Extracting Air Quality Indicator Trends Using Time Series Butterworth Filters

Giuliana Passamani*
Department of Economics and Management, University of Trento, Trento, Italy
e-mail address: giuliana.passamani@unitn.it

Paola Masotti
Department of Economics and Management, University of Trento, Trento, Italy
e-mail address: paola.masotti@unitn.it

Abstract

Air quality control is a fundamental issue, given the possible effects of exposure to air pollution on human health. Understanding the temporal variation mechanism of the aerosol and gaseous pollutants is therefore an important question. The approach we follow starts with the estimation of a latent common factor that we consider a pollution indicator from which we extract its trend component. The data set for the empirical analysis is made up of weekly time series observations covering a period of twelve years and recorded at four different monitoring sites in an Alpine Italian province[1]. In particular, the main aim is the proposal of a procedure that can be used to extract a meaningful latent dynamic stochastic process supposed to represent the generating process of air pollution for each site and to apply a model-based low pass Butterworth filter, used to isolate the low frequency component in a series. The overall main aim is to assess whether any improvement in the pollution level has been observed during the period of observation. The results show that some improvement in the level of air pollution has been achieved even if there are evident differences among the monitoring sites.

Keywords: air quality, dynamic-factor model, unobserved component model, Butterworth filter.

1. Introduction

Urban air pollution is a severe problem in many countries for the well documented adverse effects on human health, it causes a wide spectrum of health effects ranging from eye irritation to death (Pascal et al., 2013; EC, 2013). In the European Union (EU) a number of regulatory actions on ambient air quality allowed to significantly reduce air pollutant emissions and impacts nevertheless health and environmental impacts of air pollution are still significant "In 2010, annual premature mortalities amounted to over 400000 and 62% of the EU area was exposed to eutrophication" (EC, 2013). Furthermore, the Commission has estimated that in 2010 external costs of air pollution health impacts were in the range 330-940 € bn (depending on whether the low or high range of possible impact valuations is considered) (EC, 2013). The EU has adopted a number of Ambient Air Quality Directives (AAQD) which regulate air pollutant emissions and their concentrations in the ambient air as well as minimum standards for assessing and managing air quality in the EU Member States.

In this paper the ambient air quality of the Province of Trento (Italy) is assessed considering four pollutants considered the most dangerous for human health in the European Union (EEA, 2014). Particulate Matter (PM) pollution exposure causes respiratory disease, cardiovascular disease, asthma, premature death (in the EU are estimated to be over 450000) (EEA, 2014), anthropogenic sources of PM include combustion processes, mainly domestic solid fuel combustion, industrial activities and traffic. Ozone (O_3) causes breathing problems, asthma, and lung diseases. Short-term exposure to high levels of O_3 has been related with an increase of the mortality and hospital recovery for respiratory and cardiovascular diseases (EC, 2013). Tropospheric ozone derives from chemical reactions among primary pollutants (NO_X, methane (CH_4) and non-methane volatile organic compounds (NMVOC), and (CO)) with solar radiation. Nitrogen oxides (NO_2 and NO) cause serious respiratory effects

[1] The data set for the empirical analysis has been provided by "Agenzia Provinciale per la Protezione dell'Ambiente (APPA)" of the Province of Trento (Italy).

especially for children, the elderly, and asthmatics. Sources of NOx are high temperature combustion processes such as those occurring in car engines and power plants.
Data have been collected according to Directive 2008/50/EC. They consist of time series observations on four pollutants: PM_{10}, NO_2, NO and O_3. They are obtained from continuous measurements of each pollutant, the unit of measurement is μg m^{-3}. The original daily observations correspond to the twelve-years' period starting in January 2002 and ending in December 2013. For averaging out some outliers that could be present in daily observations, the data set has been transformed in weekly observations. The four monitoring sites we consider are: (i) Trento Parco S. Chiara, a residential area of the main city; (ii) Rovereto Largo Posta, in its the city center; (iii) Borgo Valsugana, a small town; (iv) Riva del Garda, a touristic town at Garda Lake. Figure 1 shows the box-plot representations of the observations: as we can observe, there are some differences in the median value of PM_{10}, NO, NO_2 and O_3 across the sites and their distribution is asymmetric, with some outliers for PM_{10} and NO.

