CYCLE EXTRACTION

A comparison of the Phase-Average Trend method,
the Hodrick-Prescott and Christiano-Fitzgerald filters

Ronny Nilsson and Gyorgy Gyomai
OECD

Abstract

This paper reports on revision properties of different de-trending and smoothing methods (cycle estimation methods), including PAT with MCD smoothing, a double Hodrick-Prescott (HP) filter and the Christiano-Fitzgerald (CF) filter. The different cycle estimation methods are rated on their revision performance in a simulated real time experiment. Our goal is to find a robust method that gives early turning point signals and steady turning point signals. The revision performance of the methods has been evaluated according to bias, overall revision size and signal stability measures.

In a second phase, we investigate if revision performance is improved using stabilizing forecasts or by changing the cycle estimation window from the baseline 6 and 96 months (i.e. filtering out high frequency noise with a cycle length shorter than 6 months and removing trend components with cycle length longer than 96 months) to 12 and 120 months.

The results show that, for all tested time series, the PAT de-trending method is outperformed by both the HP or CF filter. In addition, the results indicate that the HP filter outperforms the CF filter in turning point signal stability but has a weaker performance in absolute numerical precision. Short horizon stabilizing forecasts tend to improve revision characteristics of both methods and the changed filter window also delivers more robust turning point estimates.

Contents

1. Introduction

The methodology used for the OECD Composite Leading Indicator System (CLI) has remained largely unchanged since CLIs were first published in December 1981. The system is based on the “growth cycle” approach where cycles are measured on a deviation-from-trend basis. Therefore the selection of a well behaving de-trending method is crucial for the quality of the leading indicator. Currently the phase-average trend method (PAT) – constructed by the National Bureau of Economic Research (NBER) in the U.S. and further developed by the OECD – is used for trend estimation.

An internal OECD study was conducted to compare the properties of the PAT method and the Hodrick-Prescott (HP) filter in 2002. The study took a similar approach and arrived at a similar conclusion to that of Zarnowitz and Ozyildirim [2006] [1]. The OECD study concluded that the two de-trending methods extract similar cycles, lead to similar turning points and in most cases have highly correlated cyclical amplitudes. The PAT method was evaluated to perform better in the presence of level shift outliers, and to adapt better to variations in cyclical amplitudes in different series.

The PAT method has two operational modes: fully automated; and manual turning point insertion. The OECD uses the latter in its CLI production cycle. As a consequence the tests were carried out with manual turning point insertion, and they indicated that the PAT cycle estimates may be sensitive to turning point updates, and they give biased results if the turning point updates are not carried out in a timely manner. However the sensitivity of the different methods in real-time applications was not tested in the 2002 study.

Notwithstanding the sensitivity issue of PAT, several other reasons were identified that led us to seek alternative approaches. The PAT method is not very transparent, not widely used by economists, its automated version often gives unreasonable results, (in its present implementation) has time series length limitations, and uses ad-hoc built-in parameters that are non-modifiable but determine the average extracted cycle length. Moreover the non-automatic version relies on manual turning point insertion, which can be subjective.

This paper reports the results of a sensitivity study in a simulated real time experiment of different de-trending and smoothing methods, this time including PAT with Months for Cyclical Dominance (MCD) smoothing, a double Hodrick-Prescott (HP) filter and the Christiano-Fitzgerald (CF) filter. We rely on the frequency domain interpretation of the time series for the parameterization of these two latter filters and we use the filters with different de-trending specifications: a low frequency cut-off at 8 and 10 years and smoothing of high frequency irregularities at 6 and 12 months. We introduce the methods in more detail in Section 2 of the paper.

The study covers time series of both multiplicative and additive nature, containing different amounts of noise. We describe the dataset in detail in section 3.

The sensitivity analysis has been carried out in a simulated real time experiment, which corresponds to the quasi real time experiment performed by Orphanides and van Norden [2002]. Our study focuses on fewer methods (all our three methods are non-parametric and

[1] However the OECD study reviewed only the PAT and HP methods unlike Zarnowitz and Ozyldirim who reviewed others in addition.

robust); but with a focus on a wider range of revision measures. The main difference in revision measures is that while Orphanides and van Norden [2002] only compare cumulated revisions (i.e. total revision between the initial estimate of a cyclical value and the final estimate), we analyze the characteristics of revisions between consecutive vintages of the same cyclical value estimate. Section 4 contains the description of the experiment.

The different de-trending methods are rated based on their revision performance over consecutive vintages and in particular their ability to indicate or date cyclical turning points that are not later revised. In our baseline experiment, we try to identify the best de-trending method by evaluating how the HP and CF filters operate on the input time series with smoothing and de-trending specifications set to 6 and 96 months to be comparable to the parameters fixed in the PAT method, which is used as the benchmark. We measure bias in early revisions, overall revision size and most importantly turning point signal stability. The performance evaluation of the different de-trending and smoothing methods can be found in section 5.

In addition to the baseline experiment, in a second phase, we investigate how revision performance changes or improves when we use stabilizing forecasts before de-trending. We use the TRAMO module of the TRAMO-SEATS seasonal adjustment procedure to provide automated ARIMA forecasts for various forecast horizons in a simulated real time set-up. Finally, we evaluate the effects of increasing smoothness and allowing for longer horizon trends with parameters set to 12 and 120 months respectively. Section 6 reports on the results of these modifications compared to the baseline scenario.

We conclude that both the CF filter and HP filter outperform the currently used OECD de-trending methodology (PAT with MCD smoothing). The choice between the CF-filter and HP-filter depends on the application. The HP-filter is more suitable for applications where turning-point signals are more important whereas CF-filter is preferred where higher numerical precision is required, reflecting its relatively small cumulative revisions. Because the OECD system of CLIs' main objective is to identify cyclical turning points in a timely and stable manner, the double HP-filter fits best OECD's purpose.

2. Cycle extraction methods

The OECD CLI system uses the "deviation-from-trend" approach. This means that in the construction phase of the CLIs co-movements and similarities in patterns between the reference series and individual CLI components are evaluated between smoothed and de-trended versions of these series. This makes the cycle extraction (the equivalent of de-trending and smoothing) a crucial step in the CLI construction and production process. Therefore we deal with competing cycle extraction methods in considerable detail.

We can approach the cycle extraction exercise slightly differently than the distinct steps of trend removal and smoothing. Instead of observing the series in the time domain, we can treat the series as a complex sinusoid, built from simple sine waves of different wave length. The trend part of the series is comprised by the low frequency (high wave length) sinusoids, whereas the noise is formed by a set of high frequency sinusoids See Pollock [2006] for a thorough introduction to the related mathematical concepts of this decomposition.

Once we have the translation of our series from the time domain to the frequency domain, we can single out the cycles we are interested in, and eliminate the components whose wave length is too long (trend) or too short (noise). Much depends on the optimal cycle length an issue on which there is considerable debate: What is a business cycle? How long should a cycle be? Or, more closely related to the de-trending exercise, what is the cycle length that we consider too short or too long to treat meaningfully as a business cycle? The early papers in cyclical analysis characterize movements between 1.5 and 8 years as the cycle length of interest. Some more recent papers however argue that modern economic cycles may last longer, and cyclical fluctuations are smaller. (For example see Agresti and Mojon [2001] who endorse 10 years as the upper boundary for the business cycles in Europe.) The de-trending and smoothing methods chosen therefore should be aligned with our prior expectations on business cycles.

We will present how the Phase Average Trend method, the Hodrick-Prescott filter and the Christiano-Fitzgerald filter operate on the input time series to yield the pure cycle.

