Journal of Economic Theory and Econometrics, Vol. 33, No. 3, Sep. 2022, 1–10

Disentangling Trend and Seasonality in Panel Data: An Empirical Analysis of Food Product Sales*

Yongok Choi† Minsoo Jeong‡

Abstract This paper provides a novel approach to extract the trend and seasonal components from panel data consisting of individual entries showing both strong trend and seasonality. For such a data set, the usual principal component analysis generally fails to disentangle them. In the paper, we suggest a methodology to separately identify them using the Hodrick-Prescott filter that is commonly and widely used to remove trends in various economic data. We apply our methodology to a food product sales panel data and show that it effectively disentangles the trend and seasonal components in the data set.

Keywords trend, seasonality, panel data, principal component analysis, Hodrick-Prescott filter

JEL Classification C10, C23, C55, C81

*This paper was completed while Minsoo Jeong visited the Department of Economics, Indiana University, whose hospitality Minsoo Jeong gratefully acknowledges. We thank Yoosoon Chang, Changsik Kim, Joon Y. Park and Kaylyn Y. Park for many useful comments and discussions. We are also grateful to anonymous referees for many helpful suggestions, which have greatly improved the original version of this paper.

†School of Economics, Chung-Ang University.

‡Corresponding author. Department of Economics, Yonsei University Mirae Campus, Yonseidae-gil 1, Wonju, Gangwon, 26493, Republic of Korea. Email: mssjong@yonsei.ac.kr.

Received March 23, 2022, Revised September 1, 2022, Accepted September 28, 2022

1. INTRODUCTION

In applied work, we often find it necessary to identify and extract trend and seasonality for various reasons. Trend and seasonality may be the objects of interest themselves, or they may also be regarded as noises that should be removed or controlled. We have plenty of solutions and tools to identify and extract trend and seasonality in time series, although their performances are not always satisfactory. However, it appears to be much less known how to effectively deal with trend and seasonality in panels.

From the past, there have been many concerns on uncareful deseasonalization. For example, Singleton (1988) noted that seasonally unadjusted data may be preferred depending on cases especially when economic participants behave based on the seasonally unadjusted economic variables. Hansen and Sargent (1993) further examined the cases for which seasonal adjustment is recommended or not when estimating rational expectations models. Ghysels and Perron (1993) showed that seasonally adjusting data can lead to a bias in the unit root test. Recently, it has drawn more attention in terms of the panel data models. Alvarez (2004) considers a dynamic panel regression model to control for the individual seasonalities in panel data of short time length. Ucar and Guler (2010) suggest a modified version of an economic growth model to address the problem of using individually deseasonalized data. Nieto *et al.* (2016) point out the drawbacks of using individually deseasonalized data and propose the use of a dynamic common factor model with an autoregressive structure to overcome them. Investigating common components has been a major issue in panel data analysis as noted in Greenaway-McGrevy *et al.* (2012) and Zhou and Zhang (2016).

In the paper, we propose a methodology to identify and extract common trend and seasonality in a given panel data set. Our methodology relies on the principal component analysis (PCA) and the Hodrick-Prescott (HP) filter – the most prevalently used approaches to identify and extract the common temporal variation in a panel and the trend of a time series, respectively. The reader is referred to Jolliffe (2002) and Hodrick and Prescott (1997) for more detailed discussions on the PCA and the HP filter, respectively. We estimate seasonality in a panel by utilizing the linear combination of the leading principal components (PCs) and the HP filter. For each linear combination of the leading PCs, its trend is identified by the HP filter, and the seasonality is estimated by minimizing the trend of the linear combination and subtracting the minimized trend from the linear combination. The extracted seasonality is then used to obtain the deseasonalized panel, and the HP trend of its leading PC is defined as our estimated trend. The trend thus extracted can be interpreted as the HP trend of the linear combination of the leading PCs that is orthogonal to the extracted seasonality and has the maximum temporal variation.

