbol{P}\boldsymbol{y}_n$, we have

$$\lim_{\lambda_n\to\infty}\widehat{\boldsymbol{x}}_n = \boldsymbol{\Pi}_n\widehat{\boldsymbol{\beta}} + \boldsymbol{Q}\boldsymbol{U}_n\lim_{\lambda_n\to\infty}\widehat{\boldsymbol{\gamma}}_n = \boldsymbol{P}\boldsymbol{y}_n. \quad \textbf{(A.14)}$$

Remark A.5. (A.14) is an expected result from the definition of the gHP$_n$ filter. This is because, given Lemma A.1(i), when $\lambda_n \to \infty$, (4) becomes

$$\min_{\boldsymbol{x}_n \in \mathbb{S}(\boldsymbol{\Pi}_n)} \|\boldsymbol{y}_n - \boldsymbol{x}_n\|^2, \quad \textbf{(A.15)}$$

[5] See also Yamada (2015, 2018c).

https://doi.org/10.1017/S0266466621000189 Published online by Cambridge University Press

where $\mathbb{S}(\boldsymbol{\Pi}_n)$ denotes the column space of $\boldsymbol{\Pi}_n$. By letting $\boldsymbol{x}_n = \boldsymbol{\Pi}_n\boldsymbol{\beta}$, (A.15) is represented as $\min_{\boldsymbol{\beta}\in\mathbb{R}^2} \|\boldsymbol{y}_n - \boldsymbol{\Pi}_n\boldsymbol{\beta}\|^2$.

From Lemma A.1(i), we obtain the following result:

LEMMA A.6. $(\boldsymbol{I}_n + \lambda_n \boldsymbol{D}_n'\boldsymbol{D}_n)^{-1}\boldsymbol{\Pi}_n = \boldsymbol{\Pi}_n$.

Proof. Given $\boldsymbol{D}_n\boldsymbol{\Pi}_n = \boldsymbol{0}$ from Lemma A.1(i), it follows that $(\boldsymbol{I}_n + \lambda_n \boldsymbol{D}_n'\boldsymbol{D}_n)\boldsymbol{\Pi}_n = \boldsymbol{\Pi}_n$. Premultiplying it by $(\boldsymbol{I}_n + \lambda_n \boldsymbol{D}_n'\boldsymbol{D}_n)^{-1}$ completes the proof. ■

Given $\boldsymbol{\Pi}_n = [\boldsymbol{\iota}_n, \boldsymbol{\tau}_n]$, Lemma A.6 immediately leads to

$$(\boldsymbol{I}_n + \lambda_n \boldsymbol{D}_n'\boldsymbol{D}_n)^{-1}\boldsymbol{\iota}_n = \boldsymbol{\iota}_n. \tag{A.16}$$

(A.16) shows that each row of the smoother matrix, $(\boldsymbol{I}_n + \lambda_n \boldsymbol{D}_n'\boldsymbol{D}_n)^{-1}$, sums to unity. Given that $(\boldsymbol{I}_n + \lambda_n \boldsymbol{D}_n'\boldsymbol{D}_n)^{-1}$ is symmetric, transposing (A.16) leads to

$$\frac{1}{n}\boldsymbol{\iota}_n'\widehat{\boldsymbol{x}}_n = \frac{1}{n}\boldsymbol{\iota}_n'(\boldsymbol{I}_n + \lambda_n \boldsymbol{D}_n'\boldsymbol{D}_n)^{-1}\boldsymbol{y}_n = \frac{1}{n}\boldsymbol{\iota}_n'\boldsymbol{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_{t_i} \tag{A.17}$$

and accordingly $\boldsymbol{\iota}_n'(\boldsymbol{y}_n - \widehat{\boldsymbol{x}}_n) = 0$.

Consider the case where $\boldsymbol{y}_n$ belongs to the column space of $\boldsymbol{\Pi}_n$. Then, there exists $\boldsymbol{\eta}$ such that $\boldsymbol{y}_n = \boldsymbol{\Pi}_n\boldsymbol{\eta}$. Accordingly, from Lemma A.6, if $\boldsymbol{y}_n$ belongs to the column space of $\boldsymbol{\Pi}_n$, it follows that

$$\widehat{\boldsymbol{x}}_n = (\boldsymbol{I}_n + \lambda_n \boldsymbol{D}_n'\boldsymbol{D}_n)^{-1}\boldsymbol{y}_n = \boldsymbol{y}_n. \tag{A.18}$$

Furthermore, again from Lemma A.6, we have $(\boldsymbol{I}_n + \lambda_n \boldsymbol{D}_n'\boldsymbol{D}_n)^{-1}\boldsymbol{P} = \boldsymbol{P}$, which leads to

$$\begin{aligned}(\boldsymbol{I}_n + \lambda_n \boldsymbol{D}_n'\boldsymbol{D}_n)^{-1}(\boldsymbol{y}_n - \boldsymbol{P}\boldsymbol{y}_n) &= \widehat{\boldsymbol{x}}_n - (\boldsymbol{I}_n + \lambda_n \boldsymbol{D}_n'\boldsymbol{D}_n)^{-1}\boldsymbol{P}\boldsymbol{y}_n \\ &= \widehat{\boldsymbol{x}}_n - \boldsymbol{P}\boldsymbol{y}_n.\end{aligned}$$

Thus, we obtain

$$\widehat{\boldsymbol{x}}_n = \boldsymbol{P}\boldsymbol{y}_n + (\boldsymbol{I}_n + \lambda_n \boldsymbol{D}_n'\boldsymbol{D}_n)^{-1}(\boldsymbol{y}_n - \boldsymbol{P}\boldsymbol{y}_n). \tag{A.19}$$

Remark A.7. (i) (A.19) implies that $\widehat{\boldsymbol{x}}_n$ consists of $\boldsymbol{P}\boldsymbol{y}_n$, which is a linear trend estimated by ordinary least squares, and $(\boldsymbol{I}_n + \lambda_n \boldsymbol{D}_n'\boldsymbol{D}_n)^{-1}(\boldsymbol{y}_n - \boldsymbol{P}\boldsymbol{y}_n)$, which represents a low-frequency part of the residuals, $\boldsymbol{y}_n - \boldsymbol{P}\boldsymbol{y}_n$.

By applying the Sherman–Morrison–Woodbury (SMW) formula (Seber, 2008) to the smoother matrix in (8), we obtain

$$(\boldsymbol{I}_n + \lambda_n \boldsymbol{D}_n'\boldsymbol{D}_n)^{-1} = \boldsymbol{I}_n - \boldsymbol{D}_n'\left(\boldsymbol{D}_n\boldsymbol{D}_n' + \lambda_n^{-1}\boldsymbol{I}_{n-2}\right)^{-1}\boldsymbol{D}_n. \tag{A.20}$$

Remark A.8. (i) Given Lemma A.1(i), we can also prove Lemma A.6 by postmultiplying (A.20) by $\boldsymbol{\Pi}_n$. (ii) An alternative somewhat long proof of (A.20) is given in Section A.7.3.

Given (A.20), it follows that

$$\widehat{\boldsymbol{x}}_n = \boldsymbol{y}_n - \boldsymbol{D}_n'\widehat{\boldsymbol{\theta}}_n, \tag{A.21}$$

where

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}_n \in \mathbb{R}^{n-2}} \|\boldsymbol{y}_n - \boldsymbol{D}_n' \boldsymbol{\theta}_n\|^2 + \lambda_n^{-1} \|\boldsymbol{\theta}_n\|^2$$
$$= \left(\boldsymbol{D}_n \boldsymbol{D}_n' + \lambda_n^{-1} \boldsymbol{I}_{n-2}\right)^{-1} \boldsymbol{D}_n \boldsymbol{y}_n. \tag{A.22}$$

We remark that (A.22) is also a ridge regression estimate. Given that $\lambda_n^{-1} \to \infty$ as $\lambda_n \to 0$, it follows from (A.22) that $\lim_{\lambda_n \to 0} \widehat{\boldsymbol{\theta}}_n = \boldsymbol{0}$. Thus, we have

$$\lim_{\lambda_n \to 0} \widehat{\boldsymbol{x}}_n = \boldsymbol{y}_n - \boldsymbol{D}_n' \lim_{\lambda_n \to 0} \widehat{\boldsymbol{\theta}}_n = \boldsymbol{y}_n. \tag{A.23}$$

A.3. More Details on the gHP$_T$ Filter (I)

In this subsection, we give more details on the gHP$_T$ filter.

Given (10) and $\boldsymbol{S}\boldsymbol{y}_T = \boldsymbol{y}_n$, if $\boldsymbol{y}_T \neq \widehat{\boldsymbol{x}}_T$, it follows that

$$\|\boldsymbol{D}_T \boldsymbol{y}_T\|^2 = \lambda_T^{-1} f_T(\boldsymbol{y}_T) > \lambda_T^{-1} f_T(\widehat{\boldsymbol{x}}_T) > \|\boldsymbol{D}_T \widehat{\boldsymbol{x}}_T\|^2. \tag{A.24}$$

Then, given (11) and (A.24), $\widehat{\boldsymbol{x}}_T$ is a linear smoother of $\boldsymbol{y}_T$.

Concerning the smoother matrices of the gHP$_T$ filter in (11) and (12), we obtain the following result:

LEMMA A.9. *(i)* $\boldsymbol{\Pi}_T = (\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T' \boldsymbol{D}_T)^{-1} \boldsymbol{S}' \boldsymbol{\Pi}_n$ *and (ii)* $\boldsymbol{\Pi}_n = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T' \boldsymbol{D}_T)^{-1} \boldsymbol{S}' \boldsymbol{\Pi}_n$.

