Comput Econ
DOI 10.1007/s10614-016-9628-6

A Semi-Parametric Non-linear Neural Network Filter: Theory and Empirical Evidence

Panayotis G. Michaelides[1] · Efthymios G. Tsionas[2,6] · Angelos T. Vouldis[3] · Konstantinos N. Konstantakis[4] · Panagiotis Patrinos[5]

Accepted: 13 October 2016

Abstract In this work, we decompose a time series into trend and cycle by introducing a novel de-trending approach based on a family of semi-parametric artificial neural networks. Based on this powerful approach, we propose a relevant filter and show that the proposed trend specification is a global approximation to *any* arbitrary trend. Furthermore, we prove formally a famous claim by Kydland and Prescott (1981, 1997) that over long time periods, the average value of the cycles is zero. A simple procedure for the econometric estimation of the model is developed as a seven-step algorithm, which relies on standard techniques, where all relevant measures may be computed routinely. Next, using relevant DGPs, we compare and show by means of Monte Carlo simulations that our approach is superior to Hodrick–Prescott (HP) and Baxter and King (BK) regarding the generated distortionary effects and the ability to operate in various frequencies, including changes in volatility, amplitudes and phase.

This paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the author and do not necessarily reflect those of the ECB.

✉ Panayotis G. Michaelides
pmichael@central.ntua.gr

1 Laboratory of Theoretical and Applied Economics, School of Applied Mathematical and Physical Sciences, National Technical University of Athens, Heroon Polytechneiou 9, 157.80, Zografou Campus, Athens, Greece

2 Lancaster University, Lancaster, UK

3 European Central Bank & Bank of Greece, Athens, Greece

4 National Technical University of Athens, Athens, Greece

5 Department of Electrical Engineering (ESAT-STADIUS), Optimization in Engineering Center (OPTEC), Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Leuven (Heverlee), Belgium

6 Athens University of Economics and Business, Athens, Greece

Published online: 02 November 2016

In fact, while keeping the structure of the model relatively simple, our approach is perfectly capable of addressing the case of stochastic trend, in the sense that the generated distortionary effects in the near unit root case are minimal and, by all means, considerably fewer than those generated by HP and BK. Application to EU15 business cycles clustering is presented and the empirical results are consistent with the rigorous theoretical framework developed in this work.

Keywords Neural networks · Filtering · Clustering · EU

1 Introduction

Ever, since the seminal work of Burns and Mitchell (1946), the primary objective in a business cycles framework is to study the fluctuations of a time series around a trend. Despite the fact that throughout the last decades a number of—often contradicting—quantitative techniques have been proposed in order to extract the cyclical component of a time series, what seems to remain elusive in the literature, is an appropriate universal and global technique to be used in order to assess business cycles.

Probably the most popular approach in the literature regards business cycles as fluctuations around a trend, the so-called "deviation cycles" (Lucas 1977). In this context, trend estimation is of outmost importance, because it is necessary for the extraction of the cyclical component and for the propagation of shocks (Nelson and Plosser 1982). In fact, the more accurate the trend estimation, the more reliable the business cycle series extracted. Therefore, reliable trend estimates of a time series are very crucial because they can assist in addressing relevant issues and constitute, therefore, a very important task for researchers.

Thus far, for the extraction of the cyclical component of a time series, researchers assume that the trend specification of the time series follows a certain pattern i.e. linear, exponential etc. Nevertheless, this is an ad-hoc assumption which totally ignores the inherent characteristics of the time series at hand i.e. the existence of fat tails, long memory, etc. To this end, in what follows we formally establish a novel methodological framework that takes into consideration the inherent non-linearities of the time series. More specifically, in this work, we decompose a time series into trend and secular component by introducing a novel de-trending approach based on a family of artificial neural networks (ANNs). So far, ANNs have found limited applications in *Economics*. However, they have very important advantages, such as increased flexibility, excellent approximation properties, and instead of fitting the data with a pre-specified model, they let the dataset itself serve as evidence to support the model's approximation of the underlying model (Santin et al. 2004). Thus, ANNs are quite flexible and attractive when the theoretical trend specification is not known *a priori* (Zhang and Berardi 2001).

In this work, instead of fitting the time series data with a pre-specified trend equation, we utilize an ANN specification and let the dataset itself serve as evidence to support the model's approximation of the underlying trend. Also, by exploiting the excellent approximation properties of Neural Networks we prove formally that the proposed trend specification is a global approximation to *any* arbitrary trend. So far, a famous claim by Hodrick and Prescott (1981, 1997) states that the "conceptual framework is that over long time periods, their average is near zero". In this work, we prove formally (mathematically and statistically) that the produced cyclical component, by means of our proposed approach, does indeed disappear in the long run, or in other words, its mean value is equal to zero.

Next, using a number of relevant data generating processes (DGPs) we investigate: (a) the ability of the proposed neural network filter (NNF) to extract the cyclical component of an artificially generated time series that exhibits cycles in a wide range of frequency domains; (b) the ability of NNF to extract cycles that incorporate changes in volatility, amplitudes and phase shifts; (c) the distortionary effect of the cycles produced by NNF with regard to the artificially generated cycles. As a next step, the results of the aforementioned simulations are compared with Baxter–King (BK) and Hodrick–Prescott (HP) filters, and the Monte Carlo results suggest that the performance of NNF is superior in all cases.

Lastly, our proposed technique is confronted with real-world data to assess its ability to model satisfactorily various situations of interest. In this context, from an economic viewpoint, we provide the estimation and visualization of business cycles fluctuations for output in EU15 using fuzzy clustering to study the creation of groups of countries with similar characteristics.

Given that there has been a growing interest lately in the approaches for de-trending non-stationary times series and for representing their underlying trends, we will show that our proposed technique has the following advantages when compared to the widely adopted filtering methods of Hodrick–Prescott (HP) (Hodrick and Prescott 1997) and Baxter–King (BK) (Baxter and King 1999): First, it avoids the problem of a pre-specified functional form of trend, since it lets the dataset itself serve as evidence to support the model's approximation of the underlying trend. Second, it does not require *a priori* assumptions for the smoothing parameter. Third, it is able to capture the non-linear characteristics that business cycles exhibit. Fourth, it is capable of capturing all frequency ranges and all spectrum peak locations. Fifth, the distortionary effects it creates are very limited even in the near unit root case and, sixth, using Monte Carlo techniques it is clearly superior when compared to the HP and BK using various DGP processes, including the near unit root case.[1]

The paper is structured as follows: Sect. 2 provides a literature review, Sect. 3 introduces the NNF; Sect. 4 derives the proposed filtering method and provides some helpful results; Sect. 5 investigates NNF's ability to capture the cycles generated by a number of DGPs; Sect. 6 sets out the proposed econometric implementation; Sect. 7 presents the empirical results; finally, Sect. 8 concludes.

[1] Other popular approaches include the Kalman filter. For an enlightening survey see Kim and Nelson (1999) and for a rigorous analysis of the theory regarding models with non-stationary time series see Chang et al. (2009). Also, several non-linear models have been estimated on real output growth (e.g. Terasvirta 1994). This strand of the literature assumes that output growth is measured accurately, which is quite unlikely to happen since the data contain measurement errors (e.g. Zellner 1992). Hence, sampling all the states conditional on the parameters is relevant (Giordani et al. 2007) but it is not true of threshold models (Pitt et al. 2010). Pitt et al. (2012) used the particle filter to integrate out the states. Also, Malik and Pitt (2011), using particle filtering theory, approximated the likelihood of the unobserved components.

2 Background Literature

There is a plethora of studies in the literature suggesting that business cycles of a time series exhibit nonlinear properties and thus nonlinear quantitative techniques should be employed for the thorough examination of the problem. Neftci (1984), using a finite Markov process, implemented a test to investigate whether US unemployment is characterized by sudden drops or jumps. The results provided evidence in favor of non-linearity of the time series. Falk (1986) re-evaluated the techniques used by Neftci (1984) by applying them to time series data regarding US GNP, productivity and Investment. Diebold and Rudebusch (1989), using an ARIMA specification, investigated the existence of asymmetries in US GDP time series. Scheinkman and LeBaron (1989), in order to investigate the output of stochastic systems, created a deterministic system whose chaotic output could mimic the behavior of a stochastic system. The model provided evidence in favor of the existence of non-linearities in the US stocks return data. In a seminal paper, Hamilton (1990) created a model that could incorporate discrete shifts in the growth rate of non-stationary time series. The model was tested using post-war data for US GNP, and the results provided evidence in favor of non-linearities in business cycles. According to the paper's findings, periodic shift of growth is an inherent feature of the US economy. Beaudry and Koop (1993) tested the existence of asymmetries in US GNP using an extended ARMA model and after-war data. Their results confirmed the existence of asymmetries in the time series examined.

Moreover, Balke and Fomby (1994) examined fifteen US macroeconomic time series, using Tsay's (1988) outlier specification. Their results provided evidence that outliers are strongly associated with the business cycle of the time series under investigation, which in turn confirms the non linear character of business cycles. Tanizaki and Mariano (1994) developed a simulation based non-linear filter that could be applied in non-normal and non-linear times series. The results showed that the estimates of their technique were less biased than those provided by extended Kalman filtering. Ramsey and Rothman (1996) conducted time series irreversibility test. Their results were in favor of the existence of asymmetries. Brunner (1997) made an attempt to reconcile the empirical literature regarding the existence of asymmetries in a time series by implementing the majority of statistical tests used in the literature to investigate the properties of US GNP. The results showed that the time series has non-linear characteristics.

Asymmetric persistence of US GDP time series via a variety of non-linear statistical tests was examined by Hess and Iwata (1997). Pesaran and Potter (1997) examined the non-linearity of US output creating a model that allowed for floor and ceiling effects that could alter the dynamics of growth. Their model provided evidence in favor of asymmetries within the time series, validating the non-linear character of business cycles series. Watanabe (1999) created a non-linear filter based on quasi- maximum likelihood that could yield the exact likelihood of stochastic volatility models using linear approximations. The implementation of the filtering technique on real time data yielded promising results. Psaradakis and Sola (2003) considered the issue of testing for asymmetries in business cycles. Their research showed that asymmetries are likely to be detected in practice only when they are particularly prominent.

Recently, Creal et al. (2010) created a robust band pass filter that decomposes a time series into trend and unobserved component. According to their work, the unobserved component is considered a cycle, in different amplitudes and phase shifts. The benchmark cycle of the filter was created using US data, and the implementation of their technique yielded very satisfactory results. In the same spirit, Malik and Pitt (2011) using particle filtering theory, derived the probability density function of unobserved components in state space model and, thus, managed to approximate the likelihood of these unobserved components. Their simulating results were very promising and the derived likelihood function converges asymptotically to the true likelihood function of the unobserved components. In another important work, Andreasen (2011) managed to improve the accuracy and speed of Central Difference Kalman filter for DSGE models in a Bayesian framework. Also, Guarin et al. (2013) proposed a nonlinear filter based on the Fokker–Planck equation to estimate time varying default risk. The implementation of their filter on Dow-Jones industrial average component companies yielded promising results. Again, Andreasen (2013) incorporated a quasi-maximum likelihood estimation on Central Differencing Kalman filtering, in an attempt to estimate non linear DSGE models with non-Gaussian shocks. According to the paper's findings, the estimates are consistent and asymptotically normal for DSGE models solved up to a third order.

Limited research has been done, so far, regarding the applicability of ANNs in a business cycle framework. See, for instance, the papers by Kiani (2005, 2011) who made use of ANNs in order to examine business cycles asymmetries and the fluctuations of economic activity in CIS countries. For a non parametric business cycle model that does not require the use of any functional form see Kauermann et al. (2012), whereas for an assessment of business cycles dynamics through classical linear control analysis see Wingrove and Davis (2012).

Kuan and White (1994) introduced the perspective of ANN to assess the non-linearity of a time series. The results were discussed in a broader context with regards to the non-parametric tests used in econometric literature that could incorporate the non-linear properties of business cycles. Hutchinson et al. (1994) created an ANN model for option pricing based on the asymmetry properties that the time series data exhibit. Vishwakarma (1994) created an ANN model in order to examine the business cycles turning points. The model was tested using monthly data for the US GDP for the period 1965–1989. The results showed that the turning points identified by the model were characterized by extreme accuracy compared to the official dates, whereas the results confirmed the existence of asymmetries. Brockett et al. (1994) developed a neural network model as an early warning system for predicting insurer insolvency. According to their findings, based on a sample of US property liability insurers, neural networks forecasting capabilities outperformed both the results obtained by both discriminants analysis and the National Association of Insurance Commissioners' Insurance Regulatory Information System ratings.

Serrano-Cinca (1997), utilized a feedforwrad neural network model to in attempt to classify companies on the basis of information provided by their financial statements, utilizing a Spanish dataset. The findings were then compared with those obtained by linear discriminant analysis and logistic regression, giving credit to the view that neural networks outperformed the other methods. Faraway and Chatfield (1998) investigated the predicting capabilities of a variety of neural networks models with those obtained from Box–Jenkins and Holt–Winters methods, utilizing data on US industry. According to their findings neural networks suffered from a convergence and local mimina problems, which resulted in poor out-of-sample forecasting performance when compared to previous methods. In this context, the authors suggest caution when utilizing neural networks. Adya and Collopy (1998) investigated the forecasting capabilities of neural networks (NNs) based on the effectiveness of validation and the effectiveness of implementation, utilizing a sample of 48 studies that emerged in the literature between 1988 and 1994. According to their findings, eighteen (18) studies supported the potential of NNs for forecasting and prediction. Swanson and White (1995, 1997a, b) compared the predictive power of non-linear, linear and ANN models in economic and financial time series data. Their results provided evidence in favor of the predictive power of ANNs, implying that time series exhibit non-linear properties. Gencay (1999) and Qi and Maddala (1999) tested the use of both ANN and linear models in order to determine the predictive power on both economic and financial time series data. Their results provided strong evidence in favor of ANNs. Bidarkota (1999) investigated asymmetries in the conditional mean dynamics using GDP data of four US economic sectors, finding evidence of non-linearities in some US sectors.

