Discussion paper

Inflation indicators

Co-integration in Structural Time Series models

Sara Houweling
Pim Ouwehand

March 2018

Content

Summary
Last year, Statistic Netherlands released the 'price dashboard', consisting of seventeen indicators related to inflation. We investigated the relations between the time series of the inflation indicators using structural time series models. We used the notion of co-integration to determine whether some of the series show similar behaviour over time. We found that most of them are related to other indicators within the dashboard, but only to a certain extent. We identified groups of series that have a high degree of co-movement.

Keywords
Inflation, Price dashboard, Consumer Price Index, structural time series modelling, co-integration

1. Introduction[1]

What is inflation? Last year, Statistics Netherlands has made a rigorous shift in the answer to this simple question. To explain the reason for this shift, we take a closer look at inflation. A standard way to explain inflation is as a sustained increase in the general price level. The problem however with this standard definition is that it is very difficult to measure the general price level (Kazemier, Zeelenberg and Walschots, 2017). Therefore most measures of inflation consider a basket of consumer goods and measure price changes of this basket. This approach was followed by Statistics Netherlands as well, and led to the publication of inflation as a single figure ('the inflation of February 2017 is 1.8%'). Recently this strategy was revised to acknowledge that price changes in consumer goods are not the only element of inflation. For example, price changes in consumer goods are very different from price changes in houses, which can be considered as part of inflation as well. Therefore Statistics Netherlands has created the 'price dashboard' (CBS, 2017) in which price changes are placed in a broader perspective.

In the price dashboard the economy is divided into four parts, each of which has four individual price indices. Together with the consumer price index they give an overview of price changes in the overall economy. The four economic categories that are part of the price dashboard are:

- Household consumption
- Capital market
- Real estate and investment
- Production of goods and services

In the next chapter we describe the individual indices in detail.

In this paper we investigate the relationships between the indices of the price dashboard. Which indices are leading or lagging other indices? How strong is the relation between the indices? Investigating this will lead to a better understanding of how the inflation indicators behave over time and how close their relation is.

Special attention in given to the consumer price index (CPI). Although this is no longer the only measure of inflation, consumer prices are still an important measure to policy makers. For example the European Central Bank bases its monetary policy on consumer prices. The ECB measures inflation by the HICP (Harmonized Index of Consumers Price), which is very similar to the CPI. One of the main goals of the ECB is to keep the HICP in the euro area at a level close to, but below 2% (ECB, 2011). In the Netherlands the CPI is used for example in maximum house rental fee increases and in wage negotiations. Because of the central role of the consumer prices in measuring

[1] Many thanks to Kees Zeelenberg for his support and comments. Thanks to Jan van der Brakel for reviewing this paper.

inflation, we zoom in to the behaviour of the CPI and to possible indices that are leading to the CPI. However, it is not our primary goal to predict the CPI. If we would want to predict the CPI, we would need to take factors into account from outside the dashboard as well. Other research shows that for example monetary values (M0-M3) and worldwide commodity prices are important to predict the CPI (Artis et al 1995, Seitz 1998). These are not included in the dashboard. We have limited ourselves to the indicators within the price dashboard and investigate to which extent they have predictive value for the CPI, without trying to make a prediction for the CPI.

Part of this analysis is already done in Houweling and Zeelenberg (2017). In this study we investigated the relationships between the seventeen indices by finding the lag under which correlation is maximised. This provided an answer to the question what the lead time is between any pair of indicators and how close their relation is. Furthermore, we performed an analysis on turning points to see if leading indicators could successfully predict turning points in CPI.

In this paper we use structural time series to answer the same questions in a more rigorous way. A major advantage of the structural time series approach is that it offers the possibility to determine the correlation between trend components. The trend component does not include the season and irregular fluctuation and is therefore a smoothed version of the original series. The correlation between trends measures an underlying similarity between series that is supposed to be more structural than the correlation that was calculated in Houweling and Zeelenberg (2017). If there is a strong correlation in trend between two series, the unobserved trend components are driven by common factors. This means that they have a similar underlying pattern. If two series have common trends, one series can be used to forecast the other series, for example in a now-casting model.