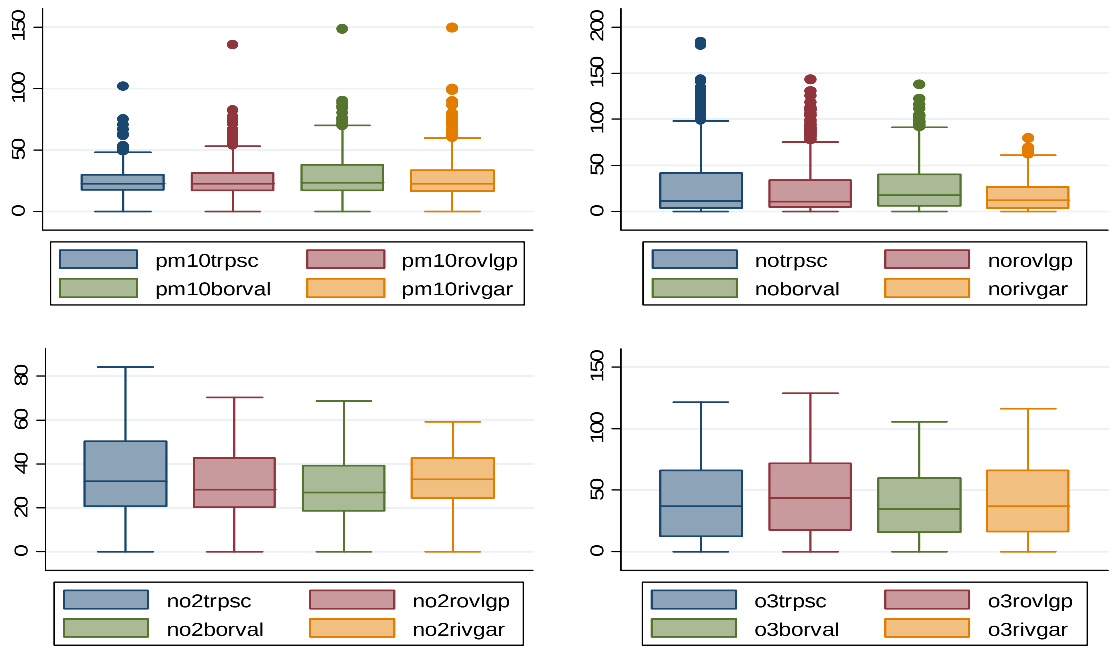

Figure 1: Box-plot of the weekly time series for the four pollutants, PM_{10}, NO, NO_2 and O_3.

The paper is organized as follows. In Section 2 we briefly discuss the dynamic factor model used to estimate for each site the common latent factor considered as a pollution indicator. Then we present the procedure proposed for extracting the trend component of the pollution indicator. In Section 3 we describe the results and in Section 4 we conclude the paper.

2. The methodological approach

In order to investigate whether a latent common pollution indicator can be detected among aerosol and gaseous pollutants for each monitoring site, we make use of dynamic factor analysis, as in Forni et al. (2000)[2] and Fontanella et al. (2007)[3]. The dynamic factor model considered is the following:

---

[2]Following the work of Stock and Watson (1989), Forni et al. (2000) use dynamic factor analysis to compute a coincident indicator for the euro zone countries, constructed as the weighted average of the common component estimated for each country.
[3] Assuming that latent effects underlying the fluctuation of observations on pollutants can be estimated, Fontanella et al. (2007) develop a dynamic structural equation model in state-space form for the analysis of environmental pollution in the Milan district, a model which is a generalization of the dynamic factor analysis proposed by Forni et al. (2000).