Qi (2001) employed ANNs in order to model non-linearities of business cycles during US recessions. Anderson and Ramsey (2002) examined how dynamical linkages between indices of industrial production of the US and Canadian economies affect business cycles oscillation as well as their synchronization properties. Andreano and Savio (2002) investigated asymmetries of business cycles time series, using Markov switching techniques on data for G-7 countries. Their results suggest that asymmetries were present in most countries except for France, Germany and UK. Clements and Krolzig (2004) investigated the existence and identification of a common growth cycle among EU countries, using a Markov vector autoregressive process. Their main finding was that there exists a common unobserved component that governs the common growth cycle. Binner et al. (2002), constructed a weighted index measure of money utilizing the "Divisia" formulation and neural networks for the economy of Taiwan. The authors compared the inflation forecasting potential based on their approach with the traditional approach of simple sum counterparts. According to their findings, neural networked based approaches were found to be superior in terms of inflation tracking than the simple sum counterparts. The same findings were in force when the forecasts of the neural network approach were compared to the Vector Error Correction models forecasts (Binner et al. 2004). The robustness of findings of the previous studies was validated by Binner et al. (2005) who utilized data on the euro area and compared both the in-sample and out of sample forecasting capabilities of their neural network specification against the univariate autoregressive integrated moving average and vector autoregressive models. According to their findings, the neural networks specification was found to be superior in all cases.

Kiani and Bidarkota (2004), using a different data set and alternate regime switching models encompassing features to account for time varying volatility, outlier and long memory, showed that business cycle asymmetries were prevalent in all the G7 countries except France and UK. Nevertheless, Kiani (2005, 2007) and Kiani et al. (2005) managed to identify the existence of asymmetries in all G-7 countries time series using NNs. Ozbek and Ozlale (2005), using a non-linear state space model and Kalman filtering, assessed the fluctuations of the Turkish economy. Aminian et al. (2006), using data on US Real Gross Domestic production and Industrial Production investigated the coefficient of determination which accurately measures the ability of linear or nonlinear models to forecast economic data. Their findings gave credit to the view that neural networks outperform linear regression models due to the inherent nonlinearities of the data. Kiani and Kastens (2006) employed ANNs in a business cycles framework in order to assess recessions. Again, Kiani (2011) investigated asymmetries of business cycles fluctuations among CIS countries using ANNs. Kauermann et al. (2012) employed a non-parametric business cycle model that does not require the use of any specific functional form.

In conclusion, a plethora of studies suggest that business cycles exhibit non-linearities and, thus, non-linear techniques are highly relevant.

3 ANNs as Global Approximators of Trend

3.1 General Formulation of ANNs

According to Pollock (2000), filtering techniques in business cycle analysis is the notion that an economic time series can be represented as the sum of a set of statistically independent components each of which has its own characteristic spectral properties and if the frequency ranges of the components are completely disjoint, then it is possible to achieve a definitive separation of the time series into its components. However, if the frequency ranges of the components overlap, then it is still possible to achieve a tentative separation in which the various components take shares of the cyclical elements of the time series. Contrary to an important strand of the literature (see, among others, Oh et al. 2008), in this work, we focus on models with non-random walk trend components.

Consider a time series denoted by $x_t, t \in T \subseteq \mathbb{R}^+$. Following Hodrick and Prescott (1981) any time series can be decomposed into a (non-stationary) trend component and a (stationary) cyclical component. Therefore, any observed time series has the following representation:

$$x_t = c_t + g_t \tag{1}$$

where x_t denotes the observed time series; c_t denotes the unobserved cyclical component of the observed time series x_t; g_t denotes the unobserved trend that the observed time series x_t exhibits , $t \in T \subseteq \mathbb{R}^+$ denotes the time subscript.[2] Here, g_t is the structure around which c_t is fluctuating. In other words, we have decomposed the structure into long phase movement g_t and shorter fluctuations c_t.

In this work, we use artificial neural networks (ANNs) to estimate the long term trend g_t. Our aim is to quantify c_t as trajectory over t as the fluctuations around g_t and

[2] There is also a seasonal component, which is removed when seasonally adjusting the dataset (Hodrick and Prescot 1981, 1997).

from this trajectory allow for an economic interpretation. Of course, the mechanism behind our approach is, in principle, the one that is being used so far by all the relevant filters in the empirical literature. Nevertheless, in our model we assume that the trend of a time series g_t is the component which comprises only non-cyclical elements in such a way that all the cyclical elements regardless of their frequency range are included in c_t.[3]

The main idea in this paper is to express the trend not as a pre-specified form based on *a priori* assumptions, but rather let the dataset itself determine the specification of the underlying trend. In other words, instead of fitting the data with a pre-specified trend, ANNs let the dataset itself serve as evidence to support the model's approximation of the trend.

ANNs are data-driven and self-adaptive, nonlinear methods that do not require specific assumptions about the underlying trend (Zhang and Berardi 2001). In mathematical terms, ANNs are collections of activation functions that relate an output variable Y to certain input variables $X' = [X_1, \ldots, X_n]$. The input variables are combined linearly to form K intermediate variables or projections $Z_1, \ldots, Z_K$: $Z_k = X'\beta_k (k = 1, \ldots K)$ where $\beta_k \in \mathbb{R}^K$ are parameter vectors. The intermediate variables are combined non-linearly to produce Y:

$$Y = \sum_{\kappa=1}^{K} a_\kappa \varphi \left(\mathrm{Z}_\kappa\right)$$

where φ is an activation function, the α_k's are parameters and m is the number of intermediate nodes (Kuan and White 1994). By combining simple units with intermediate nodes, the NN can approximate any smooth nonlinearity (Chan and Genovese 2001). As demonstrated by Hornik et al. (1989, 1990), ANNs provide approximations to a large class of arbitrary functions while keeping the number of parameters to a minimum. In other words, they are universal approximators of functions. Also, they can approximate their derivatives, a fact which justifies their success in empirical applications (Hornik et al. 1990; Brazili and Siltzia 2003).

3.2 Global Approximation of Trend

Below, we will prove mathematically that the proposed ANN specification for the trend g_t is a global approximation to *any* arbitrary trend.

Now, Theorem 1 holds.

[3] For the standard approaches, the trend of a time series is usually regarded as the component, which comprises its non-cyclical elements together with the cyclical elements of lowest frequency (Kozicki 1999). In particular, according to Pollock (2000) popular filters such as the HP, allow powerful low-frequency components to pass through into the de-trended series when they ought to be impeded by the filter and this deficiency is liable to induce spurious cycles in de-trended data series. This is one of the drawbacks of the HP filter. Of course, there exist other model-based approaches that constitute important alternatives (e.g. Harvey and Todd 1983; Hillmer and Tiao 1982; Koopman et al. 1995) which, however, impose features that are often regarded as being undesirable (Pollock 2000).

Theorem 1 *Consider* $X \subseteq \mathbb{R}^{\mathrm{N}}$ *a compact subset of* $\mathbb{R}^{\mathrm{N}}$ *for some* $N \in \mathbb{N}$ *and* $C(X)$ *is the space of all real valued continuous functions in X. Let* φ*, belonging to* $C(X)$*, be a non-constant, bounded and continuous function. Then, the family* $\mathcal{F} = \left\{F(x) \equiv \sum_{\mathrm{i}=1}^{\mathrm{N}} a_{\mathrm{i}}\, \varphi\left(w_{\mathrm{i}}{}^{\mathrm{T}} x + b_{\mathrm{i}}\right), x \in X \text{ and } F(x) \in C(X)\right\}$ *is dense in* $C(X)$ *for any compact subset of X, where:* $a_{\mathrm{i}}, b_{\mathrm{i}} \in \mathbb{R}$, $w_{\mathrm{i}} \in \mathbb{R}^{\mathrm{m}}$ *are parameters,* $\mathrm{i} \in \{1, \ldots, N\}$ *is an index and T denotes transposition.*

Proof The proof is a straightforward application of Hornik's (1991) 2nd Theorem.

We make the assumption that between any two points in time there are an infinite number of points in time. Hence, the variable time, ranges over the entire real number line or, in our context, over a subset of it : the non-negative reals. In another formulation, time is a continuous variable.

Next, in view of Definitions 1 and 2, (Appendix) (trend set and time series), the following holds (Lemma 1): □

Lemma 1 *If* $g_{t_j}, t \in T \subseteq \mathbb{R}^+, j \in J \subseteq \mathbb{R}$ *is an arbitrary time series representing trend, such that* $g_{t_j} \in \mathbb{R} \forall j \in J$ *and* $\left\{g_{t_j} : j \in J\right\}$ *is the trend set that is closed and bounded, then the trend set is a compact subset of* $\mathbb{R}$*.*

Proof See Rudin (1976).

Next, based on Lemma 1, as well as on Lemma 2, below, we will prove Theorem 2, which shows that the NN trend is a global approximation to any arbitrary time trend. □

Lemma 2 *If T is a compact set and* $F(t) \equiv \sum_{N=1}^{N} s_i \varphi(\beta_i t)$ *is: (i) non constant, (ii) bounded and (iii) continuous, then any function of the form:*$k(t) \equiv h(t) + F(t)$*, where:* $h(t)$ *is a linear function of* $t \in T$*, is: (i) bounded, and (ii) continuous.*

Proof The Proof is trivial. □

Theorem 2 *If* $\bigcup_{j \in J} g_{t\,j}$ *is the trend set of a time series, then the family of functions* $\mathcal{F} = \{F(t) \in C(\bigcup_{j \in J} g_{t\,j}) : F(t) \equiv d + ct + \sum_{i=1}^{N} a_i \varphi(\beta_i t), \alpha_i, \beta_i, d, c \in \mathbb{R}\}$ *is dense in* $C(\bigcup_{j \in J} g_{t\,j})$ *for every compact subset* $\bigcup_{\mathrm{j} \in \mathrm{J}} g_{\mathrm{tj}} \subseteq \mathbb{R}$.

Proof See Appendix. □

Theorem 2 shows that the ANN trend is a global approximation to any arbitrary time trend.

4 Construction of the Filter

4.1 Mathematical Derivation

We have seen that any time series x_t can be expressed as:

$$x_t = c_t + g_t \tag{2}$$

As stated earlier, we assume that the trend of a time series g_t is the component, which comprises only non-cyclical elements in such a way that all the cyclical elements regardless of their frequency range are included in c_t.

Since trend in time series is not *a priori* known, in view of Theorem 2,we may assume, without loss of generality, that the general representation of the global approximation (g_t) is the following:

$$g_t = a_0 + \delta t + \sum_{k=1}^{m} \alpha_k \varphi\left(\beta_k t\right) \tag{3}$$

where $a_0, \delta, \alpha_k, \beta_k \in \mathbb{R} \forall m = 1, ..N$ denote parameters and φ denotes an activation function that is a non-constant, bounded and continuous function. Therefore, the general representation of any time series described in (2), with the use of the specification in (3), can be expressed as:

$$x_t = c_t + a_0 + \delta t + \sum_{k=1}^{m} \alpha_k \varphi\left(\beta_k t\right) \tag{4}$$

Thus:

$$c_t = x_t - [a_0 + \delta t + \sum_{k=1}^{m} \alpha_k \varphi\left(\beta_k t\right)] \tag{5}$$

4.2 Properties of the Filter

(a) Linear Time Trend as Degenerate form of NNF

In the seminal contribution by Hodrick and Prescott (1981, 1997), the smoothing parameter λ is a positive number which, *ceteris paribus*, penalises variability in the trend component series. Hence, the larger the value of λ, the smoother the cyclical component series. Also, according to Hodrick and Prescott's claim (1981, 1997, p. 3) in their seminal paper, as λ becomes sufficiently large, it degenerates to the least squares fit of a linear time trend model.[4]

In what follows, we will prove formally the aforementioned claim by Hodrick and Prescott's (1981, 1997, p. 3) for our specification, which constitutes, as shown, a global approximation. More precisely, we will show that for a sufficiently large number of nodes, at the optimum, the limit of solutions to the minimization problem, is the least squares fit of a linear time trend model.

Theorem 3 (Linear time trend as degenerate form of NNF) *If* $\overline{\beta_{\bar{m}_{l_0}}} = max\{\overline{\beta_{\bar{m}_l}}, \overline{\beta_{\bar{m}_l}} \in \mathbb{R}^{\overline{m}}\}$*, then the trend approximation produced by NNF is linear, i.e.* $g_t = \gamma + \delta t$, $\forall$ *m* $\in${*1,..., M*} *and* $\forall x_{t_i}, i \in I$*, where* I *is considered to be a compact subset of* $\mathbb{R}$.

Proof See Appendix. □

[4] In this section, for reasons of notation, when we consider fixed (instead of free) parameters, then the respective parameter is denoted by an upper bar.

(b) Mean Value of the Cycle is zero

In the seminal work by Hodrick and Prescott (1981, p. 3, 1997, p. 3), which was based on the so-called Whittaker-Henderson Type A method (Whittaker 1923), it is claimed that, regarding the business cycles defined as deviations from trend, the "conceptual framework is that over long time periods, their average is near zero". In what follows, we will prove mathematically their famous claim, namely that the mean value of the cycle produced by NNF is equal to zero. Based on the definitions of (i) trend set, (ii) time series as a random variable and (iii) time series set, respectively (Definitions 1-3, Appendix), we state our results in the form of two theorems, one more general and one more case specific.

Theorem 4 (Mean value of the cycle is zero) *For any time series* $x_{t_j} \forall j \in J$ *and* $\forall t \in T \subseteq \mathbb{R}^+$ *that can be decomposed into trend and cycle as follows:* $x_{t_j} = g_{t_j} + c_{t_j}$ *the mean value of the cycle* $c_{t_j} = x_{t_j} - g_{t_j} \forall j \in J$ *is equal to zero, i.e.* $E\left(c_{t_j}\right) = 0 \, \forall j \in J$, *if the trend set* $\bigcup_{t_i} g_{t_i}$ *is a dense subset of the set of time series* $\bigcup_{t_i} g_{t_i} \subseteq \bigcup_{t_j} x_{t_j}$.

Proof See Appendix. □

The rationale behind this important finding is that if the trend set is dense in the time series set, then the expected value of any cycle defined as trend deviation is zero. Nevertheless, this property does not necessarily hold for any cycle regardless of the method employed, since the aforementioned Theorem pre-supposes the trend set to be dense in the time series set.

Theorem 5 (Mean value of the NNF cycle is zero) *The mean value of the cycle of a time series produced by NNF is equal to zero, i.e.* $E\left(c_{t_j}\right) = 0, \forall j \in J$ *provided the trend set* $\bigcup_{t_i} g_{t_i}$ *is a dense subset of the set of time series* $\bigcup_{t_i} g_{t_i} \subseteq \bigcup_{t_j} x_{t_j}$.

Proof See Appendix. □

We have proved that the cycles produced by the proposed technique have a mean value equal to zero. The rationale behind this finding is that the remaining part (the cycle) is non-negligible but, on average, equal to zero. In other words, business cycles as deviations from trend are disturbances from a growth path (positive or negative) that lead, sooner or later, to a return to the growth path. If this wasn't the case and there were some non-trivial distinctive trend in time, this would have been captured by the NNF.