Proof. Given $\boldsymbol{D}_T \boldsymbol{\Pi}_T = \boldsymbol{0}$, $\boldsymbol{S}\boldsymbol{\Pi}_T = \boldsymbol{\Pi}_n$, and Lemma 4, it follows that $(\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T' \boldsymbol{D}_T)\boldsymbol{\Pi}_T = \boldsymbol{S}'\boldsymbol{S}\boldsymbol{\Pi}_T = \boldsymbol{S}'\boldsymbol{\Pi}_n$, which leads to $\boldsymbol{\Pi}_T = (\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T' \boldsymbol{D}_T)^{-1} \boldsymbol{S}' \boldsymbol{\Pi}_n$. By premultiplying the above equation by $\boldsymbol{S}$, we obtain $\boldsymbol{\Pi}_n = \boldsymbol{S}\boldsymbol{\Pi}_T = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T' \boldsymbol{D}_T)^{-1} \boldsymbol{S}' \boldsymbol{\Pi}_n$. ■

Given $\boldsymbol{\Pi}_n = [\boldsymbol{\iota}_n, \boldsymbol{\tau}_n]$, Lemma A.9 immediately leads to

$$(\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T' \boldsymbol{D}_T)^{-1} \boldsymbol{S}' \boldsymbol{\iota}_n = \boldsymbol{\iota}_T, \tag{A.25}$$

$$\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T' \boldsymbol{D}_T)^{-1} \boldsymbol{S}' \boldsymbol{\iota}_n = \boldsymbol{\iota}_n, \tag{A.26}$$

which show that each row of the smoother matrices in (11) and (12) sums to unity. Given that $\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T' \boldsymbol{D}_T)^{-1} \boldsymbol{S}'$ is symmetric, transposing (A.26) leads to

$$\frac{1}{n}\boldsymbol{\iota}_n' \boldsymbol{S}\widehat{\boldsymbol{x}}_T = \frac{1}{n}\boldsymbol{\iota}_n' \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T' \boldsymbol{D}_T)^{-1} \boldsymbol{S}' \boldsymbol{y}_n = \frac{1}{n}\boldsymbol{\iota}_n' \boldsymbol{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_{t_i} \tag{A.27}$$

and accordingly $\boldsymbol{\iota}_n'(\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T) = 0$.

Consider the case where $\boldsymbol{y}_n$ belongs to the column space of $\boldsymbol{\Pi}_n$. Then, there exists $\boldsymbol{\eta}$ such that $\boldsymbol{y}_n = \boldsymbol{\Pi}_n \boldsymbol{\eta}$. Accordingly, from Lemma A.9, if $\boldsymbol{y}_n$ belongs to the column space of $\boldsymbol{\Pi}_n$, it follows that

$$\boldsymbol{S}\widehat{\boldsymbol{x}}_T = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T' \boldsymbol{D}_T)^{-1} \boldsymbol{S}' \boldsymbol{\Pi}_n \boldsymbol{\eta} = \boldsymbol{\Pi}_n \boldsymbol{\eta} = \boldsymbol{y}_n, \tag{A.28}$$

which implies that if $\boldsymbol{y}_n$ belongs to the column space of $\boldsymbol{\Pi}_n$, then $\boldsymbol{S}\widehat{\boldsymbol{x}}_T = \boldsymbol{y}_n$.

Furthermore, from Lemma A.9, we obtain

$$\widehat{\boldsymbol{x}}_T = \boldsymbol{\Pi}_T\widehat{\boldsymbol{\beta}} + (\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'(\boldsymbol{y}_n - \boldsymbol{P}\boldsymbol{y}_n), \quad \textbf{(A.29)}$$

$$\boldsymbol{S}\widehat{\boldsymbol{x}}_T = \boldsymbol{P}\boldsymbol{y}_n + \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'(\boldsymbol{y}_n - \boldsymbol{P}\boldsymbol{y}_n). \quad \textbf{(A.30)}$$

Proofs of (A.29) and (A.30). Given Lemma A.9, we have

$$\begin{aligned}
&(\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'(\boldsymbol{y}_n - \boldsymbol{\Pi}_n\widehat{\boldsymbol{\beta}}) \\
&\quad = (\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'\boldsymbol{y}_n - (\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'\boldsymbol{\Pi}_n\widehat{\boldsymbol{\beta}} \\
&\quad = \widehat{\boldsymbol{x}}_T - \boldsymbol{\Pi}_T\widehat{\boldsymbol{\beta}},
\end{aligned}$$

which proves (A.29). Next, premultiplying (A.29) by $\boldsymbol{S}$ leads to (A.30). ■

(A.30) implies that $\boldsymbol{S}\widehat{\boldsymbol{x}}_T$ consists of $\boldsymbol{P}\boldsymbol{y}_n$, which is a linear trend estimated by ordinary least squares, and $\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'(\boldsymbol{y}_n - \boldsymbol{P}\boldsymbol{y}_n)$, which represent a low-frequency part of the residuals, $\boldsymbol{y}_n - \boldsymbol{P}\boldsymbol{y}_n$.

Let $\boldsymbol{A}_T \in \mathbb{R}^{T\times T}$ and $\boldsymbol{U}_T \in \mathbb{R}^{T\times(T-2)}$ be matrices defined by (33) and

$$\begin{aligned}
\widehat{\boldsymbol{\alpha}}_T &= \arg\min_{\boldsymbol{\alpha}_T\in\mathbb{R}^T} \|\boldsymbol{y}_n - \boldsymbol{S}\boldsymbol{A}_T\boldsymbol{\alpha}_T\|^2 + \lambda_T\|\boldsymbol{J}_T\boldsymbol{\alpha}_T\|^2 \\
&= (\boldsymbol{A}_T'\boldsymbol{S}'\boldsymbol{S}\boldsymbol{A}_T + \lambda_T\boldsymbol{J}_T'\boldsymbol{J}_T)^{-1}\boldsymbol{A}_T'\boldsymbol{S}'\boldsymbol{y}_n. \quad \textbf{(A.31)}
\end{aligned}$$

Then, given $|\boldsymbol{A}_T| = 1 \neq 0$, it follows that

$$\widehat{\boldsymbol{x}}_T = \boldsymbol{A}_T\widehat{\boldsymbol{\alpha}}_T, \quad \textbf{(A.32)}$$

which immediately leads to

$$\boldsymbol{S}\widehat{\boldsymbol{x}}_T = \boldsymbol{S}\boldsymbol{A}_T\widehat{\boldsymbol{\alpha}}_T, \quad \textbf{(A.33)}$$

$$\boldsymbol{S}_\perp\widehat{\boldsymbol{x}}_T = \boldsymbol{S}_\perp\boldsymbol{A}_T\widehat{\boldsymbol{\alpha}}_T. \quad \textbf{(A.34)}$$

Proof of (A.32). Given $|\boldsymbol{A}_T| = 1$, $\boldsymbol{A}_T$ is nonsingular. Then, given $\boldsymbol{J}_T = \boldsymbol{D}_T\boldsymbol{A}_T$ (Paige and Trindade, 2010), it follows that

$$\begin{aligned}
\boldsymbol{A}_T\widehat{\boldsymbol{\alpha}}_T &= \boldsymbol{A}_T(\boldsymbol{A}_T'\boldsymbol{S}'\boldsymbol{S}\boldsymbol{A}_T + \lambda_T\boldsymbol{J}_T'\boldsymbol{J}_T)^{-1}\boldsymbol{A}_T'\boldsymbol{S}'\boldsymbol{y}_n \\
&= \boldsymbol{A}_T(\boldsymbol{A}_T'\boldsymbol{S}'\boldsymbol{S}\boldsymbol{A}_T + \lambda_T\boldsymbol{A}_T'\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{A}_T)^{-1}\boldsymbol{A}_T'\boldsymbol{S}'\boldsymbol{y}_n \\
&= (\boldsymbol{S}'\boldsymbol{S} + \lambda_T\boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'\boldsymbol{y}_n = \widehat{\boldsymbol{x}}_T.
\end{aligned}$$

■

Given $\boldsymbol{S}\boldsymbol{\Pi}_T = \boldsymbol{\Pi}_n$ and (A.33), from, e.g., Yamada (2017, Theorem 3.1), it follows that

$$\boldsymbol{S}\widehat{\boldsymbol{x}}_T = \boldsymbol{S}\boldsymbol{A}_T\widehat{\boldsymbol{\alpha}}_T = \boldsymbol{\Pi}_n\widehat{\boldsymbol{\beta}} + \boldsymbol{Q}\boldsymbol{S}\boldsymbol{U}_T\widehat{\boldsymbol{\gamma}}_T, \quad \textbf{(A.35)}$$

where $\widehat{\boldsymbol{\beta}} = (\boldsymbol{\Pi}_n'\boldsymbol{\Pi}_n)^{-1}\boldsymbol{\Pi}_n'\boldsymbol{y}_n$ and

$$\begin{aligned}
\widehat{\boldsymbol{\gamma}}_T &= \arg\min_{\boldsymbol{\gamma}_T\in\mathbb{R}^{T-2}} \|\boldsymbol{Q}\boldsymbol{y}_n - \boldsymbol{Q}\boldsymbol{S}\boldsymbol{U}_T\boldsymbol{\gamma}_T\|^2 + \lambda_T\|\boldsymbol{\gamma}_T\|^2 \\
&= (\boldsymbol{U}_T'\boldsymbol{S}'\boldsymbol{Q}\boldsymbol{S}\boldsymbol{U}_T + \lambda_T\boldsymbol{I}_{T-2})^{-1}\boldsymbol{U}_T'\boldsymbol{S}'\boldsymbol{Q}\boldsymbol{y}_n, \quad \textbf{(A.36)}
\end{aligned}$$

which is a ridge regression estimate. Given $\lim_{\lambda_T\to\infty}\widehat{\boldsymbol{\gamma}}_T=\mathbf{0}$ and $\boldsymbol{\Pi}_n\widehat{\boldsymbol{\beta}}=\boldsymbol{P}\boldsymbol{y}_n$, we have

$$\lim_{\lambda_T\to\infty}\boldsymbol{S}\widehat{\boldsymbol{x}}_T=\lim_{\lambda_T\to\infty}\boldsymbol{S}\boldsymbol{A}_T\widehat{\boldsymbol{\alpha}}_T=\boldsymbol{\Pi}_n\widehat{\boldsymbol{\beta}}+\boldsymbol{Q}\boldsymbol{S}\boldsymbol{U}_T\lim_{\lambda_T\to\infty}\widehat{\boldsymbol{\gamma}}_T=\boldsymbol{P}\boldsymbol{y}_n. \quad \textbf{(A.37)}$$

A.4. More Details on the gHP$_T$ Filter (II)

In this subsection, we give more details on the gHP$_T$ filter.