$$ (1) \qquad y_t = \gamma_0 + \gamma f_t + \boldsymbol{u}_t, \quad f_t = \alpha f_{t-1} + \nu_t, \quad \boldsymbol{u}_t = \boldsymbol{\Phi} \boldsymbol{u}_{t-1} + \varepsilon_t , $$

where $\mathbf{y}_t$ represents the (4×1) vector of observed variables and f_t the latent common factor. $\mathbf{u}_t$ and $\boldsymbol{\varepsilon}_t$ are (4×1) vectors of disturbances, ν_t is a scalar disturbance, $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}$ are (4×1) vectors containing the constant level parameters and the unknown factor loadings, α is the autoregressive factor parameter[4] and $\boldsymbol{\Phi}$ is a (4×4) matrix of autoregressive parameters. Model (1) can be considered as a dynamic factor model with vector autoregressive errors, where the conditional mean of the factor is assumed to vary over time according to an AR(1). It is estimated using a maximum likelihood approach implemented by writing the model in state space form and by using the stationary Kalman filter and De Jong diffuse Kalman filter for calculating the log likelihood[5]. Once the model has been estimated, we use it in order to predict the latent variable $\hat{f}_t$. The prediction method estimates the states at each time by a Kalman smoother and using all the sample information.
For extracting the trend component from the predicted factor $\hat{f}_t$ we consider the model-based procedure suggested in Harvey and Trimbur (2003). They move from unobserved components, or structural, time series models, where the parameters are estimated by maximum likelihood and the components are obtained by smoothing algorithms, and extend this class to particular types of models in order to show that the optimal estimates of trends and periodic components are implicitly given Butterworth filters. These are low frequency pass filters and mid-range frequency band pass filters specifically designed to separate out the trend and the periodic component in a series. Formally, the procedure assumes that the time series of predicted values is generated by the following unobserved components time series model:

$$ (2) \qquad \hat{f}_t = \mu_{m,t} + \psi_{n,t} + \eta_t , $$

where the trend component

$$ (3) \qquad \mu_{1,t} = \mu_{1,t-1} + \zeta_t , \; \mu_{i,t} = \mu_{i,t-1} + \mu_{i-1,t} , \; i = 2, \; \ldots, m $$

represents an *m-th* order stochastic trend, which is a random walk for m=1 and an integrated random walk for m=2. The periodic component

$$ (4) \qquad \psi_{1,t} = c(L)\kappa_t , \; \psi_{i,t} = c(L)\psi_{i-1,t} , \; i = 2, \; \ldots, n $$

represents an *n-th* order stochastic periodic component following specific dynamics. η_t, ζ_t and κ_t are white noise errors[6].
As shown in Harvey and Trimbur (2003) the optimal estimator of the trend (3) is the low pass Butterworth filter of order (m, n). The aim of time series filters is to transform the observed series into a new series for which the spectral density function is zero for unwanted frequencies and equal to the spectral density function for desired frequencies. Therefore, if we are interested in the low frequency component of a time series, the sample spectral density function of the filtered series should show it.

[4] In the work of Zuur et al. (2003) or in the recent work of Yu et al. (2015) where dynamic factor analysis has been used, this autoregressive parameter has been set equal to one to detect the common trend component of the observed response variables, following the work of Harvey (1989) on structural time series models.
[5] The number of autoregressive lags in the model specification and the covariance structure of the errors have been decided on the basis of some preliminary analysis and in order to avoid convergence problems that can arise in the optimization phase of the likelihood, when estimating the model. Convergence problems are mentioned in Stata 11 Manual.
[6] The model considered in (2), (3) and(4) is similar to the seasonal specific local linear trend model of Proietti (2004), where an observed periodic time series is uniquely decomposed into a non periodic component and a periodic one.

4. Results

The four pollutants we consider make the vector $\mathbf{y}_t' = (PM_{10},\ NO,\ NO_2,\ O_3)_t$. Using (1) we model them as linear function of an unobserved factor f_t generated by a first-order autoregressive process and we allow the errors for the observed variables to be first-order autocorrelated. Its ML estimation[7] has given interesting results: for each site the estimated parameters of PM_{10}, NO, NO_2 and O_3 on the common factor are largely significant and the estimated autoregressive parameter for the process generating f_t indicates that the latent factor is quite persistent.