The economic intuition of this finding is that an economy, despite business cycles, moves into new neighbourhoods of growth, i.e. new growth paths (positive or negative) and, thus, the kind of wave-like movement is inherent to economic growth. This implies that growth occurs despite a business cycle process and as a result these cyclical fluctuations are no barrier to economic growth, in the sense that deviations below trend are asymptotically not necessarily expressions of deep crisis or generalised breakdown, and so on. After all, several well-known economists, such as Schumpeter (1939), believed that recessions are to be followed by periods of fast growth.

5 Simulation-Based Comparison of NNF

5.1 Empirical Estimation of NNF

As seen earlier, the proposed specification, based on Eq. (3), is as follows:

$$x_t = a_0 + \delta t + \sum_{\mathrm{k=1}}^{\mathrm{m}} \alpha_k \varphi\left(\beta_k t\right) + \mathrm{u_t}$$

where x_t is the time series, m is the number of nodes, $\mathrm{u_t}$ is the error term and t is time.

We will make use of a typical activation function which is continuous, bounded, differentiable and monotonic increasing (e.g. Hornik et al. 1989, 1990), namely $\varphi(z) = \frac{1}{1+e^{-z}}$, $z \in \mathbb{R}$. For other activation functions, see Bishop (1995).[5]

In order to empirically estimate the parameters of our model, we are based on the aforementioned equation, which has an estimable form. We propose the following estimation procedure that consists of a simple seven (7) step algorithm.

Algorithm 1: NNF filtering
Step 1: For $m = 1$, $\overline{\beta^l}_k$, $k = 1, \ldots, m$, are drawn from a uniform distribution on a hyper-rectangle $\Omega \subset \mathbb{R}^m$.
Step 2: Given these parameters, estimate $a_0, \alpha_k, \delta, k = 1, \ldots, m$ by means of Ordinary Least Squares (O.L.S.) applied to the following equation:

$$x_t = \left[a_0 + \delta t + \sum_{\mathrm{k=1}}^{\mathrm{m}} \alpha_k \varphi\left(\overline{\beta_k^i}, t\right) \right] + \mathrm{u_t}$$

$t = 1 \ldots T$.
Step 3: For the estimated parameters $a_0, \alpha_k, \delta, k = 1, \ldots, m$ which can be regarded as known, consider $\beta_k^i, k = 1, \ldots, m$ as a parameter and find its value routinely using numerical analysis techniques for non-linear equations (e.g. Broyden–Fletcher–Goldfarb–Shannon method).
Step 4: For theses values of $\overline{\beta_k^i}$, $k = 1, \ldots, m$, estimate the set of parameters a_0, α_k, δ, $k = 1, \ldots,$ using OLS.
Step 5: For the whole set of parameters $a_0, \alpha_k, \delta, k = 1, \ldots, m$ and $\overline{\beta_k^i}$, $k = 1, \ldots, m$, compute a relevant criterion, such as the Schwartz Bayes Information Criterion (BIC).[6]
Step 6: Repeat steps 1–5 for $m = 2, 3, 4 \ldots$ and keep the value of m that optimizes the aforementioned criterion. For $m^* \in \{1, ..N\}$ that optimizes the criterion selected, keep the calculated values of $\overline{\beta_{m^*}}$, and the estimated values $\overline{a_{m^*}}$, $\overline{a_{0m^*}}$, $\overline{\delta_{m^*}}$ Now, for these

[5] However, in general, the empirical results are robust, regardless of the activation function used because of the typical properties they posses (Haykin 1999).

[6] For an extensive survey on methods regarding the selection of the number of nodes in neural networks or for the appropriate model selection using information criteria see, among others, Sheela and Deepa (2013) and Konishi and Kitagawa (1996), respectively.

values of $m*$, $\overline{\beta_{m^*}}$, $\overline{a_{m^*}}$, $\overline{a_{0m^*}}$ and $\overline{\delta_{m^*}}$ we get the values of c_t which are the following:

$$c_t = x_t - [\overline{a_{0m*}} - \overline{\delta_{m^*}}t - \sum_{k=1}^{m^*} \overline{\alpha_k}\,\varphi\left(\overline{\beta_k}\,t\right)] \tag{6}$$

5.2 Arbitrary Frequency Domain

As stated earlier, one of the most serious problems that the traditional filters, such as HP and BK, face is their inability to extract the cyclical component from a time series that exhibits a cycle in different frequencies from the ones dictated by their arbitrarily chosen smoothing parameter. In this context, in order to investigate the ability of our filer to approximate trends and, therefore, cycles of any arbitrary frequency we make use of the data generating process (DGP) proposed in Guay and Saint-Amant (2005):

$$y_t = \mu_t + c_t \tag{7}$$

where

$$\begin{aligned} \mu_t &= \mu_{t-1} + \varepsilon_t, \quad \varepsilon_t \quad \sim NID\left(0, \sigma_\varepsilon^2\right) \\ c_t &= \varphi_1 c_{t-1} + \varphi_2 c_{t-2} + n_t, \quad n_t \sim NID\left(0, \sigma_n^2\right) \end{aligned}$$

Equation (7) defines y_t as the sum of a permanent component μ_t,which corresponds to a random walk, and a cyclical component c_t, which corresponds to a second order autoregressive process AR(2).[7] We also assume that ε_t and n_t are uncorrelated. Thus, the following equation expresses the DGP:

$$y_t = \mu_{t-1} + \varphi_1 c_{t-1} + \varphi_2 c_{t-2} + v_t \tag{8}$$

where $\varphi_1 + \varphi_2 < 1$. The use of an AR(2) series is useful because its spectrum may have a peak in either business cycles frequencies or at zero frequency. Now, despite the fact that this process is stationary, a continuity argument provides information also for the case of non-stationary series since, in a finite sample, any non-stationary series can be approximated by a stationary process and *vice versa* (Campbell and Perron 1991).

The spectrum of the process described in (8) is given by:

$$f_y(\omega) = \frac{\sigma_v^2}{1 + \varphi_1^2 + \varphi_2^2 - 2\varphi_1(1 - \varphi_2)\,cos\omega - 2\varphi_2 cos2\omega} \tag{9}$$

and the location of its peak is given by the expression:

$$-\sigma_v^{-2} f_y^2(\omega)\, 2sin\omega[\varphi_1(1 - \varphi_2) + 4\varphi_2 cos\omega \tag{10}$$

[7] This selection of the cyclical component was made so that the peak of the spectrum in our cycle could be either at zero frequency or at business cycle frequencies.

Table 1 Filter correlation with the true cycle when $\sigma_\varepsilon/\sigma_\eta = 10$

σε / ση	φ1	φ2	CorNNF	CorHP	CorBK	NNF range	HP range	BK range
10	0	0	0.81	0.08	0.03	0.77, 0.85	−0.07, 0.21	−0.11, 0.16
10	1.2	−0.25	0.82	0.08	0.08	0.77, 0.85	−0.11, 0.28	−0.13, 0.32
10	1.2	−0.40	0.88	0.13	0.11	0.83, 0.91	−0.12, 0.36	−0.16, 0.36
10	1.2	−0.55	0.87	0.14	0.12	0.84, 0.89	−0.08, 0.33	−0.12, 0.32
10	1.2	−0.75	0.81	0.15	0.16	0.78, 0.85	−0.01, 0.44	−0.04, 0.36

Table 2 Filter correlation with the true cycle when $\sigma_\varepsilon/\sigma_\eta = 5$

σε / ση	φ1	φ2	CorNNF	CorHP	CorBK	NNF range	HPrange	BK range
5	0	0	0.86	0.15	0.05	0.84, 0.89	0.02, 0.27	−0.09, 0.20
5	1.2	−0.25	0.91	0.16	0.17	0.89, 0.93	−0.01, 0.36	−0.05, 0.38
5	1.2	−0.40	0.91	0.23	0.23	0.89, 0.94	−0.01, 0.45	−0.03, 0.47
5	1.2	−0.55	0.92	0.24	0.26	0.90, 0.94	0.01, 0.44	0.03, 0.46
5	1.2	−0.75	0.92	0.29	0.28	0.89, 0.94	0.11, 0.44	0.09, 0.45

Thus, $f_y(\omega)$ has a peak in frequencies other than zero for:

$$\varphi_2 < 0 \text{ and } \left|\frac{-\varphi_1(1-\varphi_2)}{4\varphi_2}\right| < 1 \tag{11}$$

Then, $f_y(\omega)$ has a peak in frequencies given by the expression:

$$\omega = cos^{-1}\left(\frac{-\varphi_1(1-\varphi_2)}{4\varphi_2}\right) \tag{12}$$

Therefore, in order to investigate the ability of our proposed filter to approximate low frequency cycles we make use of the DGP in Eq. (8) with φ_1 set at the value of 1.2 and different values of φ_2 in order to control for the location of the peak in the spectrum of the cyclical component. We also vary the standard error ratio for the disturbances $\sigma_\varepsilon/\sigma_n$ so as to change the relative importance of each component. Here, we have to bear in mind that the peak of the DGP in use is located in the business cycle frequencies dictated by HP and BK, when $\varphi_2 < -0.43$. The resulting time series contains 150 observations, a standard size for macroeconomic time series, while the number of iterations was set to 10,000. The smoothing parameters used for the HP and BK are set equal to 1600 and 6–32 respectively as the relevant literature suggests (e.g. Baum et al. 2006), contrarily to the NNF which is data driven. The HP and BK correlation coefficients and their respective ranges come from Guay and Saint-Amant (2005, pp. 148–151) where the produced time series also contained 150 observations, while the number of replications was equal to 500.[8]

The results of all filters are summarized in Tables 1, 2, 3, 4 and 5.

[8] Despite the difference in the number of iterations between the two procedures, in an econometric perspective, the average correlation coefficient in both procedures is robust, and the only difference lies in the

Table 3 Filter correlation with the true cycle when $\sigma_\varepsilon/\sigma_\eta = 1$

$\sigma_\varepsilon / \sigma_\eta$	φ_1	φ_2	CorNNF	CorHP	CorBK	NNF range	HP range	BK range
1	0	0	0.91	0.59	0.19	0.88, 0.92	0.49, 0.70	0.05, 0.32
1	1.2	−0.25	0.91	0.51	0.53	0.88, 0.92	0.33, 0.68	0.36, 0.71
1	1.2	−0.40	0.90	0.71	0.70	0.88, 0.92	0.56, 0.82	0.55, 0.81
1	1.2	−0.55	0.90	0.76	0.73	0.88, 0.92	0.56, 0.82	0.61, 0.83
1	1.2	−0.75	0.90	0.83	0.79	0.88, 0.92	0.75, 0.89	0.69, 0.87

Table 4 Filter correlation with the true cycle when $\sigma_\varepsilon/\sigma_\eta = 0.5$

$\sigma_\varepsilon / \sigma_\eta$	φ_1	φ_2	CorNNF	CorHP	CorBK	NNF range	HP range	BK range
0.5	0	0	0.93	0.82	0.36	0.90, 0.95	0.75, 0.88	0.25, 0.47
0.5	1.2	−0.25	0.90	0.61	0.63	0.88, 0.92	0.41, 0.79	0.45, 0.78
0.5	1.2	−0.40	0.88	0.84	0.81	0.86, 0.91	0.73, 0.92	0.71, 0.88
0.5	1.2	−0.55	0.88	0.89	0.85	0.87, 0.91	0.83, 0.94	0.78, 0.91
0.5	1.2	−0.75	0.87	0.94	0.89	0.85, 0.91	0.90, 0.96	0.83, 0.93

Table 5 Filter correlation with the true cycle when $\sigma_\varepsilon/\sigma_\eta = 0.01$

$\sigma_\varepsilon / \sigma_\eta$	φ_1	φ_2	CorNNF	CorHP	CorBK	NNF range	HP range	BK range
0.01	0	0	0.99	0.98	0.55	0.96, 1.00	0.96, 0.99	0.48, 0.63
0.01	1.2	−0.25	0.92	0.66	0.68	0.90, 0.94	0.45, 0.83	0.52, 0.82
0.01	1.2	−0.40	0.90	0.90	0.86	0.88, 0.92	0.82, 0.96	0.79, 0.92
0.01	1.2	−0.55	0.88	0.96	0.90	0.86, 0.90	0.91, 0.99	0.85, 0.94
0.01	1.2	−0.75	0.87	0.99	0.93	0.86, 0.90	0.97, 1.00	0.89, 0.96

The results suggest that irrespectively of the variance ratio and the value of autoregressive parameters used, the proposed filter (NNF) exhibits a very high correlation equal to approximately 90 % with the true cyclical component of the time series. Specifically, the proposed NNF produces robust estimates of the cycles in the series even when the variance of the cycles is very small in the series i.e. $\sigma_\varepsilon/\sigma_n > 1$, and (or) the frequency of the cycles are located close to zero, i.e. $\varphi_2 > -0.43$. This, in turn, implies that the proposed filter could be used irrespectively of the cycle frequency and location of the peak in the cycle. Hence, NNF is capable of capturing all frequency ranges and all spectrum peak locations. On the contrary, both HP and BK do well only when the cyclical component of the time series is located in their frequency domain. However, when this is not the case, their ability to approximate the cycle is very poor, in contrast to the proposed filter (NNF).

range intervals of their estimates. To this end, without loss of generality, 10,000 iterations are considered to be an asymptotic estimate. Nevertheless, our analysis is based on the average estimates.

5.3 Volatility, Amplitudes and Phase Shifts

In this section, in order to assess the ability of NNF to extract cycles that incorporate changes in volatility, amplitudes and phase shifts, we adopt the DGP in Creal et al. (2010). Analytically, we use the following data generating process:

$$y_{i,t} = \tau_{i,t} + \delta_i \psi_t^q + \varepsilon_{i,t}, \varepsilon_{i,t} \sim NID\left(0, \sigma_{i,\varepsilon}^2\right) \tag{13}$$

$$\tau_{i,t+1} = \tau_{i,t} + \beta_{i,t} \tag{14}$$

$$\beta_{i,t+1} = \beta_{i,t} + \zeta_{i,t}, \zeta_{i,t} \sim NID\left(0, \sigma_{i,\zeta}^2\right) \tag{15}$$

$$\begin{pmatrix} \psi_{t+1}^j \\ \psi_{t+1}^{+j} \end{pmatrix} = \rho \begin{pmatrix} cos\lambda & sin\lambda \\ -sin\lambda & cos\lambda \end{pmatrix} \begin{pmatrix} \psi_t^j \\ \psi_t^{+j} \end{pmatrix} + \begin{pmatrix} \psi_t^{j-1} \\ \psi_t^{+(j-1)} \end{pmatrix}, \quad j = q, q-1, \ldots \tag{16}$$

$$\begin{pmatrix} \psi_t^0 \\ \psi_t^{+0} \end{pmatrix} = \begin{pmatrix} k_t \\ k_t^+ \end{pmatrix}, \begin{pmatrix} k_t \\ k_t^+ \end{pmatrix} \sim NID\left(0, \sigma_k^2 I_2\right) \tag{17}$$

In Eqs. (13), $\tau_{i,t}$ and $\varepsilon_{i,t}$ are the idiosyncratic trend and irregular components of the i-th variable, respectively. The cyclical component ψ_t^q is specified as a smooth cyclical process where q is an integer denoting the level of smoothness. The cycle is shared by all series and is scaled for each series with δ_i.