First, we give the following result, which is crucial for the later discussions.

LEMMA A.10. *Under Assumption 1,* $\boldsymbol{\Phi}_{11}=\boldsymbol{S}_\perp\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{S}_\perp'\in\mathbb{R}^{(T-n)\times(T-n)}$ *is positive definite.*

Proof. Let $\boldsymbol{\xi}\neq\mathbf{0}$ be a $(T-n)$-dimensional column vector. Given that $\boldsymbol{S}_\perp'\in\mathbb{R}^{T\times(T-n)}$ is a matrix that consists of $(T-n)$ columns of $\boldsymbol{I}_T$, $\boldsymbol{S}_\perp'$ is of full column rank and n rows of $\boldsymbol{S}_\perp'$ are zero vectors. Under Assumption 1, because $n\geq 3$, at least three rows of $\boldsymbol{S}_\perp'$ are zero vectors. Thus, it follows that (i) $\boldsymbol{S}_\perp'\boldsymbol{\xi}\neq\mathbf{0}$ and (ii) $\boldsymbol{S}_\perp'\boldsymbol{\xi}$ has at least three zeros. For that reason, under Assumption 1, $\boldsymbol{S}_\perp'\boldsymbol{\xi}$ cannot belong to the column space of $\boldsymbol{\Pi}_T$. Therefore, given that the null space of $\boldsymbol{D}_T$ equals the column space of $\boldsymbol{\Pi}_T$, we have $\boldsymbol{D}_T\boldsymbol{S}_\perp'\boldsymbol{\xi}\neq\mathbf{0}$, which leads to $\|\boldsymbol{D}_T\boldsymbol{S}_\perp'\boldsymbol{\xi}\|^2>\mathbf{0}$. ■

The following proposition gives another representation of the gHP$_T$ filter.

PROPOSITION A.11. $\boldsymbol{S}\widehat{\boldsymbol{x}}_T$ *in (12) equals* $\widehat{\boldsymbol{\psi}}$ *which is defined as follows:*

$$\widehat{\boldsymbol{\psi}}=\arg\min_{\boldsymbol{\psi}\in\mathbb{R}^n}\|\boldsymbol{y}_n-\boldsymbol{\psi}\|^2+\lambda_T\|\boldsymbol{F}\boldsymbol{\psi}\|^2=(\boldsymbol{I}_n+\lambda_T\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{y}_n, \quad \textbf{(A.38)}$$

where $\boldsymbol{F}=\boldsymbol{R}\boldsymbol{D}_T\boldsymbol{S}'\in\mathbb{R}^{(T-2)\times n}$ *and*
$\boldsymbol{R}=\boldsymbol{I}_{T-2}-\boldsymbol{D}_T\boldsymbol{S}_\perp'\boldsymbol{\Phi}_{11}^{-1}\boldsymbol{S}_\perp\boldsymbol{D}_T'$.

Proof. Given that $\boldsymbol{\Psi}=[\boldsymbol{S}_\perp',\boldsymbol{S}']'$ is an orthogonal matrix and $\boldsymbol{\Phi}_{11}$ is nonsingular from Lemma A.10, it follows that

$$\begin{aligned}
\boldsymbol{S}\widehat{\boldsymbol{x}}_T&=\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S}+\lambda_T\boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'\boldsymbol{y}_n\\
&=\boldsymbol{S}\boldsymbol{\Psi}'(\boldsymbol{\Psi}\boldsymbol{S}'\boldsymbol{S}\boldsymbol{\Psi}'+\lambda_T\boldsymbol{\Psi}\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{\Psi}')^{-1}\boldsymbol{\Psi}\boldsymbol{S}'\boldsymbol{y}_n\\
&=[\mathbf{0},\boldsymbol{I}_n]\left(\begin{bmatrix}\mathbf{0}&\mathbf{0}\\\mathbf{0}&\boldsymbol{I}_n\end{bmatrix}+\begin{bmatrix}\lambda_T\boldsymbol{\Phi}_{11}&\lambda_T\boldsymbol{\Phi}_{12}\\\lambda_T\boldsymbol{\Phi}_{21}&\lambda_T\boldsymbol{\Phi}_{22}\end{bmatrix}\right)^{-1}\begin{bmatrix}\mathbf{0}\\\boldsymbol{I}_n\end{bmatrix}\boldsymbol{y}_n\\
&=[\mathbf{0},\boldsymbol{I}_n]\begin{bmatrix}\lambda_T\boldsymbol{\Phi}_{11}&\lambda_T\boldsymbol{\Phi}_{12}\\\lambda_T\boldsymbol{\Phi}_{21}&\boldsymbol{I}_n+\lambda_T\boldsymbol{\Phi}_{22}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{0}\\\boldsymbol{I}_n\end{bmatrix}\boldsymbol{y}_n\\
&=\left\{(\boldsymbol{I}_n+\lambda_T\boldsymbol{\Phi}_{22})-\lambda_T\boldsymbol{\Phi}_{21}(\lambda_T\boldsymbol{\Phi}_{11})^{-1}\lambda_T\boldsymbol{\Phi}_{12}\right\}^{-1}\boldsymbol{y}_n\\
&=\left\{\boldsymbol{I}_n+\lambda_T\left(\boldsymbol{\Phi}_{22}-\boldsymbol{\Phi}_{21}\boldsymbol{\Phi}_{11}^{-1}\boldsymbol{\Phi}_{12}\right)\right\}^{-1}\boldsymbol{y}_n. \quad \textbf{(A.39)}
\end{aligned}$$

We note that $\boldsymbol{\Phi}_{ij}$ for $i,j=1,2$ are defined in (A.1). Here, given that $\boldsymbol{R}$ is an orthogonal projection matrix, $\boldsymbol{\Phi}_{22}-\boldsymbol{\Phi}_{21}\boldsymbol{\Phi}_{11}^{-1}\boldsymbol{\Phi}_{12}$ can be represented as follows:

$$\begin{aligned}
&\boldsymbol{\Phi}_{22}-\boldsymbol{\Phi}_{21}\boldsymbol{\Phi}_{11}^{-1}\boldsymbol{\Phi}_{12}\\
&\quad=\boldsymbol{S}\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{S}'-\boldsymbol{S}\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{S}_\perp'(\boldsymbol{S}_\perp\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{S}_\perp')^{-1}\boldsymbol{S}_\perp\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{S}'
\end{aligned}$$

https://doi.org/10.1017/S0266466621000189 Published online by Cambridge University Press

$$= \boldsymbol{S}\boldsymbol{D}_T' \left\{ \boldsymbol{I}_{T-2} - \boldsymbol{D}_T \boldsymbol{S}_\perp' (\boldsymbol{S}_\perp \boldsymbol{D}_T' \boldsymbol{D}_T \boldsymbol{S}_\perp')^{-1} \boldsymbol{S}_\perp \boldsymbol{D}_T' \right\} \boldsymbol{D}_T \boldsymbol{S}'$$
$$= \boldsymbol{S}\boldsymbol{D}_T' \boldsymbol{R} \boldsymbol{D}_T \boldsymbol{S}' = \boldsymbol{S}\boldsymbol{D}_T' \boldsymbol{R}' \boldsymbol{R} \boldsymbol{D}_T \boldsymbol{S}' = \boldsymbol{F}'\boldsymbol{F}. \quad \textbf{(A.40)}$$

Substituting (A.40) into (A.39) leads to $\boldsymbol{S}\widehat{\boldsymbol{x}}_T = (\boldsymbol{I}_n + \lambda_T \boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{y}_n = \widehat{\boldsymbol{\psi}}$. ■

Similarly to $\boldsymbol{S}\widehat{\boldsymbol{x}}_T = \widehat{\boldsymbol{\psi}}$, we can show

$$\boldsymbol{S}_\perp \widehat{\boldsymbol{x}}_T = \widehat{\boldsymbol{\phi}}, \quad \textbf{(A.41)}$$

where

$$\widehat{\boldsymbol{\phi}} = -\boldsymbol{\Phi}_{11}^{-1}\boldsymbol{\Phi}_{12}\widehat{\boldsymbol{\psi}} = -(\boldsymbol{S}_\perp \boldsymbol{D}_T' \boldsymbol{D}_T \boldsymbol{S}_\perp')^{-1} \boldsymbol{S}_\perp \boldsymbol{D}_T' \boldsymbol{D}_T \boldsymbol{S}' \widehat{\boldsymbol{\psi}}. \quad \textbf{(A.42)}$$