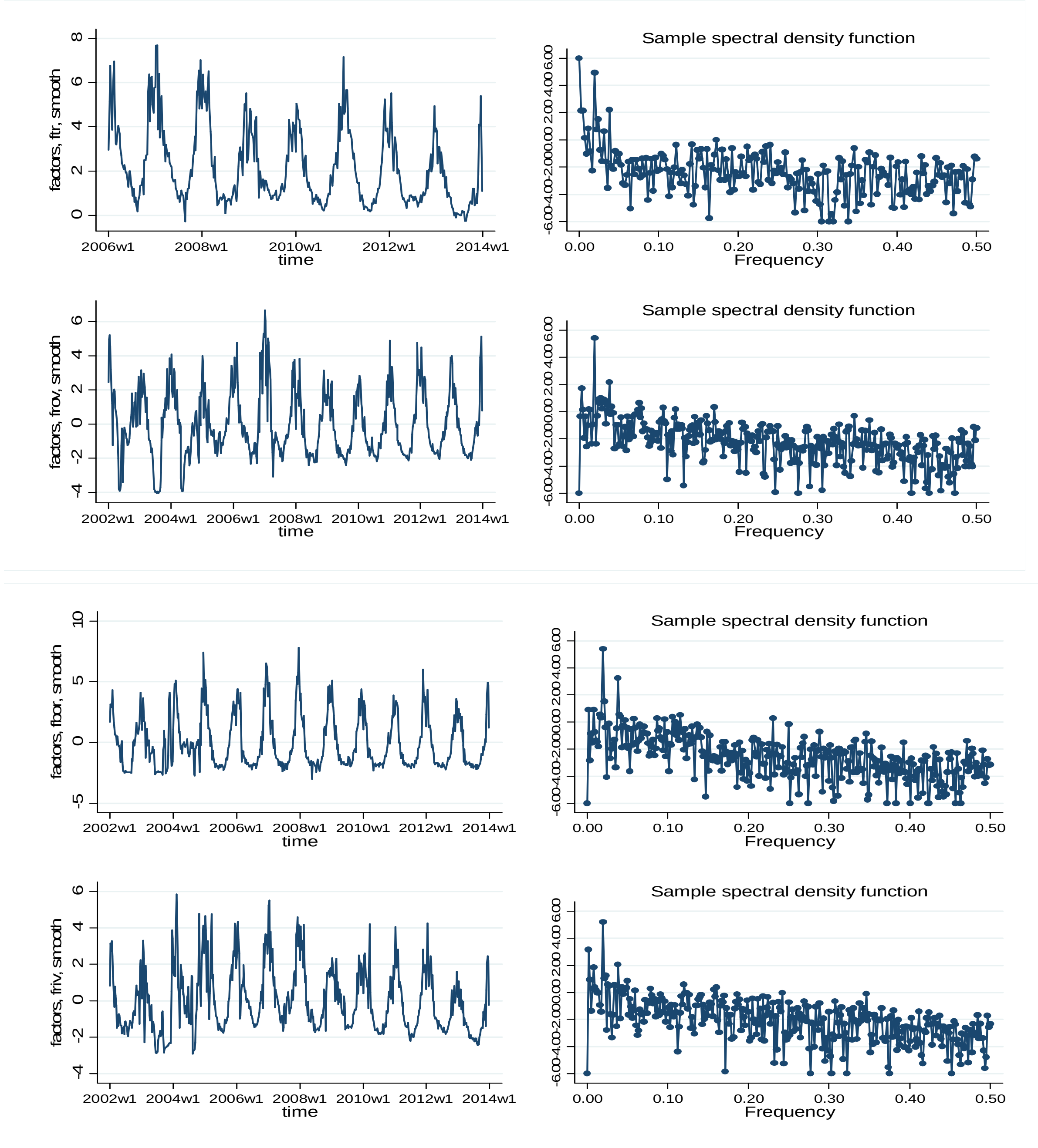

Figure 2: The estimated latent factor and its sample spectral density function.

[7] The ML estimates are not reported for reasons of space, but are available upon request.

In Figure 2 we represent the four predicted factor $\hat{f}_t$ for each monitoring site[8], as well as its sample spectral density function. As we can notice: the factors are characterised by a clear periodic behaviour; their periodograms reveal stochastic components at any frequency; it's very hard to figure out the trend component.

In order to extract the trend component of the latent factor representing the pollution indicator, we filter the time series $\hat{f}_t$ using a Butterworth filter of order 2 and defined to remove the stochastic periodic components at frequencies greater than 0.01, which correspond to yearly periods. As we can see from Figure 3 the periodograms of the resulting filtered series show that the filter has removed almost all the stochastic components and has only left the components corresponding to very low frequencies.