2.1. The Phase-Average-Trend (PAT) method

This is the method used at present in the OECD CLI system. It is the modified version of the similar (PAT) method developed by the United States NBER. This method is used in combination with the Bry-Boschan turning point detection algorithm. The resulting medium-term cycle is smoothed by the MCD method to yield the final smooth cycle.

The PAT method consists of the following set of operations:

- first estimation and extrapolation of long-term trend (75 month moving average);
- calculation of deviations from moving average trend;
- correction for extreme values;
- identification of tentative turning points and determination of cyclical phases, i.e. expansions and contractions (Bry-Boschan routine);
- new estimation of the long-term trend; We proceed by calculating averages for each phase, smoothing the sequence of phase averages over three adjacent phases. Finally these smoothed values are positioned in the centre of their corresponding phases and linearly interpolated.
- extrapolation of the long-term trend at the series ends to recover periods lost because of the centered moving averaging;
- calculation of deviations from PAT trend;

The implementation of PAT works in two modes: automated and manual (supervised) mode. The automated mode uses the turning points from the Bry-Boschan algorithm, the supervised mode accepts turning points entered by the user, and ignores Bry-Boschan values. As most of the parameters of the PAT procedure are fixed, the manual turning point setting is the way to fine-tune the system, and modify implausible cycle results. The manual turning point setting gives the analyst a very strong and precise tool to intervene in the de-trending process. At the same time this targeted intervention is one of the most criticized features of PAT. The rules or conventions that govern the intervention of the analyst are not easy to document, different analysts may come up with different turning point choices and, as a consequence, the PAT with manual turning point specification is a non-transparent, ad-hoc system.

The smoothing coupled with the PAT method is the so called MCD moving average. This procedure ensures approximately equal smoothness between series and also ensures that the month-to-month changes in each series are more likely to be due to cyclical than irregular movements. The MCD moving average time span is defined as the shortest span for which the I/C ratio is less than unity; where I and C are average absolute month-to-month changes of the irregular and trend-cycle component of the series, respectively. The maximum time span of the MCD moving average is capped at 6 months.

The PAT method in automatic mode has a tendency to select cycles between 15 and 75 month, as a direct consequence of parameters fixed in the PAT software. These cycle lengths are somewhat shorter than the cycle lengths in which we are interested: 18 to 96 months. Therefore manual intervention is inevitable, minor cycles have to be removed and the process has to be rerun in supervised mode for all practical applications. At the same time PAT is sensitive to turning point updates and is known to give biased results if the turning point updates are not carried out in a timely manner, which is a real danger in a large scale manually operated project.

A few additional operational deficiencies of PAT can be summarized as follows:

- The PAT method is not sufficiently transparent for two reasons: firstly because the algorithm which produces the cycle estimates is not available in any major econometric software, secondly because it has to be operated with manual turning point insertion, which can be a relatively subjective exercise.
- The method was developed four decades ago when computational power was more limited and software languages less developed. Methods that were designed and developed later, in parallel with the evolution of knowledge in the real business cycles field of macroeconomics were built on new foundations, and PAT has remained unchanged since. The algorithm written in the 1970’s was adapted to run on personal computers in the mid 1990’s but preserved most of the limitations of the early implementation. (Many parameters are not modifiable, series that are longer than 50 years cannot be treated.)
- As the number of countries and zones included in the OECD CLI system has increased (and it is likely to increase further) the resource intensiveness of PAT (especially turning point updates for all components) are a burden that urges OECD to shift to de-trending methods that need less maintenance after initial calibration.

2.2. The Hodrick-Prescott (HP) filter

The Hodrick-Prescott filter is one of the best known and most widely used de-trending methods by macroeconomists. The filter was first described in a working paper published in 1981. In its original form the trend estimate is a result of an optimization problem:

$$y_t = \tau_t + c_t$$
$$\min_{\tau_t} \sum_t (y_t - \tau_t)^2 + \lambda \sum_t (\tau_{t+1} - 2\tau_t + \tau_{t-1})^2$$

We decompose our initial y_t series into τ_t – the trend component and c_t the cyclical component, such that we minimize the distance between the trend and the original series and at the same time we minimize the curvature of the trend series. The trade-off between the two goals is governed by the λ parameter.

The optimization problem has a solution that can be represented by a linear transformation which is independent from y_t. (see Maravall and del Rio [2001]). This makes the filter very fast.

What was impossible with the PAT method is possible with the HP filter. We can transform the filter into the frequency domain and understand its effects on various cycles that make up the time series. In frequency domain changes to λ determine the shape of the frequency response function of the HP filter and the cut-off frequency. The frequency response function shows how the filter affects certain frequencies, it shows which frequencies are retained and which are let through. The cut-off frequency is defined as the frequency where 50% is let through and 50% is retained from the original power of the cycle. Thus we can align the λ parameter with our goal to filter out economic cycles in a certain frequency range with the help of the transformation into the frequency domain. Before the frequency domain interpretation emerged there were only rules of thumb to set the λ parameter. Rule of thumb values later proved to be in line with values that had been determined by frequency selection criteria, i.e. separating the "trend" cycles with a wavelength larger than 8 years. See for example Maravall and del Rio [2001] to learn more on how the λ parameter translates to the frequency domain.

Properties of the HP filter:

- The cut-off region is not steep; meaning that leakage from cycles just outside the target region can be significant. In engineering applications filter leakage is a sign of a poor filter. However in business cycle analysis there are arguments to support at least a small degree of desirable leakage. Since the frequency band of 1.5 to 8 years has been selected based on expert decision several decades ago, the boundaries 1.5 and 8 years should not be regarded as carved in stone. The filter leakage for example allows strong 9 year cycles to appear in the filtered series.
- It is asymmetric. With the exception of the central values the double HP filtered series are phase shifted compared to the underlying ideal cycle. Phase shifts fade out for a given observation as newer observations arrive.

We apply the HP filter twice to achieve a smoothed de-trended cycle. First we remove the long term trend by setting λ to a high value, and we preserve the business cycle frequencies and the high frequency components. Second, we apply the HP filter with a smaller λ, meaning that the cut-off frequencies are much higher, and so, preserve the trend part of the filter results. The first step is de-trending the second step smoothes.

2.3. The Christiano-Fitzgerald (CF) filter

The Christiano-Fitzgerald random walk filter is a band pass filter that was built on the same principles as the Baxter and King (BK) filter. These filters formulate the de-trending and smoothing problem in the frequency domain. Should we have continuous and/or infinitely long time series the frequency filtering could be an exact procedure. However the granularity and finiteness of real life time series do not allow for perfect frequency filtering. Both the BK and CF filters approximate the ideal infinite band pass filter. The Baxter and King version is a symmetric approximation, with no phase shifts in the resulting filtered series. But symmetry and phase correctness comes at the expense of series trimming. Depending on the trim factor a certain number of values at the end of the series cannot be calculated. There is a trade-off between the trimming factor and the precision with which the optimal filter can be approximated. On the other hand the Christiano-Fitzgerald random walk filter uses the whole time series for the calculation of each filtered data point. The advantage of the CF filter is that it is designed to work well on a larger class of time series than the BK filter, converges in the long run to the optimal filter, and in real time applications outperforms the BK filter. For details see Christiano-Fitzgerald [1999]. For these reasons we included only the Christiano-Fitzgerald filter in our study that compares different cycle detection methods.