2. Data description

All data used in this research is retrieved from the CBS price dashboard. This data is on a monthly level and ranges from January 1997 until July 2017. There are three series that were not available for the full period. The 'import industrial products' starts in January 2000 and 'import machinery' starts in January 2005. For 'new build houses there is only data available until February 2017. The full list of the seventeen price dashboard elements is as follows:

- CPI
- Household consumption
    - CPI energy
    - CPI industrial goods (excluding energy)
    - CPI foods
    - CPI services
- Financial markets
    - AEX share price
    - 3-months interest rate
    - Gold price
    - 10-year interest rate (government loans)
- Real estate and investment
    - Price index for privately owned houses
    - Price index of new build houses (proxy)
    - Producer price index of capital goods produced in the Netherlands
    - Import price index of machinery[2]
- Production of goods and services
    - Price index of imports of industrial products
    - Wage rate
    - Price of crude oil
    - Output price of industry

Below, the four segments (except CPI) are described in more detail (CBS, 2017; Kazemier, Zeelenberg and Walschots, 2017).

Household consumption

The first segment is household consumption, as measured in the CPI. Selected components in household consumption are food, energy (natural gas, electricity, et cetera), industrial goods (excluding energy) and services. The latter two form the core part of inflation of household consumption.

[2] Note that the "import price of machinery" is based on a much smaller category than "capital goods produced in the Netherlands producer price". Means of transport and computers are not included in the machinery category, but are included in the capital goods category.

Capital market

The second segment concerns the financial markets. Important indicators here are the long term interest rate (the interest rate on the newest 10-year government loan), the short term interest rate (3-month Euribor) and the index of share prices on the Dutch stock market (AEX). The two interest rate series are the only series on the dashboard that are not 12-month percentage changes.
Statistics Netherlands had some difficulty in selecting a fourth indicator, not being an interest rate. It was considered to select the exchange rate of foreign currency. However, a large part of the international trade in goods of the Netherlands is within the euro zone: 41 percent for imports in 2016 and 53 percent for exports. Moreover, on average the exchange rate of the euro against the currency of the main non-EU trade partners of European Union did not change very much during the last 15 years. Therefore, preference was given to the price of gold as the fourth indicator.

Real estate and investment

The third segment is fixed assets. Ideally, this segment contains information on the prices of private homes (new buildings and existing houses) and other buildings (offices, shops, warehouses, industrial buildings and so on), the rent of commercial property, the price of machinery, ships, aircraft, vehicles, et cetera. However, only few of these prices are available on a monthly or quarterly basis. For real estate, the price of existing privately owned dwellings and a proxy for the price of new houses have been chosen. Investments is represented by the import price of machinery, equipment and tools and the producer price of capital goods produced in the Netherlands.
The price of newly built houses is approximated by the construction costs of new buildings. The difference is that the price of new buildings also includes the profit margin of the project developer as well as the price of the land on which the property is built.
The construction cost of new buildings is only available on a quarterly basis. In order to reach a monthly series of percentage year-on-year mutations, the quarterly indices are first projected on the middle month of the relevant quarters. The values for the intermediate months are obtained by linear interpolation. For the most recent months, the value of the most recent index available is duplicated. After that, the 12-month percentage changes are calculated.

Production of goods and services

The last segment concerns the production of goods and services. As indicators Statistics Netherlands has chosen wage rates, the price of imports of goods and services (excluding energy) and the price of energy. The latter is represented by the price of crude oil. In addition, the output price of industry is important. The output prices of services are not included because they are not available on a monthly basis. However, they are strongly related to wages, and wages are included in this segment. The wage rate refers to the index of basic wage rates in collective agreements; this excludes overtime payments and additional payments.