We illustrate our methodology using food product sales of the Daesang Corporation, one of the leading food product companies in Korea. The data set shows strong seasonality as well as conspicuous trend. As expected, the leading PCs contain both noticeable trend and seasonality. We show how our methodology can be used to disentangle their trend and seasonality. Our methodology works very effectively and disentangles trend and seasonality rather nicely. The rest of the paper is organized as follows. In Section 2, we introduce our methodology to identify and extract trend and seasonality in panels. Section 3 provides empirical illustration of our methodology, and Section 4 concludes the paper.

2. METHODOLOGY

Let the panel data consisting of observations on (Y_{it}) for $i = 1, \ldots, M$ and $t = 1, \ldots, N$ be available, where i and t denote individual and time, respectively. We suppose that (Y_{it}) is given by

$$Y_{it} = \alpha_i T_t + \beta_i S_t + U_{it} \tag{1}$$

for $i = 1, \ldots, M$ and $t = 1, \ldots, N$, where (T_t) and (S_t) represent common trend and seasonality with loadings α_i and β_i for individual i, respectively, and (U_{it}) is the term including the idiosyncratic component representing various individual specific effects in (Y_{it}). In model (1), we assume that there are only two common components in the data: a common trend component which exhibits smooth long-term fluctuations, and a common seasonal component which consists of more volatile repeated fluctuations. Other than these two common components, we assume that there are no common movements residing in all i.[1] To identify the common trend and seasonality in the panel data, we use the criteria used by the PCA and the HP filter, which are the most prevalently used approaches to find common components and the trend, respectively. The HP filter usually assumes that the data consists of only trend and cyclical components, and refers to the remainder other than trend as a cyclical component. However, since the HP filter identifies the trend using a smoothness measure, the HP filter can be also used to differentiate the trend and seasonal components in model (1).

Under some circumstances, we may use only the PCA to estimate trend and seasonality in our model. For instance, if (Y_{it}) has a conspicuous time trend, then we may expect that the leading PC of (Y_{it}) closely approximates trend. By the same token, if a strong seasonal pattern exists in (Y_{it}) with no eminent trend, the leading PC of (Y_{it}) is expected to mainly represent seasonality. Obviously, however, this is not always the case. The data set we will analyze in the paper shows both strong trend and seasonality, and their two leading PCs include both trend and seasonality entangled in a complicated manner. This is demonstrated in Figure 1, where we present two leading PCs obtained from our data set. The PCs of lower orders are not presented, since they are relatively minor in magnitude and do not seem to have any prominent trend or seasonality.

In a panel data set where none of the trend and seasonality components dominates, we may not use the PCA directly to extract any of them. In the paper, we propose a methodology to estimate them using leading PCs. As will be shown in what follows, we extract seasonality (S_t) and trend (T_t) sequentially.

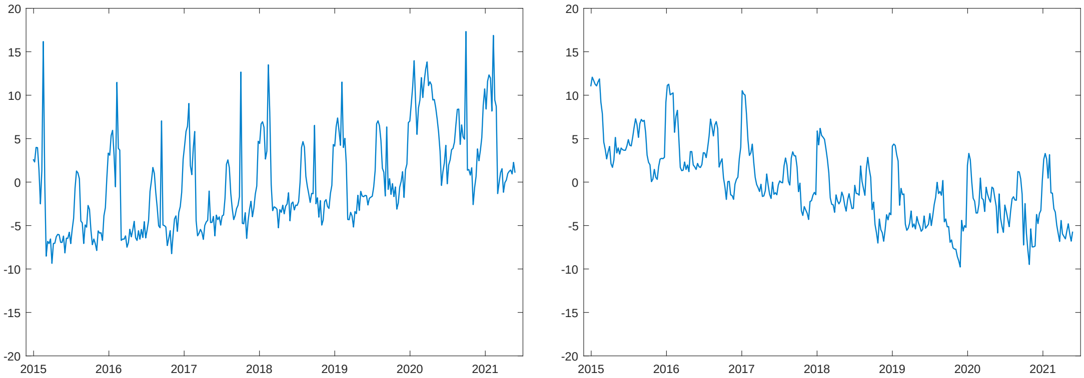
To extract (S_t), we let F_{jt}, $j = 1, \ldots, K$ for $K \geq 2$, be the j-th PC obtained from Y_{it}, and define

$$C_t(\theta) = \sum_{j=1}^{K} \theta_j F_{jt}, \tag{2}$$

where $\theta = (\theta_j)$ is a K-dimensional parameter such that $\sum_{j=1}^{K} \theta_j^2 = 1$. For the choice of K, we should include all the leading PCs containing important information to extract seasonal component correctly. We may utilize the scree plot or the test suggested in Ahn and Horenstein (2013), for

[1] If our assumption does not hold and some common irregular components exist in the data, then our estimated seasonal component $(\hat{S}_t)$ in (5) will include the common irregular components.