In Eqs. (14) and (15) the individual trend component is specified as a smooth local linear trend process, where $\beta_{i,t}$ is the growth of trend. In Eqs. (16) and (17) the dynamics of the common cycle ψ_t^q is modeled as a q-th ordered stochastic model where ρ denotes the damping parameter with the restriction of $\rho \in (-1, 1)$ so as to ensure stationarity; λ is the frequency of the cycle measured in radians with the period of the cycle equal to $\frac{2\pi}{\lambda}$. Thus, the DGP. described by Eqs. (13)–(17) generates a series that can be decomposed into trend and cycle without stochastic volatility.

Now, in order to also account for stochastic volatility in the disturbances we allow in Eq. (13) the disturbances to follow a random walk process whose innovations are a mixture of a standard Gaussian noise sequence and a stochastic indicator variable with known probabilities:

$$\sigma_{i,t,\varepsilon}^2 = \exp\left(h_{i,t,\varepsilon}\right) \tag{18}$$

$$h_{i,t+1,\varepsilon} = h_{i,t,\varepsilon} + \mathrm{K}_{i,t,\varepsilon}\omega_{i,t,\varepsilon}, \omega_{i,t,\varepsilon} NID\,(0, 1) \tag{19}$$

Similarly, we allow in Eq. (24):

$$\sigma_{i,t,k}^2 = \exp\left(h_{i,t,k}\right) \tag{20}$$

$$h_{i,t+1,k} = h_{i,t,k} + K_{i,t,k}\omega_{i,t,k}, \omega_{i,t,k} NID\,(0, 1) \tag{21}$$

Lastly, we also allow in Eq. (15):

$$\beta_{i,t+1} = \beta_{i,t} + K_{i,t,\zeta}\zeta_{i,t}, \zeta_{i,t} \sim NID\,(0, 1) \tag{22}$$

Table 6 Correlation coefficients

Model	Correlation	Range
NNF (number of nodes = 1)	0.503	0.451–0.555
NNF (number of nodes = 2)	0.884	0.781–0.974
NNF (number of nodes optimal using BIC)	0.891	0.882–0.981

Table 7 Correlation coefficients

	All-SV	No-SV	Break in ρ	Break in ρ, λ, ξ_i	Coint	No phase shifts
NNF (number of nodes = 1)	0.413–0.512	0.301–0.411	0.054–0.127	0.017–0.223	0.213–0.415	0.417–0.816
NNF (number of nodes optimal by BIC)	0.818-0.945	0.791–0.912	0.818–0.994	0.821–0.994	0.890–0.997	0.812–0.913

Thus, the DGP described by Eqs. (18)–(22) and (13), (14) and (16) generates a series that can be decomposed into trend and cycle with stochastic volatility.

Next, using the results by Creal et al. (2010, p. 706) and 10,000 iterations in the D.G.P. that does not account for stochastic volatility, we present the range of the correlation coefficients of the cycle derived by means of our proposed filtering technique (NNF) and their actual cycle dictated by the selection of posteriors in Creal et al. (2010, Table II).

The results in Table 6 suggest that the NNF exhibits a very high correlation with the actual cycle generated by the GDP and the selection of posteriors in Creal et al. (2010). Specifically, the use of BIC as a selection criterion for the optimal number of nodes in the NNF is of crucial importance, since the *ad-hoc* selection of nodes exhibits a significantly lower ability to capture the cycle.

The robustness of NNF is further confirmed in Table 7 where the range of the correlation coefficients between the cycles derived from the NNF and the actual cycles by the DGPs, based on Creal et al. (2010, Table IV), are presented.

5.4 Distortionary Effects and Near-Unit Root Behaviour

In this subsection, in order to measure the distortionary effect of NNF in comparison to HP and BK, for different values of the smoothing (tuning) parameters, following Pedersen (2001) we employ the following DGP.:

$$y_t = \varphi y_{t-1} + \varepsilon_t \tag{23}$$

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-1} + v_t \tag{24}$$

where $\varepsilon_t, v_t \sim N\left(0, \sigma^2\right)$.

Table 8 Size of distortions of filters measured for 5 autoregressive processes with 31,416 gridpoints on frequencies, multiplied by 1000

Cyclical components	AR(1) $\varphi = 0.90$	$\varphi = 0.95$
HP	73.6	44.06
BK	74.7	41.8
NNF1 (number of nodes optimal using BIC)	11.271	9.293
NNF2 (number of nodes = 2)	13.535	11.201

The AR(1) data generating process described in Eq. (23) was implemented with different values of φ. Analytically, we made use of the values of $\varphi = 0.90$ and $\varphi = 0.95$ in order to account for stationarity.

However, of great interest is also the so-called near unit root case which refers to the stochastic trend case. In other words, we would like to comparatively examine our model's behaviour in terms of distortionary effects when a stochastic trend is present. In this context, we used the values of $\varphi = 0.99$ and $\varphi = 0.999$ that resemble a near unit root process. Furthermore, the AR(2) process described in Eq. (24) was implemented using two near unit root specification for the values of $\varphi_1 = 1.3297$ and $\varphi_2 = -0.3318$.

The distortionary effect of each filter is typically measured as the absolute difference between the true and the distorted cyclical component at frequency ωmultiplied by the size of the grid on ω, using the following formula:

$$Q = \sum\nolimits_{\omega \in W} |H(\omega) - H^*(\omega)| 2S_y(\omega)\,\Delta\omega \tag{25}$$

where Δ denotes the first difference operator.

If we, now, normalize the weights ω in (25) so as to sum up to unity we get:

$$Q_\omega = \sum\nolimits_{\omega \in W} |H(\omega) - H^*(\omega)| v(\omega) \tag{26}$$

where $v(\omega)$ is the ratio of power spectral density of the input process, $S_y(\omega)$, at frequency ω over the variance of the series, i.e.

$$v(\omega) = \frac{2S_y(\omega)\,\Delta\omega}{\sum_{\omega \in W} 2S_y(\omega)\,\Delta\omega} \tag{27}$$

where $\sum_{\omega \in W} 2S_y(\omega)\,\Delta\omega$ is approximately the variance of the series.

Tables 8 and 9 present the distortionary effects by means of HP, BK and NNF using the normalized Q-stat given by Eq. (26).[9] Additionally, the NNF simulations were performed using 10,000 iterations, while the results of BK and HP come from Pedersen (2001, p. 1094).

[9] The smoothing parameters used for the HP and BK are again set equal to 1600 and 6–32 respectively, as the relevant literature suggests (e.g. Baum et al. 2006), contrarily to the NNF which is data driven.

Table 9 Near-unit root size of distortions of filters measured for 5 autoregressive processes with 31,416 gridpoints on frequencies, multiplied by 1000

Cyclical components	AR(1) $\varphi = 0.99$	AR(1) $\varphi = 0.999$	AR(2)
HP	9.331	0.092	2.86
BK	8.6	0.086	2.54
NNF1 (number of nodes optimal using BIC)	7.215	0.0053	0.325
NNF2 (number of nodes = 2)	9.343	0.0061	0.414

The results suggest that in the AR(1) process, based on the autoregressive DGP used, the NNF presents significantly fewer distortions compared to HP and BK. Furthermore, regarding the near-unit root case which corresponds to a stochastic trend, we note that in the AR(1) and AR(2) processes, the NNF outperforms the other two traditional approaches, HP and BK, by an order of magnitude, approximately. According to Phillips and Magdalinos (2008), under both strict and moderate non-stationarity, central limit theory indeed applies, while such systems may also be more realistic for practical work.

Despite the fact that both HP and BK filters perform better when they get closer to the near-unit root case, the NNF exhibits fewer distortions compared to both HP and BK, irrespectively of the (non-)stationarity characteristics of the series. This, in turn, is attributed to the fact that the smoothing/tuning parameters are specified *a priori* for HP and BK, in contrast to NNF. Actually, the fact that it is superior in all cases should be attributed to the NNF being data-driven and a global approximation. Lastly, the fact that NNF1 presents fewer distortions than NNF2 is clearly attributed to the different number of nodes used in each case.

To sum up, while keeping the structure of the proposed approach relatively simple, NNF exhibits superior performance compared to HP and BK in terms of distortionary effects, including the stochastic trend case, where the NNF outperforms the traditional approaches by an order of magnitude, approximately.

5.5 Mean Value of the Cycle

We have shown that if the trend set is dense in the time series set, then the expected value of any cycle that is defined as trend deviation is zero. Also, we have proved that the cycle produced by the NNF has a mean value equal to zero. In order to test the empirical validity of our theoretical results, in what follows we report the mean value of the cycle produced by NNF based on the various DGP processes that have been employed using 10,000 iterations each. We can see from the following Table that, as expected, the mean value of the cyclical component produced by NNF is not significantly different from zero (0), a result which implies that the rigorous theoretical framework developed, even in its strict form, is fully consistent with the simulated results (Tables 10).

Table 10 Expected versus observed values of the NNF cycle

DGP process			Expected Mean value C_t	Observed Mean value C_t	p value
Guay and Saint-Amant (2005)					
$\sigma_\varepsilon / \sigma_\eta$	φ_1	φ_2			
10	0	0	0.00	−0.032	0.515
10	1.2	−0.25	0.00	−0.044	0.612
10	1.2	−0.40	0.00	−0.003	0.781
10	1.2	−0.55	0.00	0.0045	0.677
10	1.2	−0.75	0.00	0.0376	0.981
5	0	0	0.00	−0.0036	0.627
5	1.2	−0.25	0.00	0.0047	0.335
5	1.2	−0.40	0.00	0.0038	0.871
5	1.2	−0.55	0.00	−0.0093	0.776
5	1.2	−0.75	0.00	0.0049	0.615
1	0	0	0.00	−0.0012	0.335
1	1.2	−0.25	0.00	0.0023	0.889
1	1.2	−0.40	0.00	0.0032	0.557
1	1.2	−0.55	0.00	−0.0044	0.345
1	1.2	−0.75	0.00	0.0072	0.635
0.5	0	0	0.00	−0.0033	0.454
0.5	1.2	−0.25	0.00	−0.0033	0.675
0.5	1.2	−0.40	0.00	−0.0047	0.224
0.5	1.2	−0.55	0.00	0.0046	0.456
0.5	1.2	−0.75	0.00	0.0058	0.724
0.01	0	0	0.00	0.0061	0.464
0.01	1.2	−0.25	0.00	−0.0054	0.238
0.01	1.2	−0.40	0.00	0.0018	0.895
0.01	1.2	−0.55	0.00	−0.0025	0.734
0.01	1.2	−0.75	0.00	0.0035	0.724
Pedersen (2001)					
NNF-1 node					
AR(1)					
$\varphi = 0.90$			0.00	0.0012	0.198
$\varphi = 0.95$			0.00	0.0013	0.234
$\varphi = 0.99$			0.00	0.0015	0.337
$\varphi = 0.999$			0.00	0.0017	0.387
AR(2)			0.00	0.0017	0.357
NNF BIC nodes					
AR(1)					
$\varphi = 0.90$			0.00	−0.0015	0.475
$\varphi = 0.95$			0.00	−0.0022	0.345

Table 10 continued

DGP Process	Expected Mean Value C_t	Observed Mean Value C_t	p value
$\varphi = 0.99$	0.00	0.0017	0.478
$\varphi = 0.999$	0.00	0.0016	0.464
AR(2)	0.00	−0.0003	0.497
Creal et al. (2010)			
NNF-1 node	0.00	0.0043	0.623
NNF-2 nodes	0.00	−0.0013	0.623
NNF-bic nodes	0.00	0.0015	0.676
NNF-1 node			
All SV	0.00	−0.0045	0.234
No SV	0.00	−0.0032	0.455
Break in ρ	0.00	0.0018	0.677
Break in ρ, λ, ξ	0.00	0.0021	0.453
Coint	0.00	0.0019	0.678
No Phase shifts	0.00	0.0017	0.482
NNF BIC nodes			
All SV	0.00	0.0023	0.345
No SV	0.00	0.0035	0.675
Break in ρ	0.00	−0.0017	0.546
Break in ρ, λ, ξ	0.00	−0.0022	0.567
Coint	0.00	−0.0017	0.845
No Phase shifts	0.00	0.0013	0.456

p values correspond to the robust (HAC) t-statistic of regressing actual minus predicted on a constant

5.6 Forecasting Capabilities of NNF

In this sub-section we will formally compare the forecasting capabilities of the proposed NN filtering method against the filtering techniques of BK and HP.[10] We use the baseline DGP proposed in Guay and Saint-Amant (2005), described earlier, with $\boldsymbol{\sigma_\varepsilon/\sigma_\eta} = 0.01$, $\boldsymbol{\varphi}_1 = 1.2$, $\boldsymbol{\varphi}_2 = 0.40$, where all the filters seem to produce identical results. The comparison among the filtering techniques will be based on the most popular mean value measures for forecasting, i.e. (a) the mean squared errors (MSE);(b) The root mean squared errors (RMSE); (c) the mean absolute error (MAE); and (d) the mean absolute scale errors (MASE). Additionally, for the sake of robustenss, we will use a slightly different procedure to estimating the forecasting performance of the various models. More precisely, the initial estimation period is retained but the ending points of the estimation sample are shifted by one year resulting in a new estimation sample. These samples are used to estimate model parameters and produce forecasts.

[10] We would like to thank an anonymous referee for this suggestion.

Table 11 In sample forecast for a horizon of 5 years

Filtering technique	MSE	RMSE	MAE	MASE
HP	0.134	0.352	0.139	0.127
BK	0.130	0.350	0.132	0.132
NNF-1 node	0.072	0.251	0.069	0.055
NNF-optimal number of nodes via BIC	0.033	0.165	0.028	0.044

As can be seen, the proposed model's performance is superior in both, in-sample and out-of-sample forecasts, for a horizon of up to 5 years (Tables 11, 12).

6 Econometric Estimation

Our approach should now be confronted with real word economic data to assess its ability to model satisfactorily real word situations of interest. In this context, from an economic viewpoint, we will provide the estimation and visualization of such business cycles fluctuations for output in E15 using clustering in order to study the dynamics of the European economies' business cycles.