Proof of (A.41). Given (A.1), (A.2), and Lemma A.10, it follows that

$$\begin{aligned}
\boldsymbol{S}_\perp \widehat{\boldsymbol{x}}_T &= \boldsymbol{S}_\perp (\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'\boldsymbol{y}_n \\
&= \boldsymbol{S}_\perp \boldsymbol{\Psi}' (\boldsymbol{\Psi}\boldsymbol{S}'\boldsymbol{S}\boldsymbol{\Psi}' + \lambda_T \boldsymbol{\Psi}\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{\Psi}')^{-1}\boldsymbol{\Psi}\boldsymbol{S}'\boldsymbol{y}_n \\
&= [\boldsymbol{I}_{T-n}, \boldsymbol{0}] \left(\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_n \end{bmatrix} + \begin{bmatrix} \lambda_T \boldsymbol{\Phi}_{11} & \lambda_T \boldsymbol{\Phi}_{12} \\ \lambda_T \boldsymbol{\Phi}_{21} & \lambda_T \boldsymbol{\Phi}_{22} \end{bmatrix} \right)^{-1} \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I}_n \end{bmatrix} \boldsymbol{y}_n \\
&= [\boldsymbol{I}_{T-n}, \boldsymbol{0}] \begin{bmatrix} \lambda_T \boldsymbol{\Phi}_{11} & \lambda_T \boldsymbol{\Phi}_{12} \\ \lambda_T \boldsymbol{\Phi}_{21} & \boldsymbol{I}_n + \lambda_T \boldsymbol{\Phi}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I}_n \end{bmatrix} \boldsymbol{y}_n \\
&= -(\lambda_T \boldsymbol{\Phi}_{11})^{-1} \lambda_T \boldsymbol{\Phi}_{12} \left\{ (\boldsymbol{I}_n + \lambda_T \boldsymbol{\Phi}_{22}) - \lambda_T \boldsymbol{\Phi}_{21} (\lambda_T \boldsymbol{\Phi}_{11})^{-1} \lambda_T \boldsymbol{\Phi}_{12} \right\}^{-1} \boldsymbol{y}_n \\
&= -\boldsymbol{\Phi}_{11}^{-1} \boldsymbol{\Phi}_{12} \widehat{\boldsymbol{\psi}}.
\end{aligned}$$

■

Remark A.12. (i) Let $\boldsymbol{\phi} \in \mathbb{R}^{T-n}$ and $\boldsymbol{\psi} \in \mathbb{R}^n$ such that

$$\boldsymbol{\Psi}\boldsymbol{x}_T = \begin{bmatrix} \boldsymbol{S}_\perp \\ \boldsymbol{S} \end{bmatrix} \boldsymbol{x}_T = \begin{bmatrix} \boldsymbol{\phi} \\ \boldsymbol{\psi} \end{bmatrix} \in \mathbb{R}^T. \quad \textbf{(A.43)}$$

Given that $\boldsymbol{\Psi}$ is an orthogonal matrix, we have $\boldsymbol{x}_T = \boldsymbol{S}_\perp'\boldsymbol{\phi} + \boldsymbol{S}'\boldsymbol{\psi}$. Then, (9) can be represented as follows:

$$\min_{\substack{\boldsymbol{\phi} \in \mathbb{R}^{T-n} \\ \boldsymbol{\psi} \in \mathbb{R}^n}} \|\boldsymbol{y}_n - \boldsymbol{\psi}\|^2 + \lambda_T (\boldsymbol{\phi}'\boldsymbol{\Phi}_{11}\boldsymbol{\phi} + \boldsymbol{\phi}'\boldsymbol{\Phi}_{12}\boldsymbol{\psi} + \boldsymbol{\psi}'\boldsymbol{\Phi}_{21}\boldsymbol{\phi} + \boldsymbol{\psi}'\boldsymbol{\Phi}_{22}\boldsymbol{\psi}). \quad \textbf{(A.44)}$$

Of course, $(\widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\phi}})$ minimizes the objective function in (A.44). Notice that the optimality conditions for (A.44) are

$$-(\boldsymbol{y}_n - \widehat{\boldsymbol{\psi}}) + \lambda_T (\boldsymbol{\Phi}_{21}\widehat{\boldsymbol{\phi}} + \boldsymbol{\Phi}_{22}\widehat{\boldsymbol{\psi}}) = \boldsymbol{0}, \quad \textbf{(A.45)}$$

$$\boldsymbol{\Phi}_{11}\widehat{\boldsymbol{\phi}} + \boldsymbol{\Phi}_{12}\widehat{\boldsymbol{\psi}} = \boldsymbol{0}. \quad \textbf{(A.46)}$$

(ii) The expression in (A.38) is called the Reinsch form (see, e.g., Hastie et al., 2009, p. 154). (iii) A Matlab user-defined function to calculate $\widehat{\boldsymbol{\psi}} (= \boldsymbol{S}\widehat{\boldsymbol{x}}_T)$ in (A.38) and $\widehat{\boldsymbol{\phi}} (= \boldsymbol{S}_\perp \widehat{\boldsymbol{x}}_T)$ in (A.42) is provided in the Online Supplementary Material.

https://doi.org/10.1017/S0266466621000189 Published online by Cambridge University Press

From Proposition A.11, we obtain the following result:

COROLLARY A.13.

$$S(S'S+\lambda_T D_T' D_T)^{-1}S' = (I_n+\lambda_T F'F)^{-1} \quad \textbf{(A.47)}$$

$$= I_n - F'\left(FF'+\lambda_T^{-1}I_{T-2}\right)^{-1}F. \quad \textbf{(A.48)}$$

Proof. (A.47) immediately follows from Proposition A.11. (A.48) follows from applying the SMW formula to $(I_n+\lambda_T F'F)^{-1}$. Of course, (A.48) can also be proved alternatively, as shown in Section A.7.3. ■

Concerning F, the next result holds.

LEMMA A.14. $F\Pi_n = 0$.

Proof. Given that $\Pi_n = S\Pi_T$, $\Psi'\Psi = S_\perp' S_\perp + S'S = I_T$, and $D_T\Pi_T = 0$, it follows that

$$F\Pi_n = RD_T S'S\Pi_T = RD_T(I_T - S_\perp' S_\perp)\Pi_T = -RD_T S_\perp' S_\perp \Pi_T$$
$$= -\left\{I_{T-2} - D_T S_\perp'(S_\perp D_T' D_T S_\perp')^{-1} S_\perp D_T'\right\} D_T S_\perp' S_\perp \Pi_T$$
$$= -D_T S_\perp' S_\perp \Pi_T + D_T S_\perp' S_\perp \Pi_T = 0.$$

■

Then, given Lemma A.14, postmultiplying (A.48) by Π_n, we obtain

$$(I_n+\lambda_T F'F)^{-1}\Pi_n = \left\{I_n - F'\left(FF'+\lambda_T^{-1}I_{T-2}\right)^{-1}F\right\}\Pi_n = \Pi_n. \quad \textbf{(A.49)}$$

Notice that (A.49) is an alternative representation of Lemma A.9(ii).

Also, from (A.48), it follows that

$$\widehat{\psi} = y_n - F'\widehat{\zeta}, \quad \textbf{(A.50)}$$

where

$$\widehat{\zeta} = \arg\min_{\zeta\in\mathbb{R}^{T-2}} \|y_n - F'\zeta\|^2 + \lambda_T^{-1}\|\zeta\|^2$$
$$= \left(FF'+\lambda_T^{-1}I_{T-2}\right)^{-1}Fy_n. \quad \textbf{(A.51)}$$

We remark that (A.51) is also a ridge regression estimate. Given that $\lambda_T^{-1}\to\infty$ as $\lambda_T\to 0$, it follows from (A.51) that $\lim_{\lambda_T\to 0}\widehat{\zeta} = 0$. Thus, we have

$$\lim_{\lambda_T\to 0}\widehat{\psi} = y_n - F'\lim_{\lambda_T\to 0}\widehat{\zeta} = y_n, \quad \textbf{(A.52)}$$

which implies that

$$\lim_{\lambda_T\to 0} S\widehat{x}_T = y_n. \quad \textbf{(A.53)}$$

https://doi.org/10.1017/S0266466621000189 Published online by Cambridge University Press

Given (A.37), (A.41), (A.42), and $\boldsymbol{D}_T\boldsymbol{\Pi}_T = \boldsymbol{0}$, we obtain

$$\lim_{\lambda_T\to\infty} \boldsymbol{S}_\perp \widehat{\boldsymbol{x}}_T = \boldsymbol{S}_\perp \boldsymbol{\Pi}_T (\boldsymbol{\Pi}_n'\boldsymbol{\Pi}_n)^{-1}\boldsymbol{\Pi}_n'\boldsymbol{y}_n = \boldsymbol{S}_\perp \boldsymbol{\Pi}_T \widehat{\boldsymbol{\beta}}. \tag{A.54}$$

Proof of (A.54). Given (A.37), (A.41), and (A.42), we have

$$\begin{aligned}\lim_{\lambda_T\to\infty} \boldsymbol{S}_\perp \widehat{\boldsymbol{x}}_T &= -\boldsymbol{\Phi}_{11}^{-1}\boldsymbol{\Phi}_{12} \lim_{\lambda_T\to\infty} \widehat{\boldsymbol{\psi}} = -\boldsymbol{\Phi}_{11}^{-1}\boldsymbol{\Phi}_{12}\boldsymbol{S}\boldsymbol{\Pi}_T(\boldsymbol{\Pi}_n'\boldsymbol{\Pi}_n)^{-1}\boldsymbol{\Pi}_n'\boldsymbol{y}_n \\ &= -\boldsymbol{\Phi}_{11}^{-1}\boldsymbol{\Phi}_{12}\boldsymbol{S}\boldsymbol{\Pi}_T\widehat{\boldsymbol{\beta}}.\end{aligned} \tag{A.55}$$