Figure 1: Two Leading Principal Components



Note: Presented are two leading PCs. The first and second PCs are given in the left and right panels, respectively.

which we will explain more in Section 3.1. Next, we let

$$R_t(\theta) = \underset{(R_t)}{\text{argmin}} \left(\sum_{t=1}^{N} (C_t(\theta) - R_t)^2 + \lambda \sum_{t=2}^{N-1} \big((R_{t+1} - R_t) - (R_t - R_{t-1}) \big)^2 \right) \tag{3}$$

for $t = 1, \ldots, N$, where λ is a tuning parameter of our choice. For each θ, $\big(R_t(\theta)\big)$ in (3) is the trend extracted by the HP filter from $\big(C_t(\theta)\big)$ with the penalty parameter λ. Once we obtain $\big(R_t(\theta)\big)$ for each θ, we may find the value θ^* of θ given by

$$\theta^* = \underset{\theta}{\text{argmin}} \sum_{t=1}^{N} R_t^2(\theta), \tag{4}$$

and define

$$\hat{S}_t = C_t(\theta^*) - R_t(\theta^*) \tag{5}$$

for $t = 1, \ldots, N$, where $\big(C_t(\theta^*)\big)$ is defined as $\big(C_t(\theta)\big)$ in (2) with $\theta = \theta^*$. We let $(\hat{S}_t)$ be our extracted seasonality. Note that θ^* defined in (4) is the value of θ which yields the HP trend whose sum of squares is smallest, and we use this value of θ as the loading for the leading PCs to obtain our estimate for seasonality (S_t).[2]

We may estimate trend (T_t) using our estimate $(\hat{S}_t)$ for seasonality (S_t). To obtain an estimate $(\hat{T}_t)$ of (T_t), we define

$$Z_{it} = Y_{it} - \hat{\delta}_i \hat{S}_t$$

for $i = 1, \ldots, M$ and $t = 1, \ldots, N$, where $(\hat{S}_t)$ is the estimated seasonality introduced in (5), and $\hat{\delta}_i$ is the OLS estimate for the regression coefficient in the regression of $(Y_{it})_{t=1}^{N}$ on $(\hat{S}_t)$ for each $i = 1, \ldots, M$. Using the leading PC (G_t) of (Z_{it}), we then let

$$(\hat{T}_t) = \underset{(T_t)}{\text{argmin}} \left(\sum_{t=1}^{N} (G_t - T_t)^2 + \lambda \sum_{t=2}^{N-1} \big((T_{t+1} - T_t) - (T_t - T_{t-1}) \big)^2 \right), \tag{6}$$

[2]Ideally, $C_t(\theta^*)$ consists of only the seasonal component and $R_t(\theta^*)$ is negligible because $C_t(\theta^*)$ is a linear combination that yields the smallest trend and $R_t(\theta^*)$ is the trend component of $C_t(\theta^*)$. However, since there may remain some residual trend in $C_t(\theta^*)$, we obtain our final seasonal component by subtracting $R_t(\theta^*)$ from $C_t(\theta^*)$.

where again λ denotes the penalty parameter. Our estimate $(\hat{T}_t)$ of trend (T_t) is thus defined as the HP trend of the leading PC (G_t) of (Z_{it}).[3]

Using $(\hat{S}_t)$ in (5) as our estimate for seasonality is the key novel idea of our method. We find a linear combination of the leading PCs whose trend is minimal, and define that linear combination of the leading PCs as our estimate for seasonality net of its HP trend. The leading PCs effectively summarize common temporal variations in a given panel data, and it is sensible to extract common seasonality as a linear combination of the leading PCs. Moreover, it is very natural and intuitively appealing to use the linear combination minimizing trend to estimate seasonality. To make our idea operational, we use the HP filter to identify and extract trend. As will be demonstrated in the next section, our approach works well and provides a very satisfactory estimate of seasonality (S_t).