6.1 White Noise Analysis

In order to check whether the cyclical time series derived is indeed a cycle, we proceed by implementing a white noise test. The sample autocorrelation function measures how well a time series is correlated with its own past history. In order to test for autocorrelation we use the Ljung and Box (1978) test (Q-stat) which tests the null hypothesis of white noise for a maximum lag length k:

$$Q = n\left(n+2\right)\sum_{j=1}^{h}\frac{\hat{\rho}_j^2}{n-j} \tag{28}$$

where n is the sample size, $\hat{\rho}_j$ is the sample autocorrelation at lag j, and h is the number of lags being tested.

For significance level α, the critical region for rejection of the hypothesis of randomness is $Q > x_{1-a,h}^2$ where $x_{1-a,j}^2$ is the α-quantile of the chi-square distribution with h degrees of freedom. The alternative hypothesis is that at least one of these autocorrelations is non-zero, so that the series is not white noise. In case the null hypothesis is rejected, the time series is not white noise.

6.2 Cluster Analysis

As we know, cluster analysis is a methodology used to partition a set of observations into a distinct number of clusters so that all observations within a group are similar, while observations in different groups are not similar. Its main advantage is that it creates natural groups rather than classifying them on the basis of some *ad hoc* criterion.[11]

The technique of "fuzzy" clustering will be used, because unlike the so-called "hard" clustering approach that assigns each object to only one cluster, the fuzzy clustering approach is much more suitable to analyze data where increasing indeterminateness is present. This powerful technique, assigns membership coefficients (i.e. probabilities) which express the degree of "belongingness" of each object to each of the clusters, in a way that the highest coefficient indicates the cluster to which this country is most likely to belong.

Various strategies for the determination of the number of clusters have been proposed.[12] Probably, the most common method is k-means clustering (Hartigan and Wong 1978; MacQueen 1967). Its main advantage is that the distance between any two objects is not affected by the addition of new objects in the analysis. Thus, the method of fuzzy k-means clustering is adopted (Bezdec 1981). It enables the classification of data objects that populate some multi-dimensional space into a number of different groups (clusters). In our case the data objects are the time series of GDP cycles for each EU15 economy.[13]

Let $X = \{X_1, \ldots, X_K\}$ be a set of K data objects, each of them belonging to $A \subseteq R^N$. For a given number of clusters, the method of fuzzy C-means clustering is based on the minimization of an objective function defined as:

$$J(U, Y) = \sum_{k=1}^{K} \sum_{c=1}^{C} (u_{kc})^m \|x_k - y_c\|, \quad 1 < m < +\infty$$

where $U = \{u_{kc}\}$ is the participation matrix, defining a fuzzy partition of X, $\| \cdot \|$ is the inner product, $Y = \{Y_1, \ldots, Y_C\}$ is the set of the cluster centers, $Y_c \in A, c = \{1, \ldots, C\}$ and m is a parameter representing the weight attached to membership participation.

7 Empirical Results

The data on real G.D.P. in 2000 prices are annual and come from AMECO (European Commission). In the empirical analysis that follows we make use of the original (raw) data that came directly from the AMECO database without using any transformation i.e. logarithmic or exponential, so as to ensure that the initial crucial characteristics of the data in use will not be lost.

[11] Cluster analysis has often been applied to European data (see, e.g. Jacquemin and Sapir 1995; Artis and Zhang 1997, 1998a, b).

[12] See, among others, Bock (1974), Bozdogan (1993), Engelman and Hartigan (1969).

[13] In order to avoid biased inferences due to differences in the size of each economy, the cyclical components were standardized by their standard deviation.

Table 12 Out of sample forecast for a horizon of 5 years

Filtering techniques	MSE	RMSE	MAE	MASE
HP	0.314	0.560	0.310	0.289
BK	0.310	0.544	0.317	0.414
NNF-1 node	0.065	0.250	0.061	0.055
NNF-optimal number of nodes via BIC	0.087	0.281	0.085	0.044

7.1 Stationarity Results

The results, which are available upon request by the authors, suggest non-stationarity of the original times series. However, the NNF filtered components have all been found stationary.[14]

7.2 NNF Estimation

The estimated model's p values for are shown in Table 12 and are very satisfactory (Table 13).

The number of nodes m for each country has been selected using the AIC criterion which is, in general terms, consistent with previous studies (e.g. Michaelides et al. 2010). The values of the AIC are presented in Table 14.

7.3 Cyclical Components

In this section, we extract the cyclical components of real GDP series for each economy using the proposed NN specification as well as the most popular filters in the related literature, namely the BK and HP filters. In addition, we use the correlation coefficients so as to econometrically establish any potential relation between them.

In brief, the BK Filter (Baxter and King 1999) has been used in a large number of studies, as of yet (e.g. Agresti and Mojon 2001; Stock and Watson 1999; Massmann and Mitchell 2004). The BK filter is based on the idea of constructing a band-pass linear-filter that extracts a frequency range corresponding to the minimum and maximum frequency of the business cycle, c_t^{BK}. The algorithm consists of constructing two low-pass filters. The first passes through the frequency range $[0, \omega_{max}]$, denoted $\bar{a}(L)$, where L is the lag operator, and the second through the range $[0, \omega_{min}]$, denoted $\underline{a}(L)$. Subtracting these two filters, the ideal frequency response is obtained and the de-trended time series is:

$$c_t^{BK} = [\bar{a}(L) - \underline{a}(L)]y_t$$

14 We should note of course that the filters of BK and HP do not presuppose the use of stationary time series, i.e. Sowell MLE estimations. Despite the fact that in the case of NNF this is also true, we examined the stationarity characteristics of the cyclical components extracted for reasons of robustness. We would like to thank an anonymous referee for this comment.

Table 13 NNF estimates (*Note* t-statistic in parenthesis)

Country	α_0	δ	α_1	α_2	α_3	α_4	α_5	$R^2(adj)$
Greece	27.135	−0.036	3.084	0.000	0.000	0.000	0.000	0.984
	(6.025)	(−12.741)	(0.631)	(0.000)	(0.000)	(0.000)	(0.000)	
Italy	339.576	−0.084	16.493	23.981	0.000	0.000	0.000	0.997
	(26.714)	(−53.176)	(2.199)	(3.265)	(0.000	(0.000)	(0.000)	
Spain	188.836	−6.137	30.066	33.495	0.000	0.000	0.000	0.991
	(20.552)	(−11.083)	(4.373)	(5.167)	(0.000)	(0.000)	(0.000)	
Portugal	20.857	−3.288	3.368	0.000	0.000	0.000	0.000	0.992
	(14.191)	(−22.160)	(4.547)	(0.000)	(0.000)	(0.000)	(0.000)	
France	448.832	−0.028	19.061	17.545	25.061	0.000	0.000	0.998
	(34.714)	(−16.980)	(2.901)	(2.661)	(4.564)	(0.000)	(0.000)	
Belgium	86.862	−0.061	2.925	2.131	2.431	8.642	0.000	0.996
	(19.344)	(−14.102)	(2.029)	(1.568)	(1.612)	(6.045)	(0.000)	
Luxembourg	5.117	−0.030	3.036	0.000	0.000	0.000	0.000	0.976
	(4.061)	(−19.207)	(4.778)	(0.000)	(0.000)	(0.000)	(0.000)	
Netherlands	135.502	−0.027	4.779	2.319	26.127	0.000	0.000	0.991
	(21.881)	(−34.311)	(1.514)	(0.803)	(9.834)	(0.000)	(0.000)	
Austria	58.574	−0.023	4.913	2.585	0.067	0.000	0.000	0.996
	(28.112)	(−93.015)	(3.596)	(2.167)	(0.063)	(0.000)	(0.000)	
Ireland	39.106	−0.036	11.371	24.806	0.000	0.000	0.000	0.986
	(4.283)	(−29.921)	(3.129)	(5.645)	(0.000)	(0.000)	(0.000)	
United Kingdom	383.043	−0.025	71.866	0.000	0.000	0.000	0.000	0.992
	(22.769)	(−54.179)	(7.122)	(0.000)	(0.000)	(0.000)	(0.000)	
Germany	676.411	-0.083	107.336	0.000	0.000	0.000	0.000	0.996
	(39.172)	(−3.581)	(12.355)	(0.000)	(0.000)	(0.000)	0.000	
Denmark	433.773	−0.021	34.727	15.278	0.000	0.000	0.000	0.996
	(27.082)	(−26.386)	(5.577)	(1.073)	(0.000)	(0.000)	(0.000)	
Finland	61.566	−0.038	14.190	6.530	2.552	3.871	0.000	0.989
	(10.098)	(−59.846)	(8.435)	(3.819)	(1.545)	(2.329)	(0.000)	
Sweden	977.080	−0.053	82.872	27.884	103.491	0.000	0.000	0.991
	(36.785)	(−17.024)	(3.891)	(1.820)	(4.403)	(0.000)	(0.000)	

where y_t is the original time series at hand. Based on the literature, we make use of "Burns and Mitchell" settings for business cycle frequency range, i.e. two (2) to eight (8) years and a moving average of three (3) years.

Next, we turn to the Hodrick–Prescott (HP) filter which has also been extensively used in the literature, see, among others, Artis and Zhang (1997) and Dickerson et al. (1998).The filter is based on the minimization of an arbitrary trend specification. More specifically, the trend is obtained by minimizing the fluctuations of the actual data around it, i.e. by minimizing the following function:

Table 14 AIC criterion

Country	AIC (1 node)	AIC (2 nodes)	AIC (3 nodes)	AIC (4 nodes)	AIC (5 nodes)
Greece	8.745	9.246	9.137	8.770	9.413
Italy	10.480	10.352	10.600	10.620	11.200
Spain	11.174	9.774	11.175	10.272	9.837
Portugal	6.150	6.378	6.308	6.460	6.364
France	9.894	9.989	9.857	10.387	10.700
Belgium	7.358	7.279	7.447	6.880	6.961
Luxemburg	5.627	5.667	6.100	5.870	5.820
Netherlands	9.411	8.648	8.387	9.280	8.764
Austria	6.766	6.657	6.633	6.710	6.715
Ireland	9.173	9.127	9.345	9.730	9.420
United	9.849	10.709	11.591	10.863	10.179
Germany	10.904	12.186	12.185	12.065	11.643
Denmark	10.550	10.350	10.626	10.709	10.913
Finland	7.195	7.798	7.168	6.805	7.085
Sweden	13.213	12.779	11.920	12.607	11.979

$$\sum_{t=1}^{T} \left(y_t - y_t^*\right)^2 - \lambda \sum_{t=2}^{T-1} \left[\left(y_{t+1}^* - y_t^*\right) - \left(y_t^* - y_{t-1}^*\right)\right]$$

where y_t^* is the long-term trend of the variable y and the coefficient $\lambda > 0$ determines the smoothness of the long-term trend. The smoothing parameter used for yearly data is equal to $\lambda = 6.25$.

Now, in order to get a visual inspection regarding the co-movement of the cycles extracted via our proposed NN filter and the traditional filtering techniques we computed the cyclical components of real GDP series for every economy in our dataset, see Fig. 1.

It can be easily inferred that the NNF cycles are quite very close to the BK cycles. The HP cycle often fails to follow closely their pattern. Also the linear cycle is clearly not able to follow closely their pattern. Of course, our findings based on visual inspection of the cycles are also confirmed by the correlation coefficients among the various cycles. See Fig. 2.

Also, the correlations of the GDP cycles among EU15 countries -for each filtering technique- are presented compactly in Fig. 3. Again, we can see that the NNF and the BK produce similar results.

7.4 White Noise

The results of the Ljung and Box test, which are available upon request, indicate a rejection of the null hypothesis of white noise for all the countries' cycles. In other words, the existence of cyclical regularities is a valid hypothesis from a statistical viewpoint for all EU15 countries.

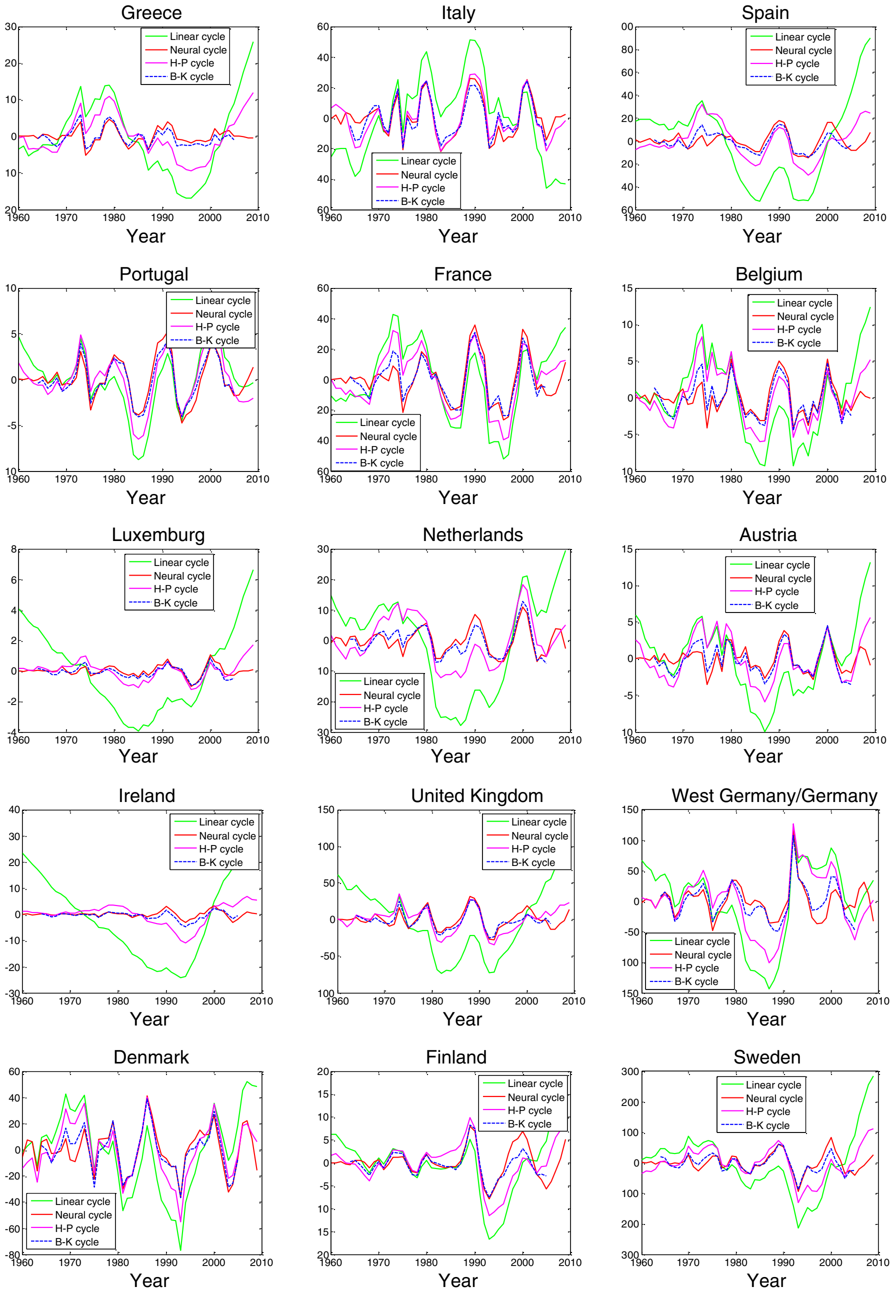


Fig. 1 Cyclical components of real GDP series

Greece

	Neural	HP	BK	Linear
Neural	1			
HP	0.438	1		
BK	0.874	0.734	1	
Linear	0.278	0.945	0.581	1