Here, given $\boldsymbol{S}'\boldsymbol{S} = \boldsymbol{I}_T - \boldsymbol{S}_\perp'\boldsymbol{S}_\perp$ and $\boldsymbol{D}_T\boldsymbol{\Pi}_T = \boldsymbol{0}$, it follows that

$$\begin{aligned}-\boldsymbol{\Phi}_{11}^{-1}\boldsymbol{\Phi}_{12}\boldsymbol{S}\boldsymbol{\Pi}_T &= -(\boldsymbol{S}_\perp\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{S}_\perp')^{-1}\boldsymbol{S}_\perp\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{S}'\boldsymbol{S}\boldsymbol{\Pi}_T \\ &= -(\boldsymbol{S}_\perp\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{S}_\perp')^{-1}\boldsymbol{S}_\perp\boldsymbol{D}_T'\boldsymbol{D}_T(\boldsymbol{I}_T - \boldsymbol{S}_\perp'\boldsymbol{S}_\perp)\boldsymbol{\Pi}_T \\ &= (\boldsymbol{S}_\perp\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{S}_\perp')^{-1}(\boldsymbol{S}_\perp\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{S}_\perp')\boldsymbol{S}_\perp\boldsymbol{\Pi}_T = \boldsymbol{S}_\perp\boldsymbol{\Pi}_T.\end{aligned} \tag{A.56}$$

Substituting (A.56) into (A.55) yields (A.54). ■

By combining (A.37) and (A.54), we obtain

$$\lim_{\lambda_T\to\infty} \boldsymbol{\Psi}\widehat{\boldsymbol{x}}_T \left(= \boldsymbol{\Psi} \lim_{\lambda_T\to\infty} \widehat{\boldsymbol{x}}_T\right) = \lim_{\lambda_T\to\infty} \begin{bmatrix}\boldsymbol{S}_\perp \\ \boldsymbol{S}\end{bmatrix}\widehat{\boldsymbol{x}}_T = \begin{bmatrix}\boldsymbol{S}_\perp \\ \boldsymbol{S}\end{bmatrix}\boldsymbol{\Pi}_T\widehat{\boldsymbol{\beta}} = \boldsymbol{\Psi}\boldsymbol{\Pi}_T\widehat{\boldsymbol{\beta}}.$$

Premultiplying the above equation by $\boldsymbol{\Psi}'$ yields

$$\lim_{\lambda_T\to\infty} \widehat{\boldsymbol{x}}_T = \boldsymbol{\Pi}_T\widehat{\boldsymbol{\beta}}. \tag{A.57}$$

On the other hand, even though $\lim_{\lambda_T\to 0} \boldsymbol{S}\widehat{\boldsymbol{x}}_T = \boldsymbol{S}\boldsymbol{y}_T (= \boldsymbol{y}_n)$, $\lim_{\lambda_T\to 0} \boldsymbol{S}_\perp\widehat{\boldsymbol{x}}_T$ is not necessarily equal to $\boldsymbol{S}_\perp\boldsymbol{y}_T$. More precisely, given (A.52) and $\boldsymbol{S}'\boldsymbol{S} = \boldsymbol{I}_T - \boldsymbol{S}_\perp'\boldsymbol{S}_\perp$, it follows that

$$\begin{aligned}\lim_{\lambda_T\to 0} \boldsymbol{S}_\perp\widehat{\boldsymbol{x}}_T &= -\boldsymbol{\Phi}_{11}^{-1}\boldsymbol{\Phi}_{12}\lim_{\lambda_T\to 0}\widehat{\boldsymbol{\psi}} = -\boldsymbol{\Phi}_{11}^{-1}\boldsymbol{\Phi}_{12}\boldsymbol{y}_n \\ &= -(\boldsymbol{S}_\perp\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{S}_\perp')^{-1}\boldsymbol{S}_\perp\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{S}'\boldsymbol{S}\boldsymbol{y}_T\end{aligned} \tag{A.58}$$

$$\begin{aligned}&= -(\boldsymbol{S}_\perp\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{S}_\perp')^{-1}\boldsymbol{S}_\perp\boldsymbol{D}_T'\boldsymbol{D}_T(\boldsymbol{I}_T - \boldsymbol{S}_\perp'\boldsymbol{S}_\perp)\boldsymbol{y}_T \\ &= \boldsymbol{S}_\perp\boldsymbol{y}_T - (\boldsymbol{S}_\perp\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{S}_\perp')^{-1}\boldsymbol{S}_\perp\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{y}_T.\end{aligned} \tag{A.59}$$

Remark A.15. (i) One of the above results, (A.57), is expected from the definition of the gHP$_T$ filter. This is because when $\lambda_T \to \infty$, (9) becomes

$$\min_{\boldsymbol{x}_T\in\mathbb{S}(\boldsymbol{\Pi}_T)} \|\boldsymbol{y}_n - \boldsymbol{S}\boldsymbol{x}_T\|^2, \tag{A.60}$$

where $\mathbb{S}(\boldsymbol{\Pi}_T)$ denotes the column space of $\boldsymbol{\Pi}_T$. Given $\boldsymbol{S}\boldsymbol{\Pi}_T = \boldsymbol{\Pi}_n$, by letting $\boldsymbol{x}_T = \boldsymbol{\Pi}_T\boldsymbol{\beta}$, (A.60) is represented as

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^2} \|\boldsymbol{y}_n - \boldsymbol{\Pi}_n\boldsymbol{\beta}\|^2. \tag{A.61}$$

(ii) Likewise, the other result, (A.59), is somewhat expected from the definition of the gHP$_T$ filter. This is because, given $\lambda_T > 0$, the gHP$_T$ filter can be alternatively represented as follows:

$$\min_{\boldsymbol{x}_T \in \mathbb{R}^T} \lambda_T^{-1} \|\boldsymbol{y}_n - \boldsymbol{S}\boldsymbol{x}_T\|^2 + \|\boldsymbol{D}_T \boldsymbol{x}_T\|^2. \tag{A.62}$$

When $\lambda_T^{-1} \to \infty$, (A.62) becomes

$$\min_{\boldsymbol{x}_T \in \mathbb{R}^T} \|\boldsymbol{D}_T \boldsymbol{x}_T\|^2, \quad \text{s.t.}\, \boldsymbol{S}\boldsymbol{x}_T = \boldsymbol{y}_n. \tag{A.63}$$

Given $\boldsymbol{S}\boldsymbol{x}_T = \boldsymbol{y}_n$, $\|\boldsymbol{D}_T \boldsymbol{x}_T\|^2$ can be regarded as a function of $\boldsymbol{S}_\perp \boldsymbol{x}_T$. More precisely, given $\boldsymbol{S}\boldsymbol{x}_T = \boldsymbol{y}_n$, it follows that

$$\begin{aligned} \boldsymbol{D}_T \boldsymbol{x}_T &= \boldsymbol{D}_T \boldsymbol{\Psi}' \boldsymbol{\Psi} \boldsymbol{x}_T = \boldsymbol{D}_T \boldsymbol{S}_\perp' \boldsymbol{S}_\perp \boldsymbol{x}_T + \boldsymbol{D}_T \boldsymbol{S}' \boldsymbol{S} \boldsymbol{x}_T \\ &= \boldsymbol{D}_T \boldsymbol{S}_\perp' \boldsymbol{S}_\perp \boldsymbol{x}_T + \boldsymbol{D}_T \boldsymbol{S}' \boldsymbol{y}_n. \end{aligned}$$

Therefore, instead of (A.63), we can consider the following minimization problem:

$$\min_{\boldsymbol{S}_\perp \boldsymbol{x}_T \in \mathbb{R}^{T-n}} \|\boldsymbol{D}_T \boldsymbol{S}_\perp' \boldsymbol{S}_\perp \boldsymbol{x}_T + \boldsymbol{D}_T \boldsymbol{S}' \boldsymbol{y}_n\|^2. \tag{A.64}$$

The solution of (A.64) is $-(\boldsymbol{S}_\perp \boldsymbol{D}_T' \boldsymbol{D}_T \boldsymbol{S}_\perp')^{-1} \boldsymbol{S}_\perp \boldsymbol{D}_T' \boldsymbol{D}_T \boldsymbol{S}' \boldsymbol{S} \boldsymbol{y}_T$ and, as shown in (A.58), it equals $\lim_{\lambda_T \to 0} \boldsymbol{S}_\perp \widehat{\boldsymbol{x}}_T$.

A.5. Another Form of the gHP$_T$ Filter

Let $\boldsymbol{y}_* \in \mathbb{R}^T$ be the vector such that

$$\begin{bmatrix} \boldsymbol{S}_\perp \\ \boldsymbol{S} \end{bmatrix} \boldsymbol{y}_* = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{y}_n \end{bmatrix}.$$

That is to say, $\boldsymbol{y}_*$ is the column vector such that the missing observations of $\boldsymbol{y}_T \in \mathbb{R}^T$ are replaced by zeros.

Consider the following minimization problem:

$$\min_{\substack{\boldsymbol{v} \in \mathbb{R}^{T-n} \\ \boldsymbol{x}_T \in \mathbb{R}^T}} f_*(\boldsymbol{v}, \boldsymbol{x}_T) = \|\boldsymbol{y}_* + \boldsymbol{S}_\perp' \boldsymbol{v} - \boldsymbol{x}_T\|^2 + \lambda_T \|\boldsymbol{D}_T \boldsymbol{x}_T\|^2. \tag{A.65}$$

We remark that (i) considering (A.65) is inspired by Schlicht (2008), (ii) f_* in (A.65) is a quadratic function, and (iii) $\boldsymbol{S}_\perp' \boldsymbol{v} \in \mathbb{R}^{T-n}$ in (A.65) represents missing observations of $\boldsymbol{y}_T$ and $\boldsymbol{S}_\perp'$ may be regarded as a matrix of dummy variables.