The required computation to obtain the estimated seasonality $(\hat{S}_t)$ becomes more convenient if we utilize the linearity of the HP filter. To show this, we let θ, $C(\theta)$ and $R(\theta)$ be the N-dimensional vectors defined from (θ_j), $\big(C_t(\theta)\big)$ and $\big(R_t(\theta)\big)$, respectively, and F be the $N\times K$ matrix whose j-th column is the $(F_{jt})_{t=1}^N$. Then we may write $C(\theta)=F\theta$, and

$$R(\theta)=HF\theta,$$

where $H=(\lambda M+I_N)^{-1}$, M is an N-dimensional square matrix given as

$$M=\begin{pmatrix}
1 & -2 & 1 & 0 & \cdots & & & & & & & & \cdots & 0\\
-2 & 5 & -4 & 1 & 0 & \cdots & & & & & & & \cdots & 0\\
1 & -4 & 6 & -4 & 1 & 0 & \cdots & & & & & & \cdots & 0\\
0 & 1 & -4 & 6 & -4 & 1 & 0 & \cdots & & & & & \cdots & 0\\
\vdots & & & & & & & \ddots & & & & & & \vdots\\
0 & \cdots & & & & & & \cdots & 0 & 1 & -4 & 6 & -4 & 1 & 0\\
0 & \cdots & & & & & & & \cdots & 0 & 1 & -4 & 6 & -4 & 1\\
0 & \cdots & & & & & & & & \cdots & 0 & 1 & -4 & 5 & -2\\
0 & \cdots & & & & & & & & & \cdots & 0 & 1 & -2 & 1
\end{pmatrix},$$

and I_N is the N-dimensional identity matrix (Kim, 2004). Therefore, θ^* is just given by the normalized eigenvector of $F'H'HF$ associated with its smallest eigenvalue. Note that $\hat{S}$ defined from $(\hat{S}_t)$ similarly as above may also be obtained by $\hat{S}=(I-H)F\theta^*$.

As for the trend estimate $(\hat{T}_t)$ obtained in (6), $(\hat{T}_t)$ is the extracted HP trend obtained from the leading PC (G_t) of the deseasonalized panel (Z_{it}). Note that $(Z_{it})_{t=1}^N$ is orthogonal to $(\hat{S}_t)$ for all $i=1,\dots,N$ and (G_t) is defined as a linear combination of $(Z_{it})_{t=1}^N$ across $i=1,\dots,M$. We may therefore see that (G_t) is orthogonal to $(\hat{S}_t)$. In fact, we may define (G_t) as a factor of (Y_{it}) obtained by taking a linear combination of $(Y_{it})_{t=1}^N$ across $i=1,\dots,M$ such that it is orthogonal to $(\hat{S}_t)$ and at the same time has the largest temporal variation.

The procedure to obtain the estimated seasonal and trend components is summarized below.

Step 1 Calculate the K leading PCs of (Y_{it}), and let F be the resulting $N\times K$ matrix.

Step 2 Calculate the HP trend components for each column of F for a given λ. Let R be the resulting $N\times K$ matrix.

[3] For the trend estimation, it is natural to let

$$(\hat{T}_t)=(F_{1t})-(\hat{S}_t) \tag{7}$$

when $K=1$, with which our trend estimate does not coincide. However, the trend estimation in (7) cannot be generalized in the case of $K\geq 2$.

Step 3 Let θ^* be the eigenvector associated with the smallest eigenvalue of $R'R$.

Step 4 Let $\hat{S} = F\theta^* - R\theta^*$. Then, $\hat{S}$ is our estimated seasonal component.

Step 5 Regress each time series (for each i) on $\hat{S}$ and obtain the residuals. Let Z be the resulting $N \times M$ matrix of residuals.

Step 6 Let $\hat{T}$ be the leading PC of Z. Then, $\hat{T}$ is our estimated trend component.