Italy

	Neural	HP	BK	Linear
Neural	1			
HP	0.926	1		
BK	0.943	0.964	1	
Linear	0.577	0.793	0.690	1

Spain

	Neural	HP	BK	Linear
Neural	1			
HP	0.542	1		
BK	0.840	0.835	1	
Linear	0.284	0.765	0.481	1

Portugal

	Neural	HP	BK	Linear
Neural	1			
HP	0.828	1		
BK	0.966	0.917	1	
Linear	0.687	0.937	0.785	1

France

	Neural	HP	BK	Linear
Neural	1			
HP	0.801	1		
BK	0.956	0.892	1	
Linear	0.636	0.953	0.748	1

Belgium

	Neural	HP	BK	Linear
Neural	1			
HP	0.651	1		
BK	0.905	0.848	1	
Linear	0.472	0.901	0.643	1

Luxemburg

	Neural	HP	BK	Linear
Neural	1			
HP	0.671	1		
BK	0.899	0.819	1	
Linear	0.174	0.557	0.208	1

Netherlands

	Neural	HP	BK	Linear
Neural	1			
HP	0.565	1		
BK	0.920	0.807	1	
Linear	0.282	0.836	0.506	1

Austria

	Neural	HP	BK	Linear
Neural	1			
HP	0.611	1		
BK	0.899	0.846	1	
Linear	0.418	0.790	0.565	1

Ireland

	Neural	HP	BK	Linear
Neural	1			
HP	0.280	1		
BK	0.803	0.774	1	
Linear	0.019	0.689	0.406	1

United Kingdom

	Neural	HP	BK	Linear
Neural	1			
HP	0.782	1		
BK	0.928	0.884	1	
Linear	0.333	0.678	0.437	1

Germany

	Neural	HP	BK	Linear
Neural	1			
HP	0.570	1		
BK	0.871	0.820	1	
Linear	0.423	0.920	0.669	1

Denmark

	Neural	HP	BK	Linear
Neural	1			
HP	0.801	1		
BK	0.946	0.903	1	
Linear	0.533	0.879	0.659	1

Finland

	Neural	HP	BK	Linear
Neural	1			
HP	0.667	1		
BK	0.909	0.841	1	
Linear	0.500	0.853	0.656	1

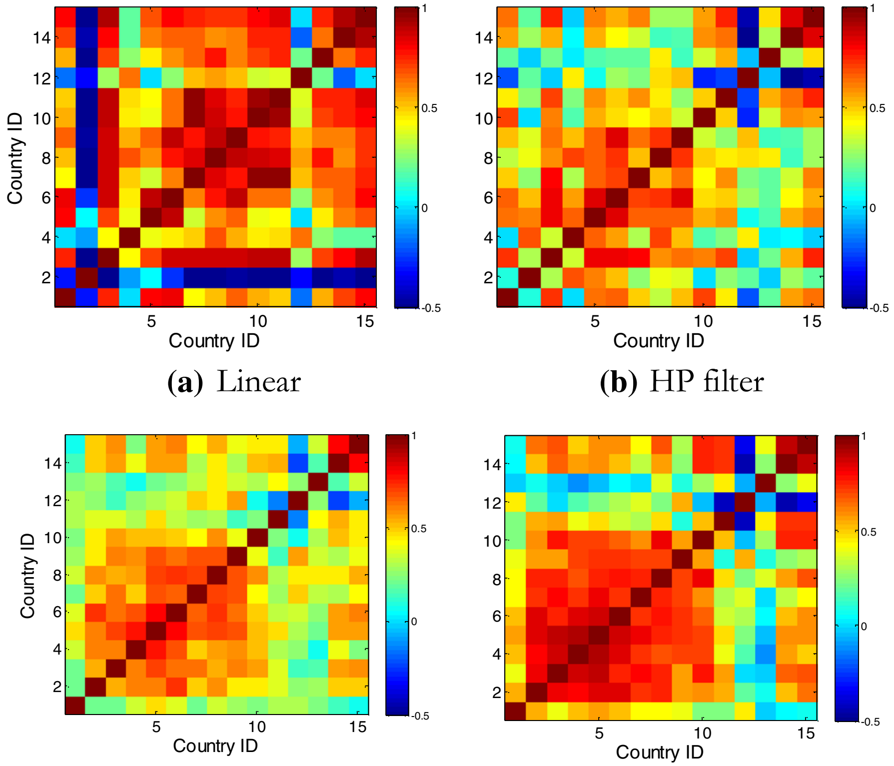
Sweden

	Neural	HP	BK	Linear
Neural	1			
HP	0.607	1		
BK	0.888	0.819	1	
Linear	0.344	0.794	0.518	1

Fig. 2 Correlation coefficients among cyclical components

7.5 Mean Value

We have proved that the cycle produced by the proposed NNF filter has a mean value equal to zero. In what follows, we report the mean value of the cycles produced by NNF based on the real world data for EU15. We can see in Table 15 that the mean value



(a) Linear (b) HP filter

(c) BK filter (d) Neural filter

Fig. 3 Correlation matrices for cycles between EU countries

of the cyclical component produced by NNF is not significantly different from zero (0), a result which implies that the rigorous theoretical framework developed earlier, even in its strict form, is fully consistent with real world empirical evidence on EU15 business cycles.

7.6 Clusters

Lastly, we conducted *k*-means (fuzzy) clustering in order to determine the groups of countries that are formed in Europe (1960–2014). We conducted the analysis using the NNF method with the secular components. The participation matrix was formed for the optimum of five (5) distinct clusters (Table 16).

An interesting empirical finding is the distinction in core and periphery counties in EU-15, a finding which is reported in the majority of studies in the relevant literature. Among others, the existence of a core of countries with similar characteristics has been documented by Bayoumi and Eichengreen, (1993), Dickerson et al. (1998), (Artis and Zhang 1998a, b), Crowley and Christi (2003), Massmann and Mitchell

Table 15 Mean values for EU15 NNF business cycles

Country		Mean value of C_t	Observed mean value of C_t	p value
1	Greece	0.00	$-1.6e-14$	1
2	Italy	0.00	$8.1e-14$	1
3	Spain	0.00	$6.9e-14$	1
4	Portugal	0.00	$9.2e-15$	1
5	France	0.00	$-1.3e-13$	1
6	Belgium	0.00	$2.8e-14$	1
7	Luxemburg	0.00	$9.9e-15$	1
8	Netherlands	0.00	$-5.3e-14$	1
9	Austria	0.00	$2.2e-14$	1
10	Ireland	0.00	$5.6e-15$	1
11	United	0.00	$1.3e-13$	1
12	Germany	0.00	$2.0e-13$	1
13	Denmark	0.00	$-6.1e-14$	1
14	Finland	0.00	$-1.4e-14$	1
15	Sweden	0.00	$4.5e-15$	1

p values correspond to the robust (HAC) t-statistic of regressing actual minus predicted on a constant

Table 16 NNF participation matrix

Country	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Austria	0.03	0.71	0.08	0.05	0.13
Belgium	0.03	0.04	0.84	0.01	0.08
Denmark	0.82	0.03	0.07	0.02	0.06
Finland	0.31	0.05	0.32	0.06	0.27
France	0.07	0.04	0.76	0.01	0.13
Germany	0.01	0.92	0.02	0.01	0.03
Greece	0.11	0.21	0.56	0.03	0.09
Ireland	0.00	0.00	0.00	1.00	0.00
Italy	0.09	0.10	0.14	0.05	0.62
Luxemburg	0.02	0.92	0.03	0.01	0.02
Netherlands	0.03	0.04	0.13	0.14	0.65
Portugal	0.03	0.03	0.04	0.03	0.87
Spain	0.64	0.12	0.16	0.02	0.06
Sweden	0.90	0.02	0.04	0.01	0.04
UnitedKingdom	0.12	0.04	0.69	0.02	0.13

(2004), Camacho et al. (2006) and Concaria and Soares (2009). See also Canzoneri et al. (1996), Bayoumi and Eichengreen (1997a, b), Taylor (1995) and Papageorgiou et al. (2010). Table 17 summarizes the first best clustering results as reported by the participation matrix presented earlier.

Table 17 Clusters based on NNF filtering

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Denmark	Austria	Belgium	Ireland	Italy
Spain	Germany	Finland		Netherlands
Sweden	Luxemburg	France		Portugal
		Greece		
		UK		

The results show that there is a core—periphery distinction, although the core seems to split into two main groups, one with Germany–Austria and another with France–Belgium and the UK. The periphery is also split (Italy and Ireland-Spain–Portugal in different groups). This pattern does not conform to a "classical" division and shows that the definition of core and periphery is more involved. In the light of recent developments after the sub-prime crisis it seems, however, that this view has indeed considerable merit. The explosion of debt hit the different economies in a different way. First, it brought to the foreground the differing Franco–German views on the future of the Eurozone. Second, it emphasized the different patterns of reaction to debt-related problems in Greece, Spain–Portugal and Italy. The fact that Ireland forms a group of its own also emerged when foreign investment exploded in Ireland and the country recovered fast from the subprime crisis. In the light of the recent crisis, the exact position of Greece in the economic 'map' of Europe, is not very clear. However, in the 1960s and part of the 1970s the country was growing fast and its industrial base was developing rapidly. As a result, it is not surprising that it is part of the core along with the UK and France. In the aftermath of the sub-prime crisis though, it is even less surprising that it belongs to the same group with the UK, which is *not* part of the Eurozone. Therefore, the clustering approach reflects various factors which are consistent, for the most part, also with what we know after the subprime crisis.

8 Conclusions

It is widely hailed that a major limitation in examining the properties of a time series in a business cycles framework is the identification of the type of trend that a time series exhibits (Woitek 1998). This is a severe limitation of the available techniques, since most filters in the literature exhibit the desirable properties only with pre-specified trend forms. In a seminal paper, Pedersen (2001) created a measure in order to quantify the level of distortion of the most popular filters. According to the results, HP and BK seem to have a quite good but sub-optimum performance, under controlled circumstances (see also Mise et al. 2005). Thus, it is apparent that in order to judge the robustness of the results of a filtering process, so far, more than one filtering techniques should be applied in a time series (Canova 1998). This increases our computational costs. On the other hand, our proposed approach allows researchers to overcome this problem since it allows, by design, the dataset itself to fit the model and not *vice versa*, as most filtering techniques suggest.

Also, it is widely accepted that business cycles in *Economics* have non-linear characteristics. As a result, relevant non-linear specifications have to be employed in order to achieve a globally credible form independent of the series in use. Thus far, for the extraction of the cyclical component of a time series, researchers assume that the trend specification of the time series follows a certain pattern, i.e. linear, exponential etc. Nevertheless, this is an ad-hoc assumption, which totally ignores the inherent characteristics of the time series at hand. To this end, in this paper, we formally establish a novel methodological framework that takes into consideration the inherent non-linearities of the time series. In this context, our approach that is intrinsically non-linear, should be considered as an appropriate alternative that is not affected by the potential nonlinearities of the dataset at hand, the so-called "neglected non-linearities" problem (Lee et al. 1993). See also Kiani (2009a, b).

Moreover, we have demonstrated formally that the cycles produced by the proposed technique have a mean value equal to zero. In other words, business cycles as deviations from trend are disturbances from a growth path (positive or negative) that lead, sooner or later, to a return to the growth path. The economic intuition of this finding is that growth occurs despite a business cycle process and as a result these cyclical fluctuations are no barrier to economic growth (positive or negative).

Next, using relevant DGPs, we have demonstrated that our approach is superior to HP and BK regarding the generated distortionary effects and the ability to operate in various frequencies, including changes in volatility, amplitudes and phase. Also, while keeping the structure of the proposed approach relatively simple, it is nevertheless capable of addressing very satisfactorily the case of stochastic trend, in the sense that the generated distortionary effects in the near unit root case are minimal and, by all means, considerably fewer than those generated by HP and BK. In this context, a relevant procedure for the econometric estimation of the NNF has been developed as a simple seven-step algorithm which relies on standard techniques and all relevant measures can be computed routinely.

Finally, the NNF was applied to real GDP time series for EU15 countries. An empirical finding of great interest is the existence of a core-periphery distinction in Europe which is consistent with the findings by other researchers. More precisely, five (5) clusters were formed based on the powerful fuzzy clustering technique.

The empirical results, which are consistent with the rigorous theoretical framework developed in this work, even in its strictest form, suggest that the proposed globally flexible Neural Network Filter (NNF) is a useful approach for expanding conventional filter theory.

Acknowledgements We are indebted to three anonymous Referees and the Editor-in-Chief, Hans Amman, for their diligent reading of the manuscript and for the constructive feedback. The first author (P.G.M.) would also like to thank Alexandros Eskenazis for a fruitful discussion on a previous version of this paper.

Appendix

Definition 1 (*Trend set*) Consider g_{t_j}, $t \in T \subseteq \mathbb{R}^+$, $j \in J \subseteq \mathbb{R}$ representing the trend of a time series $x_{t_j} \forall j \in J$, such that $g_{t_j} \in \mathbb{R}$, $\forall j \in J$. Without loss of generality, let $\bigcup_{t_j} g_{t_j} = \{g_{t_j} : g_{t_j} \textit{is the trend of } x_{t_j} \forall j \in J\} \subseteq \mathbb{R}$ be the trend set which is assumed to be closed and bounded.

Definition 2 (*Time series as random variable*) A time series model for the observed data x_{t_j}, $j \in J$ is a specification of the joint distributions, or only the means and covariances, of a sequence of random variables $\{X_t\}_{t \in T}$ of which $\{x_{t_j}\}_{t \in T}$ is postulated to be a realization.