The CF filter has a steep frequency response function at the boundaries of the filter band (i.e. low leakage); it is an asymmetric filter that converges in the long run to the optimal filter. It can be calculated as follows:

$$c_t = B_0 y_t + B_1 y_{t+1} + ... + B_{T-1-t} y_{T-1} + \tilde{B}_{T-t} y_T + B_1 y_{t-1} + ... + B_{t-2} y_2 + \tilde{B}_{t-1} y_{1,} \text{ where}$$

$$B_j = \frac{\sin(jb) - \sin(ja)}{\pi j}, j \geq 1, \text{and } B_0 = \frac{b-a}{\pi}, a = \frac{2\pi}{p_u}, b = \frac{2\pi}{p_l}$$

$$\tilde{B}_k = -\frac{1}{2} B_0 - \sum_{j=1}^{k-1} B_j$$

The parameters p_u and p_l are the cut-off cycle length in month. Cycles longer than p_l and shorter than p_u are preserved in the cyclical term c_t.

3. The dataset

The dataset used in this study covers selected time series used as components in the calculation of Composite Leading Indicators for a few OECD countries. The main source of the data is the OECD's Main Economic Indicators database.

The OECD Composite Leading Indicators System does not use a standard set of leading indicator components for all countries (35 countries), because of important differences between them in economic structure and statistical systems. Leading indicator series which perform well in both tracking and forecasting cyclical developments differ from country to country and may also change over time.

The different subject areas from which the leading indicator series (224 series are used in total, about 5-10 for each country) are chosen are set out in Table 1. Certain types of series recur regularly in the list of leading indicators for different countries. Business and consumer tendency survey series are among those most frequently used in the countries where they are available. These series concern business expectations on production, inflow of new orders, level of order books, stocks of finished goods and the assessment of the general economic situation by both businesses and consumers. The most frequently used other series are monetary and financial series such as share prices, money supply and interest rates. Series relating to stocks and orders, construction, retail sales, prices and foreign trade are also used frequently.

The selected series are set out in Table 2 and include time series of both multiplicative and additive nature containing different amounts of noise. The MCD value shown in the table gives a rough measure of the smoothing needed to reduce noise in order to highlight the cyclical properties of the series.

The selected monthly time series cover the period January 1970 – December 2007. We performed the simulated real-time experiment for the last 200 observations. Thus the shortest analyzed series were based on data for the period January 1970 – February 1991.

Table 1 Component series used in the OECD System of Composite Leading indicators

Subject area	Share of total number of components (%)	Component series (selection)
Financial series	24	Interest rates, share prices, monetary aggregates
Business tendency surveys	39	Business confidence, finished goods stocks, order books, production expectations
Consumer surveys	7	Consumer sentiment
Real quantitative indicators	30	New orders, passenger car registration/sales, construction approval/starts, hours worked, stocks, export/import

Table 2 Time series used in the study selected from components included in the OECD System of Composite Leading Indicators

Indicator	Country	Time series model	Smoothness (MCD)	Time series period	Experiment period starts
Overtime hours, manufacturing	Japan	Additive	1	January 1970 – October 2007	February 1991
Business confidence	United States	Additive	2	January 1970 – December 2007	April 1991
Consumer sentiment	United States	Additive	3	January 1978 – December 2007	April 1991
Net new orders	United States	Multiplicative	4	January 1970 – November 2007	March 1991
Import to export ratio	Japan	Multiplicative	5	January 1973 – November 2007	April 1991
New passenger car registration	United Kingdom	Additive	6	January 1970 – November 2007	March 1991

4. The experiment

OECD CLIs are published monthly and produced in a narrow time-frame. We designed our experiment to simulate the real production process, to measure the performance of the de-trending and smoothing methods under conditions in which they are supposed to operate. We started with a shortened time series (form 1970 to Mar/Apr 1991). We performed the outlier detection, de-trending and smoothing operation on this series and we standardized the resulting cyclical series. (This is the normal sequence of operations that is applied to each component in the OECD CLI construction process. An updated methodological document that describes the OECD CLI construction and production process in detail is available at: http://www.oecd.org/dataoecd/26/39/41629509.pdf.) Then we gradually increased the length of the time series, and for each new observation added, we repeated the sequence of operations and we recorded the resulting cyclical component. As a result we obtained 200 consecutive estimates (we will call them vintages) of the cyclical components of the analyzed series. (This procedure is similar to the quasi-real time scenario described in Orphanides and van Norden [2002].) In our baseline experiment we included three de-trending and smoothing specifications:

- the PAT method with the MCD smoothing,

- the double HP filter with parameters that correspond to frequency cut-off levels between 6 and 96 months,
- the CF filter with a band-pass between 6 and 96 months.

The motivation for the selection of the 6 and 96 month band reflected the built-in characteristics of the PAT method. Overall smoothness is limited by a cap at 6 months in the MCD moving average time span. Therefore we did not apply stronger smoothing (high frequency filtering) in the HP and CF case either. In the case of low frequency filters the choice was harder. The PAT method has various iterative steps and, as a result, it is difficult to establish the exact low frequency cut-off characteristics. The PAT filter however has been initially designed/calibrated to measure economic cycles that are shorter than 8 years – see Boschan and Ebanks [1978]. Thus we relied on this intended 96 month cycle for specifying the setting for the longest admitted cycles for the other two de-trending methods.

The goal in this experiment is to select the de-trending method that has fast response (identifies turning points quickly) and has good revisions characteristics (few revisions, small revisions, early revisions).

5. Evaluating the de-trending methods

We have calculated several measures to evaluate the relative performance of the three de-trending methods. The revisions for each observation *t* were obtained with the following formula:

$R_{i,t} = \hat{c}_{t,t+i} - \hat{c}_{t,t+i-1}$,

Where $R_{i,t}$ is the *i*-th revision of observation *t*, and $\hat{c}_{t,t+i}$ is the normalized estimate of the cyclical component for period *t*, by using information up to period $t+i$. Based on these revision figures we have calculated the following measures: mean revision with a Newey-West heteroscedasticity and autocorrelation (HAC) corrected t-test, mean absolute revision, standard deviation of revisions, first order autocorrelation, bias towards the centre, sign change percentage, direction change percentage.

Mean revision and mean absolute revision measures were also calculated for cumulated revisions.

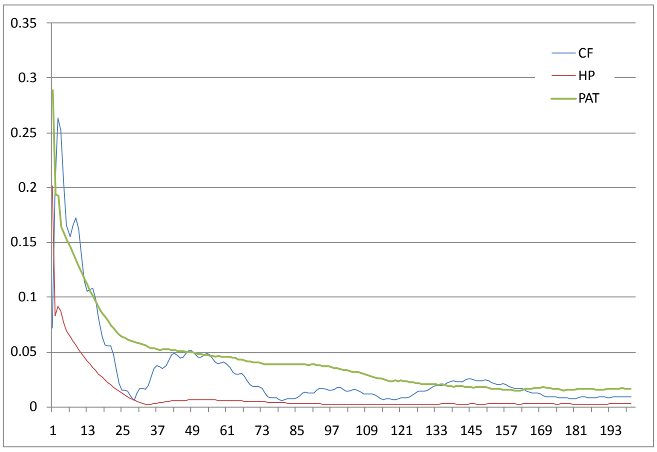
$R'_{i,t} = \hat{c}_{t,t+i} - \hat{c}_{t,t}$,

We analyzed the revision patterns for all series included in the dataset. In the Annex we provide a summary of the results for each series separately. The three de-trending methods were evaluated for each series, and they received points based on their relative performance. Based on the overall scores the order of preference for the methods would be HP, then CF and finally PAT (for further details see the tables in the Annex). The results were similar for each series type; therefore, in the following paragraph, we will present most of the results for the average of all the analyzed series.