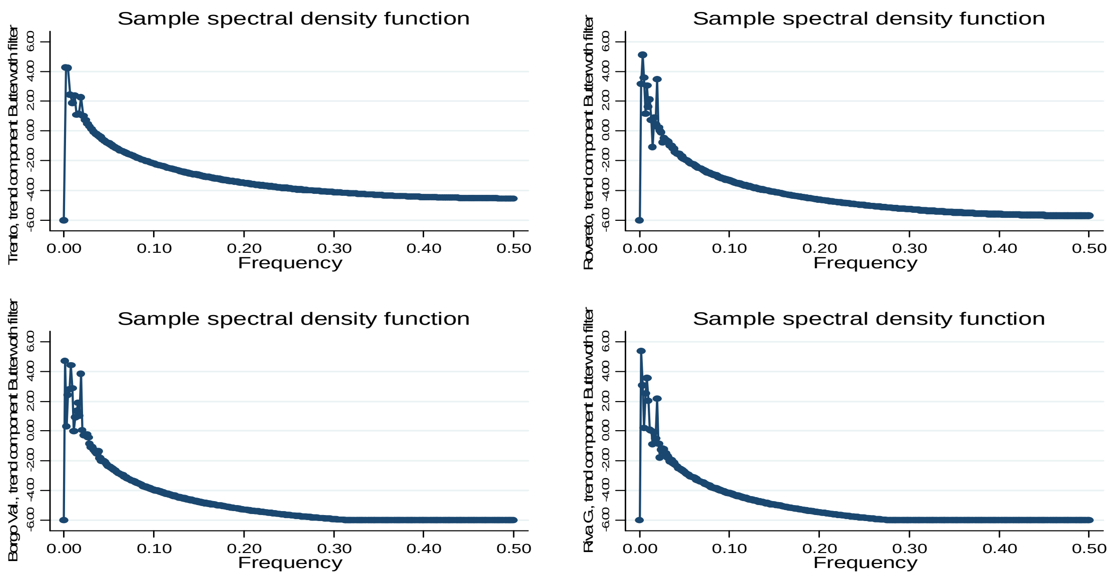

Figure 3: The sample spectral density function for the filtered trend component.

This is a reasonable choice since it leaves a trend component which is no too flat and still shows some movements across the years of observations as we can see in Figure 4. If we consider this trend component as a pollution indicator, we can say that the evolution pattern of this indicator reveals some clear different behaviours, showing some improvement in air quality in recent years, even if some worsening could be observed in every site[9].

5. Conclusions

The air quality indicator obtained using the proposed trend extraction procedure is able to summarise a complex situation in a single variable whose evolution can be compared in time and in space and the results we have obtained are interesting with respect to the aim. In our analysis we haven’t considered the possible effects on pollution of meteorological variables, because we are not actually analysing how climate changes influence the level of pollution, as they are not mainly under control. The approach followed is rather simple and enough flexible for modelling an unobserved variable as

---

[8] For the first monitoring site we estimate the factor starting from the first week in 2006, because of missing observations on PM_{10}.

[9] Because of the Butterworth filter, the filtered time series miss observations at the beginning and at the end of the series, in the sense that the first and the last filtered observations are given by a truncated filter. We should have backcasted and forecasted the input series before filtering it, if we would have reliable values for the most recent period. .