Regarding $f_*(\boldsymbol{v}, \boldsymbol{x}_T)$ in (A.65), we have the following result:

LEMMA A.16. *Under Assumption 1, the Hessian matrix of $f_*(\boldsymbol{v}, \boldsymbol{x}_T)$ in (A.65) is positive definite.*

Proof. The Hessian matrix of $f_*(\boldsymbol{v}, \boldsymbol{x}_T)$ is:

$$2\begin{bmatrix} \boldsymbol{I}_{T-n} & -\boldsymbol{S}_\perp \\ -\boldsymbol{S}_\perp' & \boldsymbol{I}_T + \lambda_T \boldsymbol{D}_T' \boldsymbol{D}_T \end{bmatrix} = 2\left(\begin{bmatrix} -\boldsymbol{S}_\perp \\ \boldsymbol{I}_T \end{bmatrix} [-\boldsymbol{S}_\perp', \boldsymbol{I}_T] + \lambda_T \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{D}_T' \end{bmatrix} [\boldsymbol{0}, \boldsymbol{D}_T] \right).$$

https://doi.org/10.1017/S0266466621000189 Published online by Cambridge University Press

Since it is evidently non-negative definite, we will show that it is nonsingular. Given that $\boldsymbol{I}_T - \boldsymbol{S}_\perp' \boldsymbol{S}_\perp = \boldsymbol{S}'\boldsymbol{S}$ and Lemma 4, we have

$$\begin{vmatrix} \boldsymbol{I}_{T-n} & -\boldsymbol{S}_\perp \\ -\boldsymbol{S}_\perp' & \boldsymbol{I}_T + \lambda_T \boldsymbol{D}_T'\boldsymbol{D}_T \end{vmatrix} = |\boldsymbol{I}_{T-n}||\boldsymbol{I}_T + \lambda_T \boldsymbol{D}_T'\boldsymbol{D}_T - \boldsymbol{S}_\perp'\boldsymbol{S}_\perp| \\ = |\boldsymbol{S}'\boldsymbol{S} + \lambda_T \boldsymbol{D}_T'\boldsymbol{D}_T| > 0,$$

which completes the proof. ∎

PROPOSITION A.17. *Let $\widetilde{\boldsymbol{v}}$ and $\widetilde{\boldsymbol{x}}_T$ denote the solutions of (A.65) such that $f_*(\boldsymbol{v},\boldsymbol{x}_T) \geq f_*(\widetilde{\boldsymbol{v}},\widetilde{\boldsymbol{x}}_T)$. Then, it follows that $\widetilde{\boldsymbol{x}}_T = \widehat{\boldsymbol{x}}_T$ and $\widetilde{\boldsymbol{v}} = \boldsymbol{S}_\perp \widehat{\boldsymbol{x}}_T$.*

Proof. Given that $\boldsymbol{S}_\perp'\boldsymbol{S}_\perp + \boldsymbol{S}'\boldsymbol{S} = \boldsymbol{I}_T$, we have

$$\|\boldsymbol{y}_* + \boldsymbol{S}_\perp'\boldsymbol{v} - \boldsymbol{x}_T\|^2 = \left\| \begin{bmatrix} \boldsymbol{S}_\perp \\ \boldsymbol{S} \end{bmatrix} (\boldsymbol{y}_* + \boldsymbol{S}_\perp'\boldsymbol{v} - \boldsymbol{x}_T) \right\|^2 = \left\| \begin{bmatrix} \boldsymbol{v} - \boldsymbol{S}_\perp \boldsymbol{x}_T \\ \boldsymbol{y}_n - \boldsymbol{S}\boldsymbol{x}_T \end{bmatrix} \right\|^2.$$

Thus, (A.65) may be represented as

$$\min_{\substack{\boldsymbol{v}\in\mathbb{R}^{T-n} \\ \boldsymbol{x}_T\in\mathbb{R}^T}} f_*(\boldsymbol{v},\boldsymbol{x}_T) = \|\boldsymbol{y}_n - \boldsymbol{S}\boldsymbol{x}_T\|^2 + \|\boldsymbol{v} - \boldsymbol{S}_\perp\boldsymbol{x}_T\|^2 + \lambda_T\|\boldsymbol{D}_T\boldsymbol{x}_T\|^2. \tag{A.66}$$

Then, $\widetilde{\boldsymbol{x}}_T$ and $\widetilde{\boldsymbol{v}}$ satisfy the following equations:

$$-\boldsymbol{S}'(\boldsymbol{y}_n - \boldsymbol{S}\widetilde{\boldsymbol{x}}_T) - \boldsymbol{S}_\perp'(\widetilde{\boldsymbol{v}} - \boldsymbol{S}_\perp\widetilde{\boldsymbol{x}}_T) + \lambda_T \boldsymbol{D}_T'\boldsymbol{D}_T\widetilde{\boldsymbol{x}}_T = \boldsymbol{0}, \quad \widetilde{\boldsymbol{v}} - \boldsymbol{S}_\perp\widetilde{\boldsymbol{x}}_T = \boldsymbol{0},$$

which leads to $\widetilde{\boldsymbol{x}}_T = \widehat{\boldsymbol{x}}_T$ and $\widetilde{\boldsymbol{v}} = \boldsymbol{S}_\perp\widehat{\boldsymbol{x}}_T$. ∎

A.6. More Details on Specifying λ_n in (2)

In this subsection, we give more details on specifying λ_n in (2).

Concerning the convex problem, (26) and (27), we have the following result.

LEMMA A.18. *Consider the convex problem given by (26) and (27). (i) There exists a global minimizer. (ii) Denote a global minimizer by $\widehat{\boldsymbol{x}}_n^*$. Then, there exists $\mu \geq 0$ such that:*

(*stationarity*) $\boldsymbol{D}_n'\boldsymbol{D}_n\widehat{\boldsymbol{x}}_n^* - \mu(\boldsymbol{y}_n - \widehat{\boldsymbol{x}}_n^*) = \boldsymbol{0}$,

(*complementary slackness*) $\mu(\|\boldsymbol{y}_n - \widehat{\boldsymbol{x}}_n^*\|^2 - \|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2) = 0$.

Proof. (i) The objective function of the problem, $\|\boldsymbol{D}_n\boldsymbol{x}_n\|^2$, is a convex function over $\mathbb{R}^n$ and thus it is a continuous function. In addition, the corresponding feasible set, $\{\boldsymbol{x}_n \in \mathbb{R}^n \,|\, \|\boldsymbol{y}_n - \boldsymbol{x}_n\|^2 \leq \|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2\}$, is a nonempty, closed, and bounded set. Thus, by the Weierstrass theorem, there exists a global minimizer [see, e.g., Theorem 2.30 of Beck (2014)]. (ii) Both $\|\boldsymbol{D}_n\boldsymbol{x}_n\|^2$ and $\|\boldsymbol{y}_n - \boldsymbol{x}_n\|^2$ are continuously differentiable convex functions over $\mathbb{R}^n$. In addition, given that $\|\boldsymbol{y}_n - \boldsymbol{x}_n\|^2 - \|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2 = -\|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2 < 0$ if $\boldsymbol{x}_n = \boldsymbol{y}_n$, Slater's condition is satisfied. Then, the lemma follows by Theorem 11.13 of Beck (2014). ∎

Given Lemma A.18, we have:

https://doi.org/10.1017/S0266466621000189 Published online by Cambridge University Press

LEMMA A.19. *If* $0 < \|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2 < \boldsymbol{y}_n'\boldsymbol{Q}\boldsymbol{y}_n$, *then (i)* $\boldsymbol{D}_n\widehat{\boldsymbol{x}}_n^* \neq \boldsymbol{0}$, *(ii)* $\mu > 0$, *(iii)* $\|\boldsymbol{y}_n - \widehat{\boldsymbol{x}}_n^*\|^2 = \|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2$, *(iv)* $\mu^{-1} = \frac{\widehat{\boldsymbol{x}}_n^{*\prime}(\boldsymbol{y}_n - \widehat{\boldsymbol{x}}_n^*)}{\|\boldsymbol{D}_n\widehat{\boldsymbol{x}}_n^*\|^2}$, *and (v)* $\widehat{\boldsymbol{x}}_n^* = (\boldsymbol{I}_n + \mu^{-1}\boldsymbol{D}_n'\boldsymbol{D}_n)^{-1}\boldsymbol{y}_n$.