5.1. Overall revision size

We have placed strong emphasis on the revision size measures: the mean absolute revision, the standard deviation of the revision, and the cumulated mean absolute revision. Based on these measures the PAT is clearly inferior to HP and CF. The cycles estimated with HP have the smallest revisions each month, however the CF cycles have smaller cumulated revisions. (In the study carried out by Orphanides and van Norden different methods are evaluated by cumulated revision measures.)

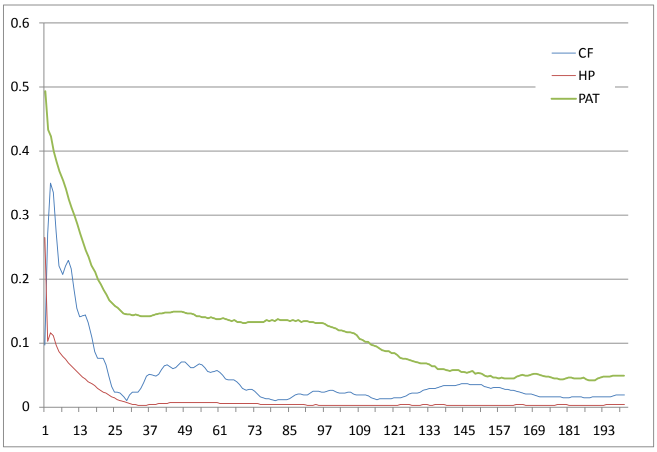
Mean absolute revision: $\sum_t \left| R_{i,t} \right| / n$



The horizontal axis correspond to *i*'s in the formulas.

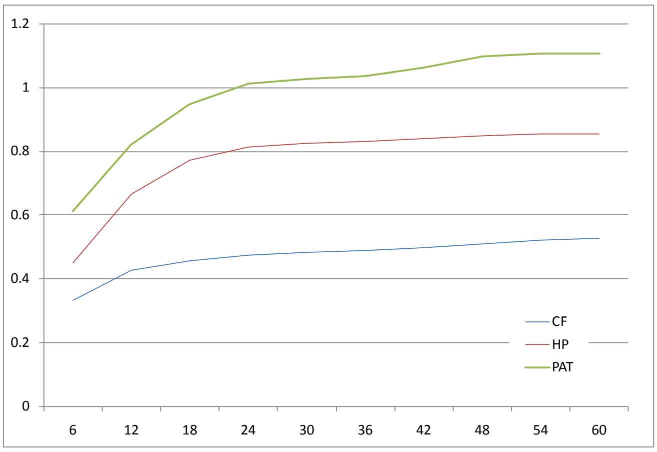
Mean absolute revision measures the overall size of revisions regardless of the potential bias that may be in the revisions. All three methods have decreasing revisions over time. The HP method outperforms CF and PAT. The HP revision sizes decrease rapidly to negligible amounts after 3 years. The PAT revisions sizes are bigger and more persistent. The CF revision sizes diminish in an oscillatory manner. There are recurring periods where the size of revisions approaches the size of PAT revisions.

Standard deviation of revisions: $\sqrt{\sum_t \left(R_{i,t} - \mu_i\right)^2 / (n-1)}$



Standard deviation measures the overall size of revisions, but corrects for the potential bias that may be in the revisions. It also tends to emphasize extremes compared to the mean absolute revision. All three methods have decreasing revisions over time, similar to the mean absolute revision results. However the performance advantage of the CF filter compared to the PAT becomes more accentuated when we use the standard deviations measure. This is due to the fact that extremely large revisions are more likely to occur in PAT's case than for the other two methods.

Cumulated absolute revision: $\sum_t \left| R'_{i,t} \right| / n$



The cumulated absolute revision measures the size of revision accumulated from the first estimate of a cycle value, without bias corrections. The cumulated revisions grow steeply in the first two years for all methods. The CF has the most favorable cumulated revisions, followed by the HP and the PAT. While the HP filter provided smaller revisions of the cyclical estimates on average, the revisions tend to be more persistent. The CF estimates' oscillatory behavior results in smaller cumulated revisions.

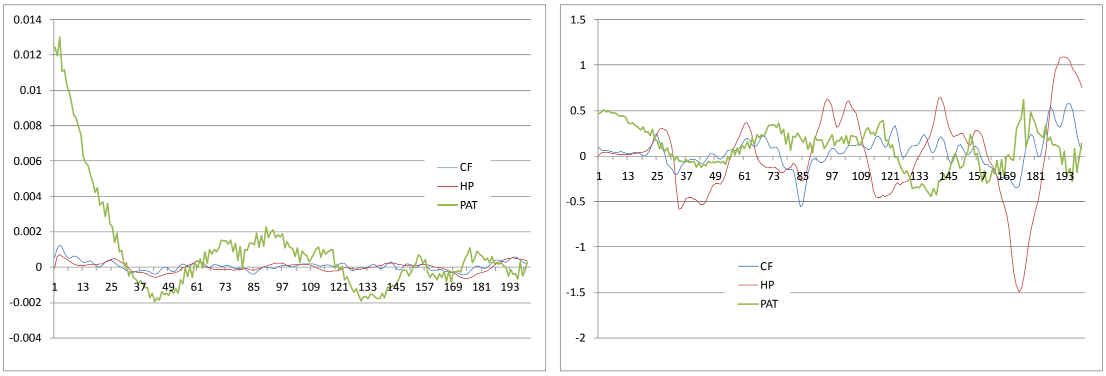
5.2. Bias and autocorrelation in revisions

In a second block we present a set of indicators aimed at evaluating the quality of the cycle estimation methods *per se*. We measured bias, autocorrelation, and conditional bias. Should these measures show significant values, it would mean that the methods are suboptimal and could be improved on by utilizing the information contained in the history of revisions.

Bias(left graph) and a Newey-West HAC corrected t-test (right graph)

Bias : $\mu_i = \frac{\sum_t R_{i,t}}{n}$

Newey-West HAC corrected t-test:μ_i/σ_{HAC}, where σ_{HAC} is the heteroskedasticity and autocorrelation corrected standard deviation of the mean revision.

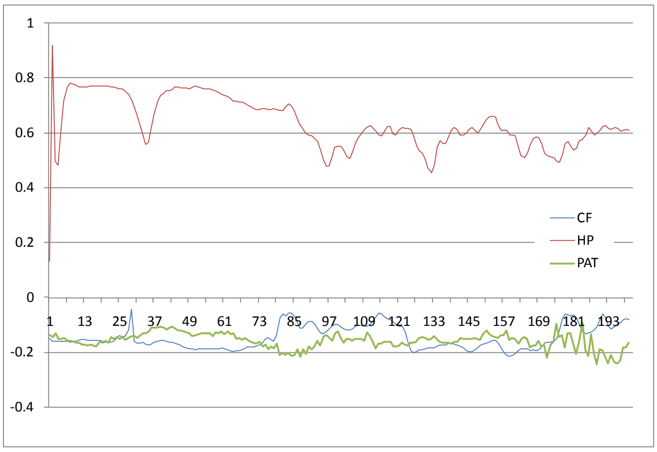


The bias values vary a lot in different time series, but in general it is largest for the PAT estimates. As it appears on the graph PAT shows a positive bias that is an order of magnitude larger than the other two methods after averaging biases for the six analyzed series.

The HAC corrected t-tests show, in several cases, that the biases in the individual series are significant. Although the regions where the series become significant are more series-specific than method-specific. This follows from the fact that after averaging the HAC measures for all series (graph on the right) the measure does not show significant bias for any method or revision period.