Definition 3 (*Time series set*) Consider x_{t_j}, $j \in J$ an arbitrary macroeconomic time series, such that $x_{t_j} \in \mathbb{R} \forall t \in T \subseteq \mathbb{R}^+$. Without loss of generality, let $\bigcup_{j \in J} x_{t_j} \subseteq \mathbb{R}$ be the time series set.

Theorem 2

Proof From Lemma 1 the trend set is compact. From Lemma 2 any function of the form : $\mathrm{F(t)} \equiv \mathrm{d} + \mathrm{ct} + \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{a_i}\, \varphi\left(\mathrm{b_i t}\right)$, $\varphi_\mathrm{i}, \mathrm{b_i}, \mathrm{d} \in \mathbb{R}$, $\mathrm{c} \in \mathbb{R} - \{0\}\}$ is non-constant, bounded and continuous. Then, from Theorem 1, the family: $\mathcal{F} = \{\mathrm{F(t)} \in \mathrm{C}\left(\bigcup_{j \in J} g_{t_i}\right) :$ $\mathrm{F(t)} \equiv \mathrm{d} + \mathrm{ct} + \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{a_i} \varphi\left(\mathrm{b_i t}\right)$, $\varphi_\mathrm{i}, \mathrm{b_i}, \mathrm{d}, \mathrm{c} \in \mathbb{R}\}$ is dense in $\mathrm{C}\left(\bigcup_{j \in J} g_{t_j}\right)$. □

Theorem 3 (Linear time trend as degenerate form of NNF)

Proof Let $x_{t_i}, i \in I$ be a time series and let $\bar{m} \in \{1, ..M\}$. Then, $\exists \overline{\beta_{\bar{m}_i}} \in \mathbb{R}^{\bar{m}}$, $\overline{a_{\bar{m}_i}} \in \mathbb{R}^{\bar{m}}+$, $\overline{a_{0\bar{m}_i}} \in \mathbb{R}$ and $\overline{\delta_{\bar{m}_i}} \in \mathbb{R}$ such that the trend of the macroeconomic time series be given by the following expression:

$$g_t = \overline{a_{0\bar{m}_i}} + \overline{\delta_{\bar{m}_i t}} + \sum_{\mathrm{k}=1}^{\bar{m}} \overline{\alpha_k} \varphi\left(\overline{\beta_k} t\right), \quad \forall i \in I$$

Now, since I is a compact subset of $\mathbb{R}$ then it is closed and bounded and there exists $i_0 \in I$ such that $\overline{\beta_{\bar{m}_{i_0}}} = \max\{\overline{\beta_{\bar{m}_i}} \in \mathbb{R}^{\bar{m}}\}$. For, this $\overline{\beta_{\bar{m}_{i_0}}}$ we have that $\overline{\delta_{\bar{m}_{i_0}}} = \max\left\{\overline{\delta_{\bar{m}_i}}, \bar{m}_i \in \{1, \ldots M\}\right\}$, while the trend of this macroeconomic time series is given by the expression:

$$g_t = \overline{a_0} + \overline{\delta_{\bar{m}_{i_0}}} t + \sum_{\mathrm{k}=1}^{\bar{m}_{i_0}} \overline{\alpha_k} \varphi\left(\overline{\beta_k} t\right), \quad i_0 \in I$$

But, since: $\overline{\beta_{\bar{m}_{i_0}}} = \max\left\{\overline{\beta_{\bar{m}_i}}, \overline{\beta_{\bar{m}_i}} \in \mathbb{R}^{\bar{m}}\right\}$, then: $\sum_{\mathrm{k}=1}^{\bar{m}_{i_0}} \overline{\alpha_k} \varphi\left(\overline{\beta_k} t\right) = \sum_{\mathrm{k}=1}^{\bar{m}_{i_0}} \overline{\alpha_k}$. In view of φ being increasing monotonic and $\varphi : \mathbb{R} \to [0, 1]$, we have that: $g_t = \overline{a_0} + \overline{\delta_{\bar{m}_{i_0}}} t + \sum_{\mathrm{k}=1}^{\bar{m}_{i_0}} \overline{\alpha_k}$.

Hence, the trend approximation is equal to the linear trend, i.e. $g_t = \gamma + \delta t$, where: $\gamma = \overline{a_0} + \sum_{\mathrm{k}=1}^{\bar{m}_{i_0}} \overline{\alpha_k}$. □

Theorem 4 (Mean value of the cycle is zero)

Proof Let x_{t_j}, $j \in J$ be a time series whose cyclical component is given by the expression:

$$c_{t_j} = x_{t_j} - g_{t_j}, \forall j \in J \text{ and } t \in T.$$

Let $\{p_{t_j}\}_{j\in J}$ be the respective probability measure assigned on each cyclical times series $c_{t_j} \forall t \in T$.

Now, provided that: $\sum_{t\in T}(x_{t_j} - g_{t_j}), p_{t_j}$ converges absolutely, i.e. $\sum_{t\in T} |(x_{t_j} - g_{t_j}), p_{t_j}| < \infty$, the expected value of our cyclical component is given by the following expression:

$$E\left(c_{t_j}\right) = E\left((x_{t_j} - g_{t_j},)\right) = \sum_{t\in T}(x_{t_j} - g_{t_j},)p_{t_j} \forall j \in J, \quad t \in T \text{ and } p_{t_j} \in [0, 1] \tag{29}$$

But, since $\bigcup_{t_i} g_{t_i}$ is a dense subset on $\bigcup_{t_j} x_{t_j}$, then by the definition of density we have that $\forall \varepsilon > 0$ and $\forall x_{t_j}$, $j \in J$ there exists g_{t_i}, $i \in I$ such that $\left|x_{t_j} - g_{t_i}\right| < \varepsilon \forall j \in J$ *and* $\forall i \in I$. Thus, $\forall \left\{\mathrm{p_{t_j}}\right\}_{\mathrm{j\in J}}$ in a relevant probability space, we have that: $|(x_{t_j} - g_{t_j})p_{t_j}| < \varepsilon p_{tj} < \varepsilon \, \forall j \in J$ and $\forall p_{t_j} \in [0, 1]$.

But, without loss of generality, for $\varepsilon_t = 1 - \frac{1}{2^t}, \forall t \in T$ we have that: $\sum_{t\in T}(1 - \frac{1}{2^t}) < \infty$.

Hence: $\sum_{t\in T} \left|(x_{t_j} - g_{t_j},)p_{t_j}\right| < \infty$ (A.14) and, therefore, Eq. (29) defines the expected value of the cyclical component of the time series.

Thus:

$$\left(c_{t_j}\right) = E\left((x_{t_j} - g_{t_j},)\right) = \sum_{t\in T}(x_{t_j} - g_{t_j},)p_{t_j} \forall j \in J, t \in T \tag{30}$$

But: $\sum_{t\in T}(x_{t_j} - g_{t_j},)p_{t_j} = (x_{1_j} - g_{1_j},)p_{1_j} + \cdots + (x_{t_j} - g_{t_j},)p_{t_j} + \ldots$ and $(x_{1_j} - g_{1_j},)p_{1_j} + \cdots + (x_{t_j} - g_{t_j},)p_{t_j} < \varepsilon_1 p_{1_j} + \cdots + \varepsilon_T p_{T_j} + \ldots$, $\forall j \in J$, $t \in T$ because of the density of $\bigcup_{t_i} g_{t_i}$ on $\bigcup_{t_j} x_{t_j}$ which implies: $\left|x_{t_j} - g_{t_i}\right| < \varepsilon \forall j \in J$ *and* $\forall i \in I$.

However: $\left|x_{t_j} - g_{t_i}\right| < \varepsilon_t \Leftrightarrow -\varepsilon_t < x_{t_j} - g_{t_i} < \varepsilon_t$, $\forall j \in J$, $\forall t \in T$ and $\forall i \in I$

$$-\sum_{t\in T} \varepsilon_t p_{t_j} \leq \sum_{t\in T} \left(x_{t_j} - g_{t_j}\right) p_{t_j} \leq \sum_{t\in T} \varepsilon_t p_{t_j} \forall j \in J \text{ and } \forall p_{t_j} \in [0, 1]$$

Now, without loss of generality, for $\varepsilon_t = 1 - \frac{1}{2^t}$, $\forall t \in T$ we have that:

$$\begin{aligned} -\sum_{t\in T} \left(1 - \frac{1}{2^t}\right) p_{t_j} &\leq \sum_{t\in T} \left(x_{t_j} - g_{t_j}\right) p_{t_j} \\ &\leq \sum_{t\in T} \left(1 - \frac{1}{2^t}\right) p_{t_j} \forall j \in J \text{ and } \forall p_{t_j} \in [0, 1] \end{aligned} \tag{31}$$

But, given that $1 - \frac{1}{2^t} > 0$ and $p_{tj} \in [0, 1]$, we have that:

$$\begin{aligned} \sum_{t\in T} \left(1 - \frac{1}{2^t}\right) p_{t_j} &\leq \sum_{t\in T} \left(1 - \frac{1}{2^t}\right) \sum_{t\in T} p_{t_j} \\ &= \sum_{t\in T} \left(1 - \frac{1}{2^t}\right) \to 0 \text{ since } : \sum_{t\in T} p_{t_j} = 1 \end{aligned} \tag{32}$$

Similarly: $-\sum_{t\in T} \left(1 - \frac{1}{2^t}\right) p_{t_j} \to 0$

Hence, given that: $-\sum_{t\in T}\varepsilon_t p_{t_j} \le \sum_{t\in T}\left(x_{t_j}-g_{t_j}\right)p_{t_j} \le \sum_{t\in T}\varepsilon_t p_{t_j}\forall j \in J$, $\forall p_{t_j} \in [0,1]$ and that: $-\sum_{t\in T}\left(1-\frac{1}{2^t}\right)p_{t_j} \to 0$ and $\sum_{t\in T}\left(1-\frac{1}{2^t}\right)p_{t_k} \to 0$ we get:

$$E\left(c_{t_j}\right)=0, \forall j \in J, t \in T \text{ and } \forall p_{t_j} \in [0,1] \tag{33}$$

□

Theorem 5 (Mean value of the NNF cycle is zero)

Proof Based on Theorem 2, $\mathcal{F} = \{F(t) \in C(X) : F(t) \equiv d + ct + \sum_{i=1}^{N} a_i\varphi\left(\beta_i t\right), \alpha_i, \beta_i, d, c \in \mathbb{R}\}$ is dense in $\bigcup_{t_i} g_{t_i}$. Hence, in view of Theorem 4, we have that: $\left(c_{t_j}\right)=0\forall j \in J$. □

References

Adya, M., & Collopy, F. (1998). How effective are neural networks at forecasting and prediction? A review and evaluation. *Journal of Forecasting*, *17*, 481–495.

Agresti, A., & Mojon, B. (2001). Some stylised facts on the euro area business cycle. ECB Working Paper 95.

Aminian, F., Suarez, E. D., Aminian, M., & Waltz, D. T. (2006). Forecasting economic data with neural networks. *Computational Economics*, *28*(1), 71–88.

Anderson, H. M., & Ramsey, J. B. (2002). U.S. and Canadian industrial production indices as coupled oscillators. *Journal of Economic Dynamics and Control*, *26*(1), 33–67.

Andreano, M. S., & Savio, G. (2002). Further evidence on business cycle asymmetries in G7 countries. *Applied Economics*, *34*(7), 895–904.

Andreasen, M. M. (2011). Non-linear DSGE models and the optimized central difference particle filter. *Journal of Economic Dynamics and Control*, *35*(10), 1671–1695.

Andreasen, M. M. (2013). Non linear DSGE models and the central difference Kalman filter. *Journal of Applied Econometrics*, *28*(6), 929–955.

Artis, M., & Zhang, W. (1997). International business cycle and the ERM: Is there a European business cycle? *International Journal of Finance and Economics*, *2*, 1–16.

Artis, M., & Zhang, W. (1998a). Core and periphery in EMU: A cluster analysis, E*UI working paper RSC* No. 98/37.

Artis, M., & Zhang, W. (1998b). Membership of EMU: A fuzzy clustering analysis of alternative criteria, *EUI working paper RSC* no. 98/52.

Balke, N. S., & Fomby, T. B. (1994). Large shocks, small shocks, and economic fluctuations: outliers in macroeconomic time series. *Journal of Applied Econometrics*, *9*(2), 181–200.

Baxter, M., & King, R. G. (1999). Measuring business cycles: Approximate band-pass filters for economic time series. *Review of Economic and Statistics*, *81*(4), 575–593.

Baum F.C. (2006). *Introduction to modern eonometrics using Stata*. Texas: Stata Press.

Bayoumi, T., & Eichengreen, B. (1993). Shocking aspects of European monetary integration. In F. Torres & F. Giavazzi (Eds.), *Adjustment and growth in the European Monetary Union*. Cambridge: Cambridge University Press.

Bayoumi, T., & Eichengreen, B. (1997a). Ever closer to heaven? An optimum-currency-area index for European countries. *European Economic Review*, *41*, 761–770.

Bayoumi, T., & Eichengreen, B. (1997b). Optimum currency areas and exchange rate volatility; theory and evidence compared. In B. Cohen (Ed.), *International trade and finance: New frontiers for research: Essays in honour of Peter Kenen*. Cambridge: Cambridge University Press.

Beaudry, P., & Koop, G. (1993). Do recessions permanently change output? *Journal of Monetary Economics*, *31*(2), 149–163.

Bezdec, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press.

Bidarkota, P. V. (1999). Sectoral investigation of asymmetries in the conditional mean dynamics of the real US GDP. *Studies in onlinear Dynamics and Econometrics*, *3*, 191–200.