3. EMPIRICAL ILLUSTRATION

3.1. DATA

To illustrate our methodology, we analyze weekly sales of the Daesang Corporation for the period of the first week of January 2015 to the fourth week of May 2021. Daesang Corporation is a general food company located in Korea, which produces, packages and distributes a wide variety of food and ingredient products. It was the second largest among general food companies in Korea in terms of total sales in 2020.[4] Our data set includes the number of units of food products sold to the retailers in a week, from Monday to Sunday, for each product. The items included in the data set are very heterogeneous. There are total 4,126 products that are produced by the company, out of which we select 497 products, excluding items discontinued or newly listed during the period of our study. Our data set shows some strong and complex seasonal patterns, as well as an obvious overall upward trend. There are two reasons for its strong and complex seasonality. The company deals with many seasonal products whose sales mainly or often exclusively occur in specific seasons. Moreover, there are conspicuous calendar effects in the sales of the company. The sales of many products the company carries show big surges in a couple of weeks before some traditional holiday weeks, followed by abrupt drops during the actual weeks of the holidays.

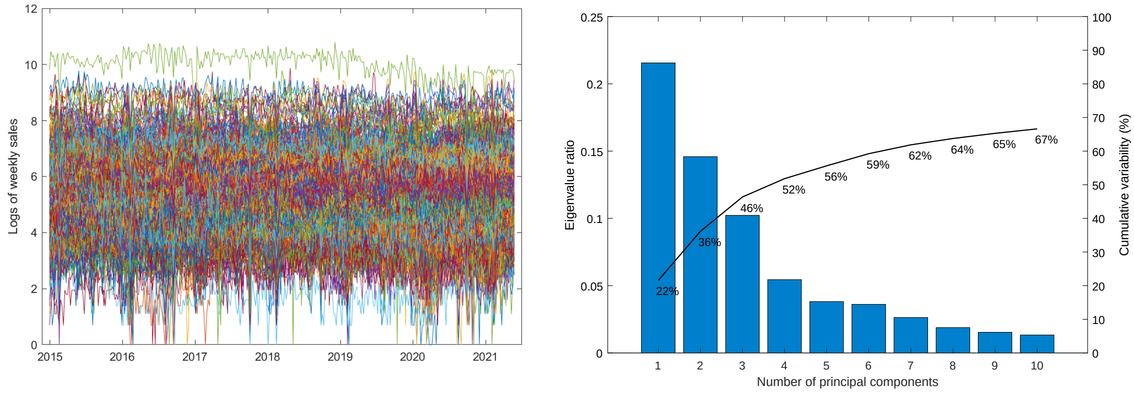
Figure 2 presents the logs of weekly sales, and the proportions of the total temporal variation explained by the leading 10 PCs of the demeaned logs of weekly sales. The first three PCs are eminent. They explain 22%, 14% and 10% of the total variation, respectively, and 46% of the total variation jointly. The test of Ahn and Horenstein (2013), both of their ER and GR statistics, also yields the presence of three factors unambiguously. Therefore, we may use up to three leading PCs to identify and extract common trend and seasonality. In our analysis, however, we only use the two leading PCs, since our main objective is to disentangle trend and seasonality and the third PC has no meaningful seasonality. As will be shown later, the detrended third PC yields a spectral density that has a main peak at an obscure frequency. This is in sharp contrast with the seasonality extracted from the two leading PCs whose spectrum has an unambiguously strong seasonal pattern. In fact, the third PC indeed has neither any notable seasonality nor conspicuous trend. The trend of the third PC is almost negligible, and the estimated spectral density of the detrended third PC does not have any interpretable seasonal pattern.

3.2. EXTRACTED SEASONALITY AND TREND

The extracted seasonality and trend are presented in Figure 3. For the presented results, we use the two leading PCs, and set the penalty parameter λ at $\lambda = 6.25 \times 52^4$ as suggested by Ravn and

[4]See *Food and food service statistics 2020*, published by the Ministry of Agriculture, Food and Rural Affairs, and Korea Agro-Fisheries and Food Trade Corporation (in Korean).