We can note however that the PAT method has larger t-test values for the revisions in the first two years, whereas the HP method tends to have close to significant biases in later revisions. This late revision bias is less worrying since the size of the revisions is negligible in later periods. The calculation of the HAC corrected t-tests is described in: di Fonzo [2005]

Autocorrelation $\rho_i = \sqrt{\Sigma_t (R_{i,t} - \mu_i)(R_{i,t-1} - \mu_i)/(\sigma_i^2 n)}$

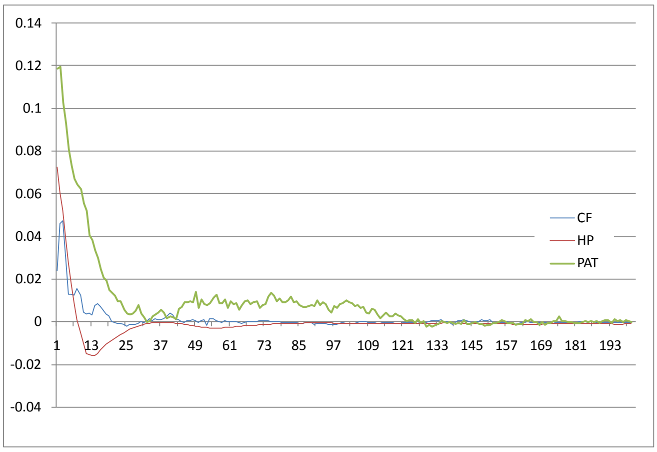


High first order autocorrelation signals that the de-trending method is not optimal in the sense that useful information is contained in past revisions, which can help predict current and upcoming revisions. In other words there is room to improve the cycle estimates for methods with high AC.

The Hodrick-Prescott method shows strong positive first order autocorrelation, which is more accentuated when the de-trending is performed on a relatively smooth series, and is weaker, but still considerable, with series having smaller signal to noise ratios. The autocorrelation patterns are less clear for the CF and PAT methods and they are different for the tested series.

Conditional (central) bias

$$\text{bias}_i = \sum_t \text{sgn}\left(\hat{c}_{t,t+i-1} - 100\right) R_{i,t}/n$$



The conditional bias measures the average revision size for each method; we treat revisions that occur above the trend with opposite sign to those that occur below the trend. Positive values therefore mean a revision-bias towards the long term trend, and negative values mean revisions away from the trend.

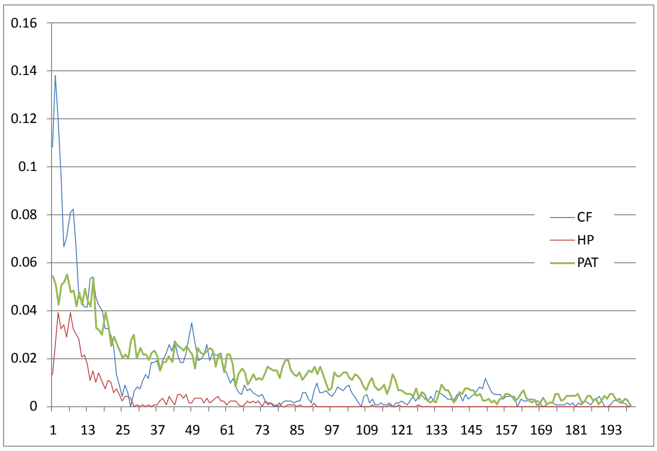
The graph shows that the revisions are mostly biased towards the centre, and this conditional bias is an order of magnitude larger than the unconditional bias. The CF method behaves the best and the HP is also relatively small compared to PAT, although after 6-9 months of revisions towards the centre it overshoots and revisions in the second year move away from the long term trend.

5.3. Signals

The ultimate goal of our CLIs is to accurately predict turning points in economic activity. Therefore the de-trending methods should be aligned to this goal, and besides having good revision characteristics as measured in the first two blocks they should emit a steady signal. A third block of measures captures how much the CLI relevant signals are revised.

Sign Change:

$$\#\left\{\text{sgn}\left(\hat{c}_{t,t+i-1} - 100\right) \neq \text{sgn}\left(\hat{c}_{t,t+i} - 100\right)\right\}/n$$



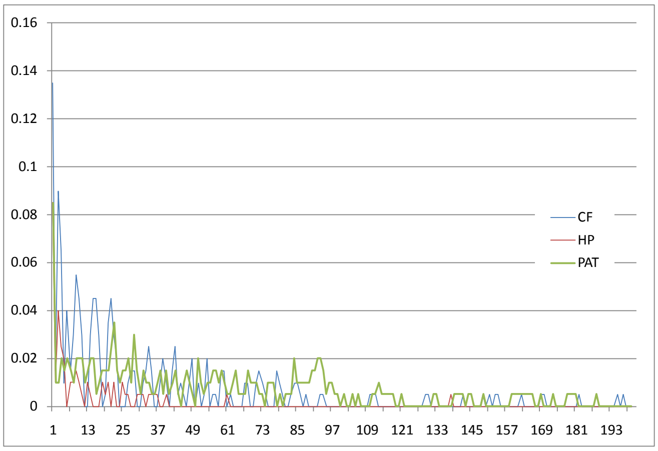
When determining cyclical phases whether the cyclical value is above or below trend values is key information. The “sign change” graph shows how many times the initial estimate has been revised to shift from below trend to above trend or vice versa.

The HP method performs best in all revision segments. The CF method has a high percentage of sign changes: in the first 6 months the chance of a sign revision is over 10%. PAT also has a relatively high sign change percentage and the number of occurrences decreases slowly.

The direction change measure is a percentage measure, like the sign change measure. It shows how many times the cyclical series have changed to increasing from decreasing or vice versa.

Direction Change

$$\#\{\mathrm{sgn}(\hat{c}_{t,t+i-1} - \hat{c}_{t-1,t+i-1}) \neq \mathrm{sgn}(\hat{c}_{t,t+i} - \hat{c}_{t-1,t+i}100)\}/n$$



Direction changes are typical only in the first few estimates; they quickly drop below 4 percent for all methods. Nonetheless the ordering is similar to that observed with sign changes. The HP method performs best, followed by PAT in general, and with some exceptions the CF method scores weakly in this test.

Producing the whole turning point estimation history for all time series goes beyond the scope of this paper, therefore we only analyze in detail the “USA net new orders” series. The following three graphs show the identified turning-points for each series vintage.

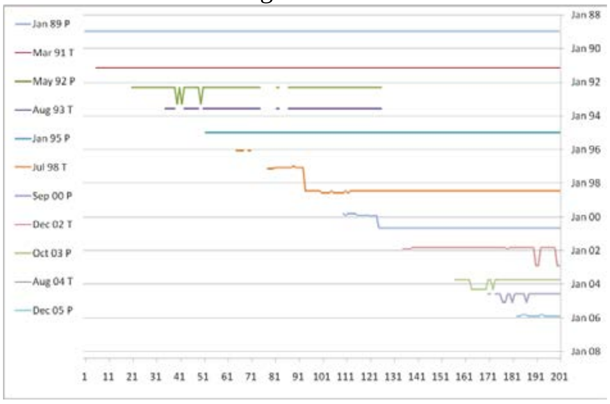
The horizontal axis contains the vintages; the vertical axis has the turning point dates. The ideal graph would show a perfect triangle, with straight horizontal lines. This would mean that the turning points are identified quickly, and after first observation, their location is not changed. The HP method comes closest to this ideal. The turning points identified with the CF and the PAT methods often oscillate. The turning point estimates change from vintage to vintage for a long period until they stabilize. These oscillations are smaller in size for the CF method, meaning that the estimated turning points often change only +/- 1 month. The PAT method has larger oscillations, and there are unexpected jumps in the turning point estimates even 100 vintages (i.e. more than 8 years) after their first appearance.