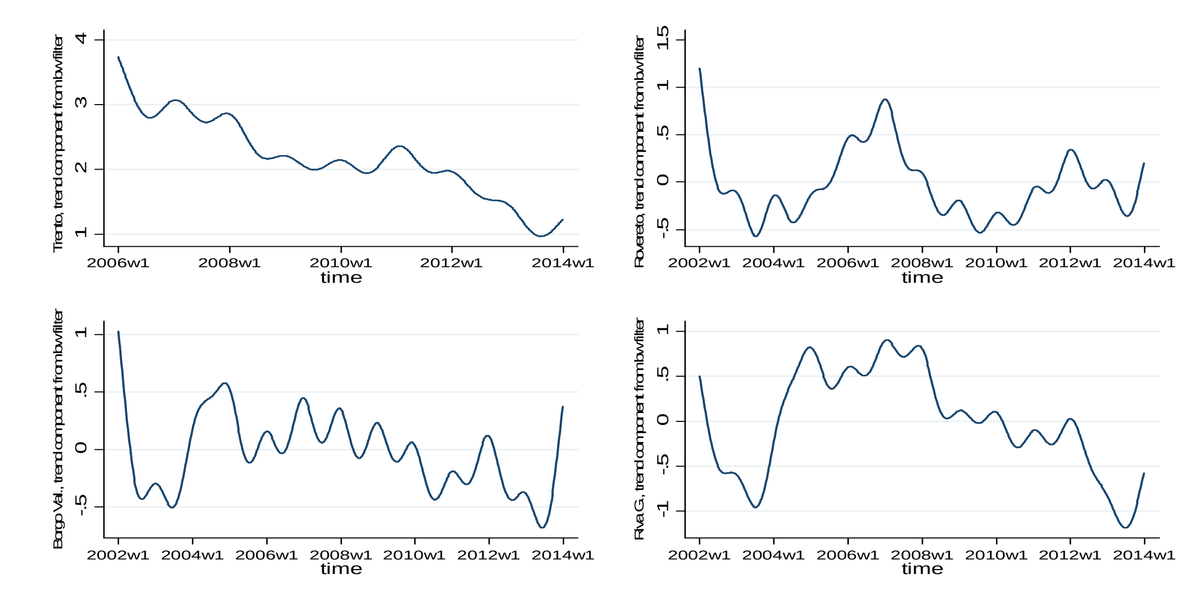

Figure 4: The trend component of the predicted latent factor for the four monitoring sites.

pollution, given the observations on pollutants. What is more important is that we can give a model based interpretation of the proposed procedure and this could be useful if we want to consider the question of pollution forecast outside the sample The model could be further improved by considering jointly the monitoring sites and by showing, eventually, spatial differences in the pollution indicators.

References

Fontanella, L., Ippoliti, L. and Valentini, P. (2007). Environmental Pollution Analysis by Dynamic Structural Equation Models. Environmetrics, 18, 265-83.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The Generalized Dynamic-Factor Model: Identification and Estimation. The Review of Economics and Statistics, 82, 540-54.

Harvey, A. C. (1989). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge UK, Cambridge University Press.

EC, (2013). Commission staff working document. Impact assessment. SWD(2013)531. (http://ec.europa.eu/environment/archives/air/pdf/Impact_assessment_en.pdf accessed 11/02/2015)

EEA (European Environment Agency), (2014). Air quality in Europe — 2014 report. EEA Report No 5/2014. (http://www.eea.europa.eu/publications/air-quality-in-europe-2014 accessed 19/02/2015)

Harvey, A. C. and Trimbur T. M. (2003). General Model-based Filters for Extracting Cycles and Trends in Economc Time Series. The Review of Economics and Statistics, 85 (2), 244-255.

Pascal M., Corso M., Chanel O., Declercq C., Badaloni C., Cesaroni G., Henschel S., Meister K., Haluza D., Martin-Olmedo P., Medina S. and on behalf of the Aphekom group (2013). *Assessing the public health impacts of urban air pollution in 25 European cities: Results of the Aphekom project*. Science of the Total Environment 449, 390–400.

Proietti, T. (204). Seasonal Specific Structural Time Series. Studies in Nonlinear Dynamics & Econometrics, 8 (2), Article 16.

Stock, J. H. and Watson, M. W. (1989). New Indexes of Coincident and Leading Economic Indicators., in Blanchard O. J. and Fisher S. ed. NBER Macroeconomics Annual, 4, 351-409.

Yu, H.-L., Lin, Y.-C. and Kuo, Y.-M. (2015). A Time Series Analysis of Multiple Ambient Pollutants to Investigate the Underlying Air Pollution Dynamics and Interactions. Chemosphere, http://dx.doi.org/10.1016/j.chemosphere.2014.12.007.

Zuur, A.F., Fryer R.J., Jolliffe, I.T., Dekker, R. and Beukema J.J. (2003). Estimating Common Trends in Multivariate Time Series Using Dynamic Factor Analysis. Environmetrics, 14, 665-685.