Proof. (i) Given (a) $\widehat{\boldsymbol{x}}_n^*$ is a solution of the convex problem given by (26) and (27), and (b) $\|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2 < \boldsymbol{y}_n'\boldsymbol{Q}\boldsymbol{y}_n$, we have

$$\|\boldsymbol{y}_n - \widehat{\boldsymbol{x}}_n^*\|^2 \leq \|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2 < \boldsymbol{y}_n'\boldsymbol{Q}\boldsymbol{y}_n. \tag{A.67}$$

Suppose that $\widehat{\boldsymbol{x}}_n^*$ belongs to the column space of $\boldsymbol{\Pi}_n$. Then, there exists $\boldsymbol{\eta} \in \mathbb{R}^2$ such that $\widehat{\boldsymbol{x}}_n^* = \boldsymbol{\Pi}_n\boldsymbol{\eta}$ and accordingly, given $\boldsymbol{y}_n'\boldsymbol{Q}\boldsymbol{y}_n = \|\boldsymbol{y}_n - \boldsymbol{\Pi}_n\widehat{\boldsymbol{\beta}}\|^2$, (A.67) becomes

$$\|\boldsymbol{y}_n - \boldsymbol{\Pi}_n\boldsymbol{\eta}\|^2 \leq \|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2 < \|\boldsymbol{y}_n - \boldsymbol{\Pi}_n\widehat{\boldsymbol{\beta}}\|^2,$$

which is a contradiction. Therefore, $\widehat{\boldsymbol{x}}_n^*$ does not belong to the column space of $\boldsymbol{\Pi}_n$, and hence, $\boldsymbol{D}_n\widehat{\boldsymbol{x}}_n^* \neq \boldsymbol{0}$. (ii) Given that $\boldsymbol{D}_n\widehat{\boldsymbol{x}}_n^* \neq \boldsymbol{0}$, premultiplying the stationarity condition by $\widehat{\boldsymbol{x}}_n^{*\prime}$ yields

$$\mu\widehat{\boldsymbol{x}}_n^{*\prime}(\boldsymbol{y}_n - \widehat{\boldsymbol{x}}_n^*) = \|\boldsymbol{D}_n\widehat{\boldsymbol{x}}_n^*\|^2 > 0. \tag{A.68}$$

Given that $\mu \geq 0$, (A.68) leads to $\mu > 0$. (iii) Given that $\mu > 0$, by the complementary slackness condition, we have $\|\boldsymbol{y}_n - \widehat{\boldsymbol{x}}_n^*\|^2 = \|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2$. (iv) Given that $\mu > 0$, by the stationarity condition, we have $\widehat{\boldsymbol{x}}_n^* = (\boldsymbol{I}_n + \mu^{-1}\boldsymbol{D}_n'\boldsymbol{D}_n)^{-1}\boldsymbol{y}_n$. ■

Moreover, we have the following result:

LEMMA A.20. *If* $\boldsymbol{y}_n$ *does not belong to the column space of* $\boldsymbol{\Pi}_n$, *then (i)* $\|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2 > 0$ *and (ii)* $\|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2 < \boldsymbol{y}_n'\boldsymbol{Q}\boldsymbol{y}_n$.

Proof. (i) From (11), we have

$$\boldsymbol{S}'\boldsymbol{y}_n = (\boldsymbol{S}'\boldsymbol{S} + \lambda_T\boldsymbol{D}_T'\boldsymbol{D}_T)\widehat{\boldsymbol{x}}_T = \boldsymbol{S}'\boldsymbol{S}\widehat{\boldsymbol{x}}_T + \lambda_T\boldsymbol{D}_T'\boldsymbol{D}_T\widehat{\boldsymbol{x}}_T.$$

Accordingly, if $\boldsymbol{S}\widehat{\boldsymbol{x}}_T = \boldsymbol{y}_n$, then we have $\lambda_T\boldsymbol{D}_T'\boldsymbol{D}_T\widehat{\boldsymbol{x}}_T = \boldsymbol{0}$. Given that $\lambda_T > 0$ and $\boldsymbol{D}_T'$ is of full column rank, we have $\boldsymbol{D}_T\widehat{\boldsymbol{x}}_T = \boldsymbol{0}$, which implies that there exists $\boldsymbol{\xi} \in \mathbb{R}^2$ such that $\widehat{\boldsymbol{x}}_T = \boldsymbol{\Pi}_T\boldsymbol{\xi}$. Premultiplying $\widehat{\boldsymbol{x}}_T = \boldsymbol{\Pi}_T\boldsymbol{\xi}$ by $\boldsymbol{S}$ leads to

$$\boldsymbol{y}_n = \boldsymbol{S}\widehat{\boldsymbol{x}}_T = \boldsymbol{S}\boldsymbol{\Pi}_T\boldsymbol{\xi} = \boldsymbol{\Pi}_n\boldsymbol{\xi}.$$

Therefore, if $\boldsymbol{y}_n$ does not belong to the column space of $\boldsymbol{\Pi}_n$, then $\boldsymbol{S}\widehat{\boldsymbol{x}}_T \neq \boldsymbol{y}_n$, which implies $\|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2 > 0$. (ii) Suppose that $\widehat{\boldsymbol{x}}_T = \boldsymbol{\Pi}_T\widehat{\boldsymbol{\beta}}$. Then, from (A.29), we obtain

$$(\boldsymbol{S}'\boldsymbol{S} + \lambda_T\boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'(\boldsymbol{y}_n - \boldsymbol{P}\boldsymbol{y}_n) = \boldsymbol{0}. \tag{A.69}$$

Given that $(\boldsymbol{S}'\boldsymbol{S} + \lambda_T\boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'$ is of full column rank, (A.69) leads to $(\boldsymbol{y}_n - \boldsymbol{P}\boldsymbol{y}_n) = \boldsymbol{Q}\boldsymbol{y}_n = \boldsymbol{0}$. Hence, $\boldsymbol{y}_n$ belongs to the column space of $\boldsymbol{\Pi}_n$. Therefore, if $\boldsymbol{y}_n$ does not belong to the column space of $\boldsymbol{\Pi}_n$, it follows that $\widehat{\boldsymbol{x}}_T \neq \boldsymbol{\Pi}_T\widehat{\boldsymbol{\beta}}$, which leads to

$$f_T(\boldsymbol{\Pi}_T\widehat{\boldsymbol{\beta}}) > f_T(\widehat{\boldsymbol{x}}_T) > \|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2. \tag{A.70}$$

The first (resp. second) inequality in (A.70) follows from Proposition 8(i) (resp. $\lambda_T\|\boldsymbol{D}_T\widehat{\boldsymbol{x}}_T\|^2 > 0$). Accordingly, given that $f_T(\boldsymbol{\Pi}_T\widehat{\boldsymbol{\beta}}) = \boldsymbol{y}_n'\boldsymbol{Q}\boldsymbol{y}_n$, we have $\boldsymbol{y}_n'\boldsymbol{Q}\boldsymbol{y}_n > \|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2$. ■

The above lemmata lead to the following result:

PROPOSITION A.21. *If $\boldsymbol{y}_n$ does not belong to the column space of $\boldsymbol{\Pi}_n$, then (i) $\widehat{\boldsymbol{x}}_n^*$ satisfies the equality given by $\|\boldsymbol{y}_n - \widehat{\boldsymbol{x}}_n^*\|^2 = \|\boldsymbol{y}_n - \boldsymbol{S}\widehat{\boldsymbol{x}}_T\|^2$ and (ii) $\widehat{\boldsymbol{x}}_n^*$ equals $\widehat{\boldsymbol{x}}_n$ that is estimated with*

$$\lambda_n = \frac{\widehat{\boldsymbol{x}}_n^{*\prime}(\boldsymbol{y}_n - \widehat{\boldsymbol{x}}_n^*)}{\|\boldsymbol{D}_n\widehat{\boldsymbol{x}}_n^*\|^2}. \tag{A.71}$$

Proof. It follows from Lemmata A.19 and A.20. ■

A.7. Miscellaneous Proofs

A.7.1. *Proof of Lemma 4.* Let $\boldsymbol{\xi} \neq \boldsymbol{0}$ be a T-dimensional column vector. If $\boldsymbol{\xi}$ belongs to the column space of $\boldsymbol{\Pi}_T$, then $\|\boldsymbol{S}\boldsymbol{\xi}\|^2 > 0$, even though $\|\boldsymbol{D}_T\boldsymbol{\xi}\|^2 = 0$. This is because, under Assumption 1, $\boldsymbol{S}\boldsymbol{\Pi}_T$ is of full column rank. Otherwise, $\|\boldsymbol{D}_T\boldsymbol{\xi}\|^2 > 0$ and $\|\boldsymbol{S}\boldsymbol{\xi}\|^2 \geq 0$. Therefore, in both cases, $\boldsymbol{\xi}'(\boldsymbol{S}'\boldsymbol{S} + \lambda_T\boldsymbol{D}_T'\boldsymbol{D}_T)\boldsymbol{\xi} = \|\boldsymbol{S}\boldsymbol{\xi}\|^2 + \lambda_T\|\boldsymbol{D}_T\boldsymbol{\xi}\|^2 > 0$, which indicates that $2(\boldsymbol{S}'\boldsymbol{S} + \lambda_T\boldsymbol{D}_T'\boldsymbol{D}_T)$ is positive definite.

A.7.2. *Proof of Proposition 7.* Let $\boldsymbol{e}_1$ (resp. $\boldsymbol{e}_T$) be the first (resp. last) column of $\boldsymbol{I}_T$. Then, given that only y_1 and y_T are observable, it follows that $\boldsymbol{S} = [\boldsymbol{e}_1, \boldsymbol{e}_T]'$ and accordingly we have

$$\boldsymbol{\Pi}_n = \boldsymbol{S}\boldsymbol{\Pi}_T = \begin{bmatrix} 1 & 1 \\ 1 & T \end{bmatrix}, \quad \boldsymbol{y}_n = \boldsymbol{S}\boldsymbol{y}_T = \begin{bmatrix} y_1 \\ y_T \end{bmatrix}.$$

Given (A.29) and $\boldsymbol{P} = \boldsymbol{\Pi}_n(\boldsymbol{\Pi}_n'\boldsymbol{\Pi}_n)^{-1}\boldsymbol{\Pi}_n' = \boldsymbol{I}_2$ when $|\boldsymbol{\Pi}_n| \neq 0$, we have

$$\widehat{\boldsymbol{x}}_T = \boldsymbol{\Pi}_T\widehat{\boldsymbol{\beta}} + (\boldsymbol{S}'\boldsymbol{S} + \lambda_T\boldsymbol{D}_T'\boldsymbol{D}_T)^{-1}\boldsymbol{S}'(\boldsymbol{y}_n - \boldsymbol{P}\boldsymbol{y}_n) = \boldsymbol{\Pi}_T\widehat{\boldsymbol{\beta}}. \tag{A.72}$$

In addition, again given $|\boldsymbol{\Pi}_n| \neq 0$, we have

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{\Pi}_n'\boldsymbol{\Pi}_n)^{-1}\boldsymbol{\Pi}_n'\boldsymbol{y}_n = \boldsymbol{\Pi}_n^{-1}\boldsymbol{y}_n = \begin{bmatrix} 1 & 1 \\ 1 & T \end{bmatrix}^{-1} \begin{bmatrix} y_1 \\ y_T \end{bmatrix} = \begin{bmatrix} \frac{Ty_1 - y_T}{T-1} \\ \frac{y_T - y_1}{T-1} \end{bmatrix}. \tag{A.73}$$

Substituting (A.73) into (A.72) completes the proof.