It is also striking that the de-trending method selection affects the final turning point list (See the list of dates on the left hand side of the following graphs). Although there are similarities (common points) in the estimated turning points, in cases where the original time series and cycles are not smooth enough, the simplified Bry-Boschan (BB) method was not robust. The simplified BB turning point identification method that we used does not smooth the time series before finding tentative locations within the cycle series. It finds the local minima and maxima within the estimated cycle series to mark turning points. In our experiment its parameters were set to find peaks/troughs that are maximum/minimum values in their 5 month neighborhood, and to respect the minimum phase length criterion of 9 month and minimum cycle length of 18 month. The lack of robustness in the turning point detection routine calls for further investigation, and OECD plans to carry out further research to improve the stability/robustness of the simplified BB method.

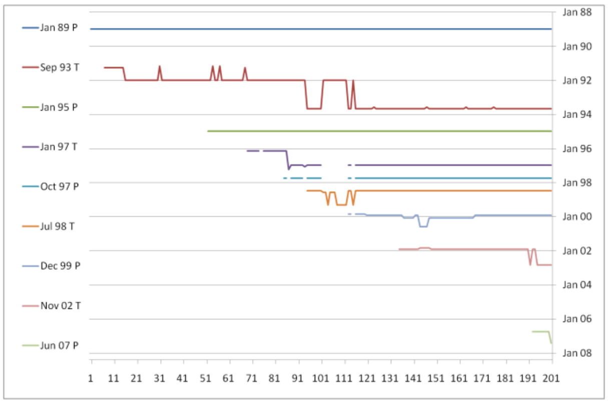
A possible way forward is to add a step to the method that does local rearrangements in the turning points (similar to the one applied in the original BB routine), or another approach would be to take into account the amplitudes of the cycles as done in Harding-Pagan [2003]. We also tested the effects of further smoothing in the cycle extraction method on TP robustness. The results are summarized in a later part of this paper.

Turning point estimation history:

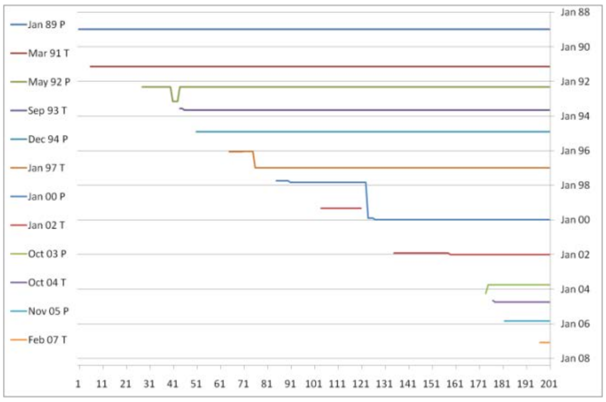
With the Christiano-Fitzgerald method:



With the PAT method:



With the Hodrick Prescott method:



6. Adjustments to the baseline experiment

The statistics and graphs in the baseline experiment clearly show that the PAT method is inferior to both the HP and CF method. Therefore the PAT method was not included in the following part of this study. The HP method proved to be better in terms of turning point signal stability and had smaller month to month revisions, but showed a surprisingly high and steady autocorrelation. The CF method scored worse in most revision measures except cumulated revisions.

Trying to improve further on the revision properties of the CF-filter and HP-filter, in this part of the paper we will test the consequences of using stabilizing forecasting on the de-trending method, and we will analyze the trade-off between early vintage and late vintage revisions that is involved in calculating smoother cycle series.

6.1. Shift in the filtered band to contain cycles between 12 and 120 month

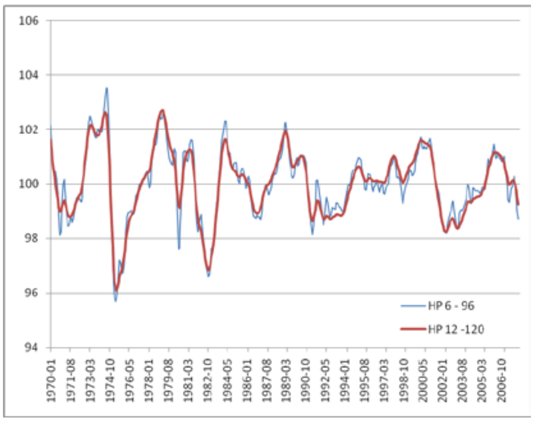
First we present the results of calculating smoother cycles by jointly increasing the upper and lower bands of the two filters. Remember that in the baseline scenario the high frequency cutoff point was set at 6 months and the low frequency cutoff was set at 96 months. In the alternative scenario these have been raised to 12 months and 120 months. The motivation for increasing the smoothness of the cyclical estimate and to increase the length of accepted cycles came from the observation that business cycles dampened: their amplitudes decreased, cycles became longer and harder to spot because of higher frequency variation in economic activity after the 90s.

The difference between the baseline scenario and the alternative scenario is not significant, according to most of the measures. The characteristics of the HP and CF filters are unaltered and their relative performance is unchanged according to each measure. The response of the HP method to the filter band change is mixed in terms of the cumulative revisions; 4 times out of the 6 series we tested there was a decrease in revision size, but the remaining 2 showed contrarian evolution. The first order autocorrelation remained high and steady. The CF filter has slightly worse sign and directional revision percentages after the filter-band change, but, at the same time, slightly improved mean absolute revision statistics. However the main advantage of using greater smoothing and allowing longer cycles is that turning point estimation stability improved for both methods. We illustrate this increased stability in the following points:

1. The estimated cyclical patterns are easier to spot just by looking at them; the identified turning-points are much less dependent on the turning-point selection algorithm and its parameterization. The from the higher degree of smoothing yields fewer short lived cycles, fewer local minima and maxima, that could mislead the turning point selection algorithm. Therefore, although our illustration uses only one series, the results are valid more generally.

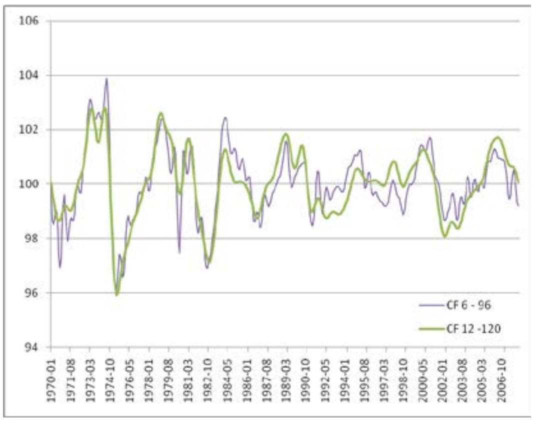
Final cycle estimate

With the Hodrick-Prescott method
6-96 vs. 12-120



"USA net new orders" series

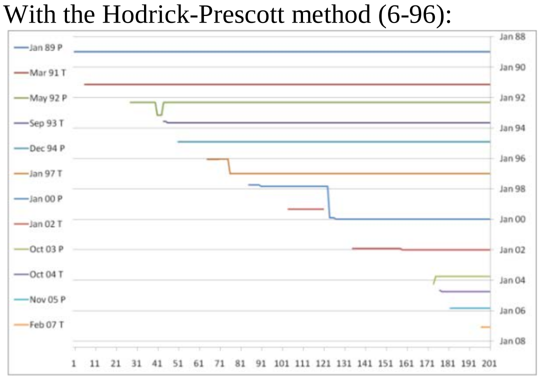
With the Christiano-Fitzgerald method
6-96 vs. 12-120



2. We also created the simulated real-time TP identification graphs for the "USA net new orders" series. These show that a minor volatility appears in early TP signals, however the likelihood of a major TP revision later decreases. From the perspective of the need for stable early TP detection in real time this trade-off is well worth to be taken.