Binner, J. M., Gazely, A. M., & Chen, S. (2002). Financial innovation and Divisia money in Taiwan: A neural network approach. *The European Journal of Finance*, *8*, 238–247.
Binner, J. M., Gazely, A. M., Chen, S., & Chie, B. (2004). Financial innovation and Divisia money in Taiwan: Comparative evidence from neural network and vector error-correction forecasting models. *Contemporary Economic Policy*, *22*(2), 213–224.
Binner, J. M., Bissoondeeal, R. K., Elger, T., Gazely, A. M., & Mullineux, A. W. (2005). A comparison of linear forecasting models and neural networks: An application to Euro inflation and Euro Divisia. *Applied Economics*, *37*(6), 665–680.
Bishop, C. M. (1995). Neural networks for pattern recognition. New York: Oxford University Press.
Brockett, P., Cooper, W., Golden, L., & Pitaktong, U. (1994). A neural network method for obtaining an early warning of insurer insolvency. *The Journal of Risk and Insurance*, *61*(3), 402–424.
Bock, H. H. (1974). *AutomatischeKlassifikation (Clusteranalyse)*. Goettingen: Vandenhoek&Ruprecht.
Bozdogan, H. (1993). Choosing the number of component clusters in the mixture model using a new informational complexity criterion of the inverse Fisher information matrix. In O. Opitz, B. Lausen, & R. Klar (Eds.), *Information and classification* (pp. 40–54). Berlin: Springer.
Brazili, A., & Siltzia, B. (2003). Risk related non-linearities is exchange rates: Evidence from a panel of Central and Eastern European Countries. *Open Economies Review*, *14*, 135–155.
Brunner, A. D. (1997). On the dynamic properties of asymmetric models of real GNP. *The Review of Economics and Statistics*, *79*(2), 321–352.
Burns, A. F., & Mitchell, W. C. (1946). *Measuring business cycles*. New York: National Bureau of Economic Research.
Camacho, M., Perez-Quiros, G., & Saiz, L. (2006). Are European business cycle close enough to be just one?*CEPR discussion papers*, No. 4824.
Campbell, J. Y. & Perron, P. (1991). Pitfalls and opportunities: What macroeconomists should know about unit roots. In O. Blanchard, & S. Fischer (Eds.), *NBER Macroeconomics Annual 1991*, Vol. 6.
Canova, F. (1998). Detrending and business cycle facts. *Journal of Monetary Economics*, *41*, 475–512.
Canzoneri, M., Valles, J., & Vinals, J. (1996). Do exchange rates move to address international macroeconomic imbalances? *CERP discussion papers*, No. 1948. London Venter of Economic Policy Research.
Chan, N. H., & Genovese, R. C. (2001). A comparison of linear and nonlinear statistical techniques in performance attribution. *IEEE Transactions on Neural Networks*, *12*(4), 922–928.
Chang, Y., Miller, I. J., & Park, J. Y. (2009). Extracting a common stochastic trend: Theory with some applications. *Journal of Econometrics*, *150*, 231–247.
Clements, M. P., & Krolzig, H.-M. (2004). Can regime-switching models reproduce the business cycle features of US aggregate consumption, investment and output? *International Journal of Finance & Economics*, *9*(1), 1–14.
Concaria, L. A., & Soares, M. J. (2009). Business cycle synchronization across the Euro area: A wavelet analysis, *NIPE* working papers.
Creal, D., Koopman, S. J., & Zivot, E. (2010). Extracting a robust US business cycle using a time-varying multivariate model-based bandpass filter. *Journal of Applied Econometrics*, *25*(4), 695–719.
Crowley P., & Christi, C. (2003). European Union Studies Association (EUSA), Biennial Conference, (8th), March 27–29.
Dickerson, A., Gibson, H., & Tsakalotos, E. (1998). Business cycle correspondence in the European Union. *Economica*, *25*, 51–77.
Diebold, F. X., & Rudebusch, G. D. (1989). Long memory and persistence in aggregate output. *Journal of Monetary Economics*, *24*(2), 189–209.
Engelman, L., & Hartigan, J. A. (1969). Percentage points of a test for clusters. *Journal of American Statistical Association*, *64*, 1647.
Falk, B. (1986). Further evidence on the asymmetric behavior of economic time series over the business cycle. *Journal of Political Economy*, *94*(5), 1096–1109.
Faraway, J., & Chatfield, C. (1998). Time series forecasting with neural networks: A comparative study using the airline data. *Applied Statistics*, *47*(2), 231–250.
Gencay, R. (1999). Linear, non-linear and essential foreign exchange rate prediction with simple technical trading rules. *Journal of International Economics*, *47*(1), 91–107.
Giordani, P., Kohn, R., & van Dijk, D. (2007). A unified approach to nonlinearity, structural change, and outliers. *Journal of Econometrics*, *137*, 112–133.
Guarin A., Liu, X., & Ng, W. L. (2013). Recovering default risk from CDS spreads with a nonlinear filter. *Journal of Economic Dynamics and Control* (forthcoming October 2013).

Guay, A., & Saint-Amant, P. (2005). Do the Hodrick-Prescott and Baxter-King filters provide a good approximation of business cycles? Annalesd'Economieet de Statistique. *ENSAE, issue*, *77*, 133–155.
Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, *45*(1–2), 39–70.
Hanke, M. (1999). Neural networks versus Black-Scholes: An empirical comparison of the pricing accuracy of two fundamentally different option pricing methods. *Journal of Computational Intelligence in Finance*, *5*, 26–34.
Hartigan, J. A., & Wong, M. A. (1978). Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics*, *28*, 100–108.
Harvey, A. C., & Todd, P. H. (1983). Forecasting economic time series with structural and BoxJenkins models: A case study. *Journal of Business and Economic Forecasting*, *1*, 299–307.
Harvey, A. C., & Jaeger, A. (1993). Detrending: Stylized facts and the business cycle. *Journal of Applied Econometrics*, *8*, 231–247.
Haykin, S. (1999). *Neural Networks*. New Jersey: Prentice-Hall.
Hess, G. D., & Iwata, S. (1997). Asymmetric persistence in GDP? A deeper look at depth. *Journal of Monetary Economics*, *40*(3), 535–554.
Hillmer, S. C., & Tiao, G. C. (1982). An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association*, *77*, 63–70.
Hodrick, R. J., & Prescott, E. C. (1981). Postwar U.S. business cycles: An empirical investigation. Carnegie Mellon University, discussion paper no. 451.
Hodrick, R. J., & Prescott, E. C. (1997). Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit, and Banking*, *29*(1), 1–16.
Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, *4*(2), 251–257.
Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*, 359–366.
Hornik, K., Stinchcombe, M., & White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks*, *3*, 551–560.
Hutchinson, J. M., Lo, A. W., & Poggio, T. (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *Journal of Finance*, *49*, 851–889.
Jacquemin, A., & Sapir, A. (1995). *Is a European hard core credible?* (p. 1242). CEPR discussion paper: A statistical analysis.
Kauermann, G., Teuber, T., & Flaschel, P. (2012). Exploring US business cycles with bivariate loops using penalized spline regression. *Computational Economics*, *39*(4), 409–427.
Kiani, K. M. (2005). Detecting business cycle asymmetries using artificial neural networks and time series models. *Computational Economics*, *26*(1), 65–85.
Kiani, K. M. (2007). Asymmetric business cycle fluctuations and contagion effects in G7 countries. *International Journal of Business and Economics*, *6*(3), 237–253.
Kiani, K. M. (2009a). Asymmetries in macroeconomic time series in eleven Asian economies. *International Journal of Business and Economics*, *8*(1), 37–54.
Kiani, K. M. (2009b). Neural networks to detect nonlinearities in time series: Analysis of business cycle in France and the United Kingdom. *Applied Econometrics and International Development*, *9*(1), 67–76.
Kiani, M. K. (2011). Fluctuations in economic and activity and stabilization policies in the CIS. *Computational Economics*, *37*(2), 193–220.
Kiani, K. M., & Bidarkota, P. V. (2004). On business cycle asymmetries in G7 countries. *Oxford Bulletin of Economics and Statistics*, *66*(3), 333–351.
Kiani, K., & Kastens, T. L. (2006). Using macro-financial variables to forecast recessions: An analysis of Canada, 1957–2002. *Applied Econometrics and International Development 2005*, *6*(3), 97–106.
Kiani, K. M., Bidarkota, P. V., & Kastens, T. L. (2005). Forecast performance of neural networks and business cycle asymmetries. *Applied Financial Economics Letters*, *1*(4), 205–210.
Koopman, S. J., Harvey, A. C., Doornick, J. A., & Shephard, N. (1995). *STAMP 5.0: Structural time series analyser modeller and predictor: The manual*. London: Chapman & Hall.
Konishi, S., & Kitagawa, G. (1996). Generalized information criteria for model selection. *Biometrika*, *83*(4), 875–890.
Kozicki, S. (1999). Multivariate detrending under common trend restrictions: Implications for business cycles research. *Journal of Economic Dynamics and Control*, *23*, 997–1028.

Kuan, C. M., & White, H. (1994). Artificial neural networks: An econometric perspective. *Econometric Reviews*, *13*, 1–91.

Kim, C.-J., & Nelson, C. R. (1999). Friedman's plucking model of business fluctuations: Tests and estimates of permanent and transitory components. *Journal of Money, Credit and Banking*, *31*(3), 317–334.

Lee, T., White, H., & Granger, C. W. J. (1993). Testing for neglected nonlinearity in time series models. *Journal of Econometrics*, *56*, 269–290.

Ljung, G., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*, 297–303.

Lucas, R. E, Jr. (1977). Understanding business cycles. In K. Karl Brunner & A. Meltzer (Eds.), *Stabilization of the domestic and international economy*. Amsterdam: North Holland.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of 5th Berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 281–297). Berkeley: University of California Press.

Malik, S., & Pitt, M. K. (2011). Particle filters for continuous likelihood evaluation and maximazation. *Journal of Econometrics*, *165*, 190–209.

Massmann, M., & Mitchell, J. (2004). Reconsidering the evidence: Are Eurozone business cycles converging? *Journal of Business Cycle Measurement and Analysis*, *1*, 275–307.

Mise, E., Kim, T.-H., & Newbold, P. (2005). On suboptimality of the Hodrick–Prescott filter at time series endpoints. *Journal of Macroeconomics*, *27*(1), 53–67.

Michaelides, P. G., Vouldis, A. T., & Tsionas, E. G. (2010). Globally flexible functional forms: The neural distance function. *European Journal Of Operational Research*, *206*, 456–469.

Neftci, S. N. (1984). Are economic time series asymmetric over the business cycle? *Journal of Political Economy*, *92*(2), 307–328.

Nelson, C. R., & Plosser, C. I. (1982). Trends and random walks in macroeconomics time series: Some evidence and implications. *Journal of Monetary Economics*, *10*, 139–167.

Oh, K., Zivot, E., & Creal, D. (2008). The relationship between the Beveridge-Nelson decomposition and other permanent transitory decompositions that are popular in economics. *Journal of Econometrics*, *146*, 207–219.

Owners, R., & Sarte, P. D. (2005). How well do diffusion indexes capture business cycles? A spectral analysis. *Federal Reserve Bank of Richmond Economic Quarterly*, 91(Fall (4)), 23–42.

Ozbek, L., & Ozlale, U. (2005). Employing the extended Kalman filter in measuring the output gap. *Journal of Economic Dynamics and Control*, *29*(9), 1611–1622.

Papageorgiou, T., Michaelides, P. G., & Milios, J. G. (2010). Business cycles synchronization and clustering in Europe (1960–2009). *Journal of Economics and Business*, *62*(5), 419–470.

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

Pesaran, M. H., & Potter, S. M. (1997). A floor and ceiling model of US output. *Journal of Economic Dynamics and Control*, *21*(4–5), 661–695.

Phillips, P. C. B., & Magdalinos, T. (2008). Limit theory for explosively cointegrated systems. *Econometric Theory*, *24*, 865–887.

Pitt, M., Giordani, P., & Kohn, R. (2010). Bayesian inference for time series state space models. In J. Geweke, G. Koop, & H. van Dijk (Eds.), *Handbook of Bayesian econometric*. Oxford: Oxford University Pres.

Pitt, M., dos Santos Silva, R., Giordani, P., & Kohn, R. (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics*, *171*, 134–151.

Pollock, D. S. G. (2000). Trend estimation and de-trending via rational square-wave filters. *Journal of Econometrics*, *99*, 317–334.

Psaradakis, Z., & Sola, M. (2003). On detreding and cyclical assymetry. *Journal of Applied Econometrics*, *18*, 271–289.

Qi, M., & Maddala, C. S. (1999). Economic factors and the stock market: A new perspective. *Journal of Forecasting*, *18*, 151–166.

Qi, M. (2001). Predicting U.S. recessions with leading indicators via neural network models. *International Journal of Forecasting*, *17*, 383–401.

Ramsey, J. B., & Rothman, P. (1996). Time irreversibility and business cycle asymmetry. *Journal of Money, Credit and Banking*, *28*(1), 1–21.

Rudin, W. (1976). *Principles of mathematical analysis*. New York: McGraw-Hill International Edition.

Santin, D., Delgado, F., & Valino, A. (2004). The measurement of technical efficiency: A neural network approach. *Applied Economics*, *36*, 627–635.

Scheinkman, J. A., & LeBaron, B. (1989). Nonlinear dynamics and stock returns. *The Journal of Business*, *62*(3), 311–337.

Schumpeter, J. A. (1939). *Business cycles: A theoretical, historical and statistical analysis of the capitalist process*. New York: McGraw Hill.

Serrano-Cinca, C. (1997). Feedforward neural networks in the classification of financial information. *The European Journal of Finance*, *3*, 183–202.

Sheela G. K., & Deepa S. N. (2013). Review on methods to fix number of hidden neurons in neural networks. Mathematical Problems in Engineering, Vol. 2013, Article ID 425740. doi:10.1155/2013/425740.

Stock, J. H., & Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, *44*(2), 293–335.

Swanson, N. R., & White, H. (1995). A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. *Journal of Business & Economic Statistics*, *13*(3), 265–275.

Swanson, N. R., & White, H. (1997a). A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *The Review of Economics and Statistics*, *79*(4), 540–550.

Swanson, N. R., & White, H. (1997b). Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International Journal of Forecasting*, *13*(4), 439–461.

Tanizaki, H., & Mariano, R. S. (1994). Prediction, filtering and smoothing in non-linear and non-normal cases using Monte Carlo integration. *Journal of Applied Econometrics*, *9*, 163–179.

Taylor, C. (1995). *EMU 2000? Prospects for European Monetary Union*. London: Chatham House Papers, The Royal Institute of International Affairs.

Terasvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, *89*, 208–218.

Tsay, R. S. (1988). Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*, *7*, 1–20.

Vishwakarma, K. P. (1994). Reckognizing business cycles turning points by means of neural networks. *Journal of Computational Economics*, *7*, 175–185.

Watanabe, T. (1999). A non-linear filtering approach to stochastic volatility models with application to daily stock returns. *Journal of Applied Econometrics*, *14*, 101–121.

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

Wingrove, C. R., & Davis, E. R. (2012). Classical linear-control analysis applied to business-cycle dynamics and stability. *Journal of Computational Economics*, *39*(1), 77–98.

Woitek, U. (1998). Note on the Baxter–King filter. *Working Papers* 9813, Business School, Economics, University of Glasgow.

Zellner, A. (1992). Commentary. In Belagia, M. & Garfinkel, M. (Eds.), The business cycle: Theories and evidence: Proceedings of the sixteenth annual economic policy conference of the Reserve Bank of St Louis.

Zhang, G. P., & Berardi, V. L. (2001). Time series forecasting with neural network ensembles: An application for exchange rate prediction. *Journal of the Operational Research Society*, *52*, 652–664.