Figure 2: Logs of Weekly Sales and Scree Plots



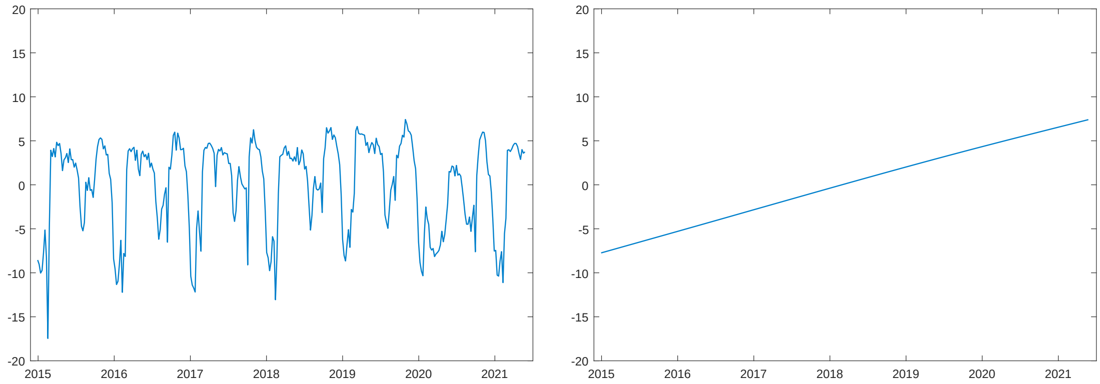
Note: The logs of weekly sales are presented in the left panel. In the right panel, the proportions and their cumulative values of the total temporal variation explained by the 10 leading principal components are presented.

Uhlig (2002) for weekly observations. The results here demonstrate that our methodology is very effective in disentangling trend and seasonality in our panel data including both strong seasonal and trending patterns as demonstrated earlier in Figure 1. The extracted seasonality has no trend, and exhibits roughly the seasonal pattern that existed commonly in the two leading PCs. The extracted trend is virtually linear, which suggests a linear growth in the quantities of product sales. As expected, it does not show any seasonal pattern. Our methodology to identify and extract trend seems to annihilate all seasonal fluctuations in our data set rather completely.

The spectral density of the extracted seasonality is provided in Figure 4. Here we also present the spectral density of the detrended third PC to check whether it has any left-over seasonality. This is necessary, since we only use the two leading PCs to extract seasonality as discussed earlier. The spectral density estimates of the extracted seasonality and the detrended third PC are computed using the signal processing toolbox in Matlab, which utilizes the Kaiser window.[5] As expected, the estimated spectral density of the extracted seasonality has sharp picks at all major seasonal frequencies, whereas that of the detrended third PC shows no such sign of the presence of seasonality. The spectral density of the extracted seasonality has major peaks at the frequencies $2\pi/52, 2\pi/26, 2\pi/17$ and $2\pi/13$ corresponding to the cycles of $52, 26, 17$ and 13 weeks, which are approximately one year, half-year, 4 months and 3 months, respectively. Amongst them, the half-year cycle is dominant. It also has other minor peaks at the frequencies $2\pi/7.4, 2\pi/5.8, 2\pi/4.72$ and $2\pi/4$ corresponding to the cycles of $7.4, 5.8, 4.72$ and 4 weeks, which are roughly one or one and half months. In contrast, the spectral density of the detrended third PC has a main peak at $2\pi/10.4$ representing the cycle of 10.4 weeks, which is hard to interpret as any seasonal cycle. It also has a wide spread of nonnegligible spectral masses in high frequency regions. Although there are some notable peaks at some seasonal frequencies, it seems evident that overall the third PC is devoid of any meaningful seasonality.