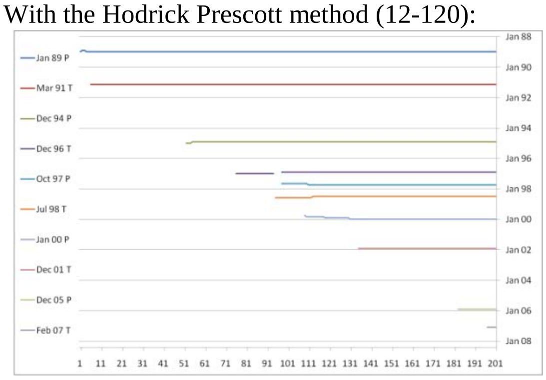
Turning point estimation history

With the Hodrick-Prescott method (6-96):



"USA net new orders" series

With the Hodrick Prescott method (12-120):

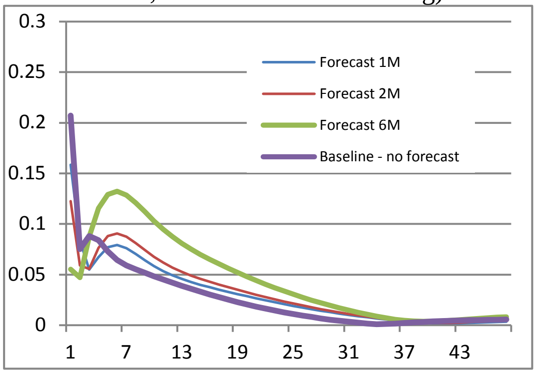


6.2. Stabilizing forecasts

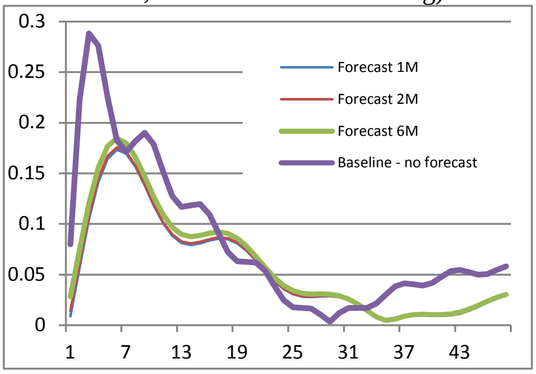
We tested the methods to see the effects of stabilizing forecasting. Our intuition was that by forecasting at each iteration we would compensate for the highly asymmetric nature of our band pass filters at the end of the time series and have beneficial effects on the stability of the cyclical estimate. Our results showed that forecasts improve revision patterns in early vintgages but at the expense of shifting and preserving the relatively high revisions at the early vintages for later vintages. As we can see from the graphs below for the HP filter the forecast horizon strongly influences the extent to which this shift occurs. Longer forecasts decrease more considerably the early vintage revisions but at the same time their impact on late vintage revisions is also large – both in size and persistence. Therefore stabilizing forecasts should only be used with short horizons for the HP filter. For the CF filter this forecast horizon dependence is less important. The beneficial effects of the stabilizing forecasts are stronger in trending time series, however these effects disappear in stationary, or cyclical series, like the business and consumer climate indicators.

Effects of forecasting before applying the filter "USA net new orders" series

With the Hodrick-Prescott method (baseline + 12-120 with 1,2 and 6 month forecasting):



With the Christiano Fitzgerald method (baseline + 12-120 with 1, 2 and 6 month forecasting)



The short horizon forecasts have an additional benefit on Hodrick-Prescott filter performance, notably they decrease significantly the first order autocorrelation that was a discomforting property of the HP filter.

7. Conclusion

Both the CF-filter and the HP-filter performs better than the Phase Average Trend filter.

Therefore the use of the HP-filter is recommended if the early, clear and steady turning point signals are a priority. In an OECD Composite Leading Indicators context this is clearly the case; a filter band of 12-120 months and the use of series specific stabilizing forecasting with forecast horizons from 0 to 2 month are put in place.

The use of the CF filter is recommended when priority is given to minimizing cumulative revisions. With the CF filter a noisy, oscillating signal arises in real time applications, but in return the initial estimates of a cyclical value are the closest to the final long term cycle value.

As a result – since the OECD CLIs aim to signal turning points, but do not attempt to give precise/exact estimates of the output gap – OECD will change its de-trending method in its CLI methodology to the double HP filter with a 12-120 month filter band specification, and a series dependent stabilizing forecasting.

References

A. Agresti, B. Mojon (2001), "Some Stylized Facts on the Euro Area Business Cycle", *ECB Working Paper series, WP No.* 95

M. Baxter, **R. G. King (1999),** "Measuring Business Cycles: Approximate Band-Pass filters for Economic Time Series", *The Review of Economics and Statistics*, *Vol.* 81, *No.* 4, pp. 575-593

C. Boschan, W.W. Ebanks (1978), "The Phase Average Trend: A New Way of Measuring Economic Growth", *Proceedings of the Business and Economic Statistics Section. American Statistical Association*, Washington, D.C.

G. Bry, C. Boschan. (1971), "Cyclical Analysis of Time Series: Selected Procedures and Computer Programs," *Technical Paper* 20, *NBER*

L. J. Christiano, T. J. Fitzgerald (1999), "The Band Pass Filter" *NBER Working Paper No.* W7257.

T. di Fonzo (2005), "The OECD Project on Revisions Analysis: First Elements for Discussion" paper submitted to the OECD STESEG meeting on 27-28 June 2005, Paris. http://www.oecd.org/dataoecd/55/17/35010765.pdf

J. D. Hamilton (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle*"*, *Econometrica*, *Vol.* 57, *No.* 2., pp. 357-384.

D. Harding, A. Pagan (2003), "A Comparison of Two Business Cycle Dating Methods", *Journal of Economic Dynamics and Control*, Elsevier, vol. 27(9), pp. 1681-1690, July

R.J. Hodrick, E.C. Prescott (1997), "Postwar U.S. Business Cycles: an Empirical Investigation", *Journal of Money Credit and Banking* 29 (1), pp. 1–16.

A. Maravall (2006), "An Application of the TRAMO-SEATS Automatic Procedure; Direct versus Indirect Adjustment", *Computational Statistics & Data Analysis No*. 50, pp. 2167 – 2190.

A. Maravall, A. del Rio (2001), "Time Aggregation and the Hodrick-Prescott Filter", *Banco de Espana Working Paper Series* 2001(08).

Newey W.K. and K.D. West (1987), "A simple positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica*, 55: 703-708.