A.7.3. *Alternative Proof of (A.20).* By premultiplying (8) by $\boldsymbol{D}_n(\boldsymbol{I}_n + \lambda_n\boldsymbol{D}_n'\boldsymbol{D}_n)$, we have

$$\boldsymbol{D}_n(\boldsymbol{I}_n + \lambda_n\boldsymbol{D}_n'\boldsymbol{D}_n)\widehat{\boldsymbol{x}}_n = (\boldsymbol{I}_{n-2} + \lambda_n\boldsymbol{D}_n\boldsymbol{D}_n')\boldsymbol{D}_n\widehat{\boldsymbol{x}}_n = \boldsymbol{D}_n\boldsymbol{y}_n.$$

In addition, premultiplying the above equation by $-\lambda_n\boldsymbol{D}_n'(\boldsymbol{I}_{n-2} + \lambda_n\boldsymbol{D}_n\boldsymbol{D}_n')^{-1}$ leads to $-\lambda_n\boldsymbol{D}_n'\boldsymbol{D}_n\widehat{\boldsymbol{x}}_n = -\lambda_n\boldsymbol{D}_n'(\boldsymbol{I}_{n-2} + \lambda_n\boldsymbol{D}_n\boldsymbol{D}_n')^{-1}\boldsymbol{D}_n\boldsymbol{y}_n$. Finally, by adding $\boldsymbol{y}_n$ to the above equation and taking $\widehat{\boldsymbol{x}}_n = \boldsymbol{y}_n - \lambda_n\boldsymbol{D}_n'\boldsymbol{D}_n\widehat{\boldsymbol{x}}_n$ into account, we obtain $\widehat{\boldsymbol{x}}_n = (\boldsymbol{I}_n + \lambda_n\boldsymbol{D}_n'\boldsymbol{D}_n)^{-1}\boldsymbol{y}_n = \{\boldsymbol{I}_n - \boldsymbol{D}_n'(\boldsymbol{D}_n\boldsymbol{D}_n' + \lambda_n^{-1}\boldsymbol{I}_{n-2})^{-1}\boldsymbol{D}_n\}\boldsymbol{y}_n$.

A.7.4. *Proof of (35).* From Proposition A.11, (A.41), and (A.42), it follows that

$$\boldsymbol{S}_\perp\widehat{\boldsymbol{x}}_T = -(\boldsymbol{S}_\perp\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{S}_\perp')^{-1}\boldsymbol{S}_\perp\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{S}'\boldsymbol{S}\widehat{\boldsymbol{x}}_T.$$

https://doi.org/10.1017/S0266466621000189 Published online by Cambridge University Press

When $\boldsymbol{S}_{\perp}$ equals $[0,0,1,0,\ldots,0]$, which is the third row of $\boldsymbol{I}_T$, $\boldsymbol{S}_{\perp}\boldsymbol{D}_T'\boldsymbol{D}_T$ is the third row of $\boldsymbol{D}_T'\boldsymbol{D}_T$. Given (19), explicitly it equals $[1,-4,6,-4,1,0,\ldots,0]$. Then, we have

$$\boldsymbol{S}_{\perp}\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{S}_{\perp}' = 6, \quad \boldsymbol{S}_{\perp}\boldsymbol{D}_T'\boldsymbol{D}_T\boldsymbol{S}' = [1,-4,-4,1,0,\ldots,0].$$

In addition, $\boldsymbol{S}\widehat{\boldsymbol{x}}_T = [\widehat{x}_{T,1}, \widehat{x}_{T,2}, \widehat{x}_{T,4} \ldots, \widehat{x}_{T,T}]'$. Combining these results yields (35).

SUPPLEMENTARY MATERIAL

To view the supplementary material for this article, please visit https://doi.org/10.1017/S0266466621000189

REFERENCES

Beck, A. (2014) *Introduction to Nonlinear Optimization Theory, Algorithms, and Applications with MATLAB* (Society for Industrial and Applied Mathematics).

Cornea-Madeira, A. (2017) The explicit formula for the Hodrick–Prescott filter in a finite sample. *Review of Economics and Statistics* 99(2), 314–318.

CVX Research, Inc. (2011) *CVX: Matlab Software for Disciplined Convex Programming, Version 2.0*. URL: http://cvxr.com/cvx.

de Jong, R. M. & N. Sakarya (2016) The econometrics of the Hodrick–Prescott filter. *Review of Economics and Statistics* 98(2), 310–317.

Gómez, V. (2001) The use of Butterworth filters for trend and cycle estimation in economic time series. *Journal of Business and Economic Statistics* 19(3), 365–373.

Grant, M. & S. Boyd (2008) Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, & H. Kimura (eds.), *Recent Advances in Learning and Control*, pp. 95–110. Springer.

Green, P. J. & B. W. Silverman (1994) *Nonparametric Regression and Generalized Linear Models: A roughness Penalty Approach*. Chapman and Hall/CRC Press.

Hamilton, J. D. (2018) Why you should never use the Hodrick–Prescott filter. *Review of Economics and Statistics* 100, 831–843.

Harvey, A. C. & T. M. Trimbur (2003) General model-based filters for extracting cycles and trends in economic time series. *Review of Economics and Statistics* 85(2), 244–255.

Hastie, T., R. Tibshirani, & J. Friedman (2009) *The Elements of Statistical Learning*, *2nd ed*. Springer.

Hodrick, R. J. & E. C. Prescott (1997) Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking* 29(1), 1–16.

Hoerl, A. E. & R. W. Kennard (1970) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12(1), 55–67.

Iacobucci, A. & A. Noullez (2005) A frequency selective filter for short-length time series. *Computational Economics* 25, 75–102.

King, R. G. & S. T. Rebelo (1993) Low frequency filtering and real business cycles. *Journal of Economic Dynamics and Control* 17(1–2), 207–231.

Mills, T. C. (2003) *Modelling Trends and Cycles in Economic Time Series*. Palgrave Macmillan.

OECD (2012) *OECD System of Composite Leading Indicators*. http://www.oecd.org/std/leading-indicators/41629509.pdf.

Paige, R. L. & A. A. Trindade (2010) The Hodrick–Prescott filter: A special case of penalized spline smoothing. *Electronic Journal of Statistics* 4, 856–874.

Pedersen, T. M. (2001) The Hodrick–Prescott filter, the Slutzky effect, and the distortionary effect of the filters. *Journal of Economic Dynamics and Control* 25(8), 1081–1101.

Phillips, P. C. B. (2005) 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), 1–26.

Phillips, P. C. B. & S. Jin (2020) Business cycles, trend elimination, and the HP filter. *International Economic Review*; doi: https://doi.org/10.1111/iere.12494.

Phillips, P. C. B. & Z. Shi (2020) Boosting: Why you can use the HP filter. *International Economic Review*; doi: https://doi.org/10.1111/iere.12495.

Robinson, G. K. (1991) That BLUP is a good thing: The estimation of random effects. *Statistical Science* 6, 15–51.

Ruppert, D., M. P. Wand & R. J. Carroll (2003) *Semiparametric Regression*. Cambridge University Press.

Sakarya, N. & R. M. de Jong (2020) A property of the Hodrick–Prescott filter and its application. *Econometric Theory* 36(5), 840–870.

Schlicht, E. (2008) Trend extraction from time series with structural breaks and missing observations. *Journal of the Japan Statistical Society* 38(2), 285–292.

Seber, G. A. F. (2008) *A Matrix Handbook for Statisticians*. Wiley.

Weinert, H. L. (2007) Efficient computation for Whittaker–Henderson smoothing. *Computational Statistics and Data Analysis* 52(2), 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(2), 121–128.

Yamada, H. (2017) The Frisch–Waugh–Lovell theorem for the lasso and the ridge regression. *Communications in Statistics – Theory and Methods* 46(21), 10897–10902.

Yamada, H. (2018a) Several least squares problems related to the Hodrick–Prescott filtering. *Communications in Statistics – Theory and Methods* 47(5), 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(2), 102–105.

Yamada, H. (2018c) A trend filtering method closely related to ℓ_1 trend filtering. *Empirical Economics* 55(4), 1413–1423.

Yamada, H. (2020a) A note on Whittaker–Henderson graduation: Bisymmetry of the smoother matrix. *Communications in Statistics – Theory and Methods* 49(7), 1629–1634.

Yamada, H. (2020b) A smoothing method that looks like the Hodrick–Prescott filter. *Econometric Theory* 36(5), 961–981.

Yamada, H. & R. Du (2019) A modification of the Whittaker–Henderson method of graduation. *Communications in Statistics – Theory and Methods* 48(15), 3795–3800.

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(5), 1–10.