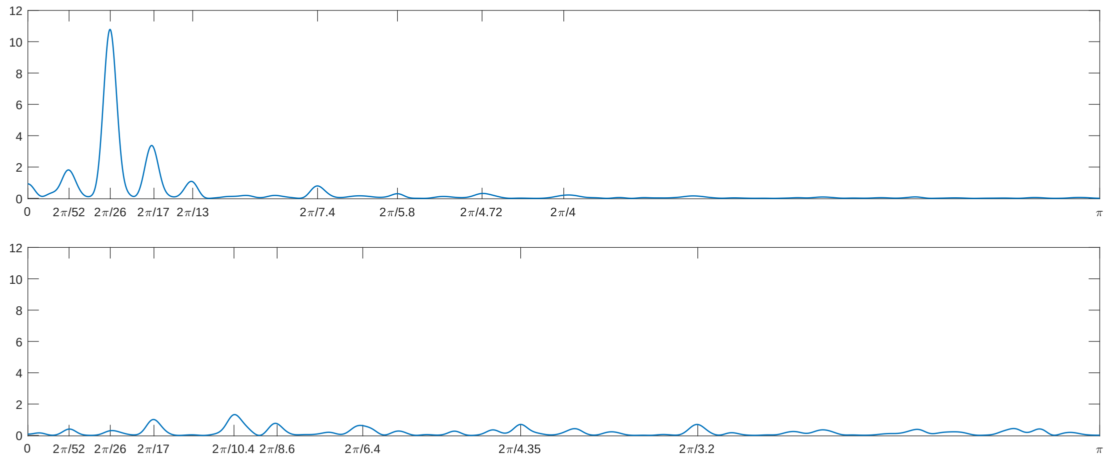
[5]See, for example, Oppenheim and Schafer (2009).

Figure 3: Extracted Seasonality and Trend



Note: The extracted seasonality and trend are presented in the left and right panels, respectively.

Figure 4: Estimated Spectral Densities



Note: The logs of weekly sales are presented in the left panel. In the right panel, the proportions and their cumulative values of the total temporal variation explained by the 10 leading principal components are presented.

4. CONCLUSION

In this paper, we develop a methodology to disentangle trend and seasonality in a panel data set, which has strong seasonality as well as noticeable trend. For such a data set, leading PCs are expected to show both trending behavior and seasonal fluctuations, and we need to further analyze them to separate them from each other. Our methodology relies on the PCA and the HP filter. Both are used extensively in the economic data analysis. We use the PCA to find the common temporal fluctuations in a panel, and apply the HP filter to find the linear combination of them which yields minimal trend. The resulting linear combination of leading PCs, net of its HP trend, is proposed as the extracted seasonality. The extracted trend is defined sequentially as the HP trend of the leading PC of the given data set after being deseasonalized using the extracted seasonality. We use the panel data set of food product sales and show that our methodology truly works well in disentangling trend and seasonality.

REFERENCES

Ahn, S. C. and A.R. Horenstein (2013). "Eigenvalue ratio test for the number of factors," *Econometrica* 81, 1203-1227.

Alvarez, J. (2004). "Dynamics and seasonality in quarterly panel data: An analysis of earnings mobility in Spain," *Journal of Business and Economic Statistics* 22, 443-456.

Ghysels, E. and P. Perron (1993). "The effect of seasonal adjustment filters on tests for a unit root," *Journal of Econometrics* 55(1-2), 57-98.

Greenaway-McGrevy, R., C. Han, and D. Sul (2012). "Standardization and estimation of the number of factors for panel data," *Journal of Economic Theory and Econometrics* 23(2), 79-88.

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

Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer-Verlag, New York.

Kim, H. (2004). "Hodrick-Prescott filter," *Working Paper.*

Nieto, F. H., D. Peña, and D. Saboyá (2016). "Common seasonality in multivariate time series," *Statistica Sinica*, 1389-1410.

Oppenheim, A. V. and R. W. Schafer (2009). *Discrete-Time Signal Processing*, Prentice-Hall, Englewood Cliffs, New Jersey.

Ravn, M. O. and H. Uhlig (2002). "On adjusting the Hodrick-Prescott filter for the frequency of observations," *Review of Economics and Statistics* 84, 371-380.

Hansen, L. P. and T. J. Sargent (1993). "Seasonality and approximation errors in rational expectations models," *Journal of Econometrics* 55(1-2), 21-55.

Singleton, K. J. (1988). "Econometric issues in the analysis of equilibrium business cycle models," *Journal of Monetary Economics* 21(2-3), 361-386.

Ucar, N. and H. Guler (2010). "Testing stochastic income convergence in seasonal heterogeneous panels," *Economic Modelling* 27, 422-431.

Zhou, Q. and Y. Zhang (2016). "Common correlated effects estimation of unbalanced panel data models with cross-sectional dependence," *Journal of Economic Theory and Econometrics* 27(4), 25-45.