A. Orphanides and S. van Norden (2002) "The unreliability of output-gap estimates in real time" *The Review of Ecomimics and Statistics*, Vol. 84(4), pp. 569-583

D.S.G. Pollock (2006), "Statistical Fourier Analysis: Clarifications and Interpretations", *available online at the Queen Mary University of London*: http://www.le.ac.uk/users/dsgp1/SIGNALS/FIVESTAT.pdf

OECD Statistics Note, October 2007, "Changes to the OECD's Composite Leading Indicator", available at: http://www.oecd.org/std/leading-indicators/39430336.pdf

V. Zarnowitz and A. Ozyildirim (2006), "Time series decomposition and measurement of business cycles, trends and growth cycles", *Journal of Monetary Economics*, Vol. 53 pp. 1717–1739

Further documents related to the OECD CLI system can be found on the OECD Business Cycle Analysis webpage: http://stats.oecd.org/mei/default.asp?rev=2

Annex A

Revision results for New Car Registration over 200 vintages

	PAT	HP	CF	Rating scores		
				PAT	HP	CF
Mean	Negative bias over first 30 vintages	No bias	No bias	1	3	3
Mean Abs. Dev.	Strong over first 125 vintages	Strong over first 30 vintages	Strong over first 20 vintages then oscillating	1	3	3
Standard Dev.	High over first 150 vintages	High over first 12 vintages	High over first 20 vintages then oscillating	1	3	2
Simple mean test						
First order AC	Not significant	Significant	Significant	3	2	2
Newey-West stdev.	High over first 150 vintages	High over first 20 vintages	High over first 20 vintages	1	3	3
HAC mean test						
Sign Change	High over first 90 vintages	High over first 12 vintages	Very high over first 20 vintages then oscillating	1	3	2
Relative Revision	High over first 36 vintages	High over first 12 vintages	High over first 25 vintages then oscillating	1	3	2
Direction Change	High over first 3 vintages	High over first 3 vintages	High over first 3 vintages then oscillating	3	3	2
Cumulated Abs Revision	High over first 24 vintages	Medium over first 24 vintages	Low over first 24 vintages	1	2	3
Cumulated Revision	Not significant	Not significant	Not significant	3	3	3
Total score				**16**	**28**	**25**

Revision results for Overtime Hours, manufacturing over 200 vintages

	PAT	HP	CF	Rating scores		
				PAT	HP	CF
Mean	Negative bias over first 50 vintages	No bias	No bias	1	3	3
Mean Abs. Dev.	Strong over first 90 vintages	Strong over first 20 vintages	Strong over first 20 vintages	1	3	3
Standard Dev.	High over first 120 vintages	High over first 20 vintages	High over first 20 vintages	1	3	3
Simple mean test						
First order AC	Not significant	Significant	Significant	3	2	2
Newey-West stdev.	High over first 120 vintages	High over first 20 vintages	High over first 20 vintages	1	3	3
HAC mean test						
Sign Change	High over first 10 vintages	High over first 10 vintages	High over first 10 vintages	3	3	3
Relative Revision	High over first 80 vintages	High over first 20 vintages	High over first 20 vintages then oscillating	1	3	2
Direction Change	High over first 3 vintages	Slightly high over first 3 vintages	Very high over first 3 vintages then oscillating	2	3	1
Cumulated Abs Revision	High	High	Medium	2	2	3
Cumulated Revision	Significant	Significant	Not significant	2	2	3
Total score				**17**	**27**	**26**

Revision results for Ratio Imports to Exports over 200 vintages

	PAT	HP	CF	Rating scores		
				PAT	**HP**	**CF**
Mean	Negative bias over first 30 vintages	Negative bias over first 10 vintages	No bias	1	2	3
Mean Abs. Dev.	Strong over first 90 vintages	Strong over first 20 vintages	Strong over first 20 vintages	1	3	3
Standard Dev.	High over first 90 vintages	High over first 20 vintages	High over first 20 vintages	1	3	3
Simple mean test						
First order AC	Significant	Significant	Significant	3	3	3
Newey-West stdev.	High over first 120 vintages	High over first 20 vintages	High over first 20 vintages	1	3	3
HAC mean test						
Sign Change	High over first 30 vintages	Slightly high over first 10 vintages	High over first 20 vintages	1	3	2
Relative Revision	High over first 80 vintages	High over first 5 vintages	High over first 20 vintages then oscillating	1	3	2
Direction Change	High over first 3 vintages	Slightly high over first 3 vintages	High over first 3 vintages then oscillating	2	3	1
Cumulated Abs Revision	High	High	Medium	2	2	3
Cumulated Revision	Significant	Not significant	Not significant	1	3	3
Total score				**14**	**28**	**26**

Revision results for Business climate indicator (PMI) over 200 vintages

	PAT	HP	CF	Rating scores		
				PAT	**HP**	**CF**
Mean	Positive bias over first 30 vintages then negative bias over next 30 vintages	Negative bias over first 10 vintages	No bias	1	3	2
Mean Abs. Dev.	Strong over first 50 vintages	Strong over first 10 vintages	Strong over first 20 vintages	1	3	2
Standard Dev.	High over first 60 vintages	High over first 10 vintages	High over first 20 vintages	1	3	2
Simple mean test						
First order AC	Significant	Significant	Not significant	2	1	3
Newey-West stdev.	High over first 60 vintages	High over first 15 vintages	High over first 20 vintages	1	3	2
HAC mean test						
Sign Change	High over first 20 vintages	Slightly high over first 5 vintages	High over first 10 vintages	1	3	2
Relative Revision	High over first 20 vintages	High over first 5 vintages	High over first 15 vintages then oscillating	1	3	2
Direction Change	Slightly high over first 3 vintages then oscillating	High over first 3 vintages	High over first 3 vintages then oscillating	2	3	1
Cumulated Abs Revision	Very high	High	Medium	1	3	2
Cumulated Revision	Significant	Not significant	Not significant	1	2	3
Total score				**12**	**27**	**21**

Revision results for Consumer sentiment over 200 vintages

	PAT	HP	CF	Rating scores		
				PAT	**HP**	**CF**
Mean	Positive bias over first 50 vintages then negative bias over next 50 vintages	Positive bias over first 10 vintages	Negative bias over first 15 vintages	1	3	2
Mean Abs. Dev.	Strong over first 30 vintages	Strong over first 10 vintages	Strong over first 20 vintages	1	3	2
Standard Dev.	High over first 90 vintages	High over first 5 vintages	High over first 20 vintages	1	3	2
Simple mean test						
First order AC	Not significant	Significant	Not significant	3	1	3
Newey-West stdev.	High over first 90 vintages	High over first 10 vintages	High over first 15 vintages	1	3	2
HAC mean test						
Sign Change	High over first 20 vintages	Slightly high over first 5 vintages	Very high over first 10 vintages	2	3	2
Relative Revision	High over first 15 vintages	Slightly high over first 5 vintages	High over first 15 vintages then oscillating	2	3	1
Direction Change	Slightly high over first 3 vintages then oscillating	Slightly high over first 3 vintages	High over first 3 vintages then oscillating	2	3	1
Cumulated Abs Revision	Very high	High	Medium	1	2	3
Cumulated Revision	Significant	Not significant	Not significant	1	3	3
Total score				**15**	**27**	**21**

Revision results for Net New Orders – durable goods over 200 vintages

	PAT	HP	CF	Rating scores		
				PAT	**HP**	**CF**
Mean	Positive bias between 20th and 50th vintages	Positive bias over first 5 vintages	Negative bias over first 10 vintages	1	3	2
Mean Abs. Dev.	Strong over first 20 vintages	Strong over first 10 vintages	Strong over first 20 vintages	2	3	2
Standard Dev.	High over first 90 vintages	High over first 5 vintages	High over first 20 vintages	1	3	2
Simple mean test						
First order AC	Not significant	Significant	Significant	3	2	2
Newey-West stdev.	High over first 90 vintages	High over first 20 vintages	High over first 20 vintages	1	3	3
HAC mean test						
Sign Change	Low over all vintages	Low over all vintages	High over first 10 vintages	3	3	1
Relative Revision	High over first 60 vintages	Slightly high over first 5 vintages	High over first 20 vintages then oscillating	1	3	2
Direction Change	High over first vintage then oscillating	Slightly high over first 2 vintages	High over first 10 vintages then oscillating	2	3	1
Cumulated Abs Revision	Very high	High	Medium	1	2	3
Cumulated Revision	Not significant	Not significant	Not significant	3	3	3
Total score				**18**	**28**	**21**