3. Model

The model we are using is a standard structural time series model described first by Harvey (1989) and later by Durbin and Koopman (2012) and Koopman and Commandeur (2007). In a structural time series model the series can be decomposed into for instance a trend component, seasonal component and an irregular component. All components can follow a stochastic process. In this model the observations y_t can be expressed as a trend μ_t plus a seasonal component γ_t and an irregular term ε_t. This is called the observation equation:

$$y_t = \mu_t + \gamma_t + \varepsilon_t \qquad \text{where } \varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$$

The trend μ_t has a level component and a slope component ν_t. The level and the slope are not observed and are therefore called unobserved components. The state equations for the unobserved level and slope are:

$$\mu_{t+1} = \mu_t + \nu_t + \xi_t \qquad \text{where } \xi_t \sim NID(0, \sigma_\xi^2)$$
$$\nu_{t+1} = \nu_t + \zeta_t \qquad \text{where } \zeta_t \sim NID(0, \sigma_\zeta^2)$$

The disturbance terms ξ_t and ζ_t in the above equations are mutually independent and make sure that the level and slope can vary over time. So unlike a linear regression model where the slope is fixed, the slope in a structural time series model can change over time.

For the seasonal component γ_t we choose to use the trigonometric seasonal state equations (Durbin and Koopman, 2012):

$$\gamma_t = \sum_{j=1}^{[\frac{s}{2}]} \gamma_{j,t}^{+}$$
$$\begin{pmatrix} \gamma_{j,t+1}^{+} \\ \gamma_{j,t+1}^{*} \end{pmatrix} = \begin{pmatrix} \cos\lambda_j & \sin\lambda_j \\ -sin\lambda_j & \cos\lambda_j \end{pmatrix} \begin{pmatrix} \gamma_{j,t+1}^{+} \\ \gamma_{j,t+1}^{*} \end{pmatrix} + \begin{pmatrix} \omega_{j,t}^{+} \\ \omega_{j,t}^{*} \end{pmatrix}$$

Where λ_j denotes the frequency of the seasonal pattern with (for a monthly pattern):

$$\lambda_j = \frac{2\pi j}{12} \qquad \text{for } j = 1,2, \dots, 6.$$

The disturbance terms are normally, independently distributed:

$$\begin{pmatrix} \omega_{j,t}^{+} \\ \omega_{j,t}^{*} \end{pmatrix} \sim NID\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_\omega^2 I_2 \right), \quad j = 1, \dots, [s/2]$$

The reason we use a trigonometric seasonal pattern instead of a dummy seasonal pattern is that the model-fit with a trigonometric season was clearly better[3], since it allows for more flexible seasonal patterns.

[3] There was a substantial amount of autocorrelation for the CPI series when we used the dummy seasonal model.

The model is estimated using the Stamp (Structural Time Series Analyser Modeller and Predictor) module in the OxMetrics software package. As a first step, it will use its default maximum likelihood estimation method to estimate the variance of all stochastic disturbance terms. In the second step the Kalman filter and a smoothing algorithm are applied to estimate the smoothed state variables at every point in time. All of this is basic Stamp functionality; we did not make any changes to the default configuration. Details on these procedures can be found in Koopman et al. (2007).

Within this framework we will be testing whether all components need to be included for the individual series. For example, not all series have a seasonal pattern. Therefore not all series have the seasonal component γ_t included in their optimal structural time series model. Moreover, there are some series that do have a seasonal pattern, but the seasonal pattern barely varies over time. In this case we can omit the seasonal disturbance terms $\omega_{j,t}^{+}$ and $\omega_{j,t}^{*}$ from the model, but leave the seasonal component γ_t in. In this case there will be a seasonal pattern that is the same in each year. This is called a "deterministic seasonal component", whereas the model with the seasonal disturbance terms included has a "stochastic seasonal component". Just like there are series without seasonal pattern or seasonal variance, there are series without a slope or without variance in slope. In this case we can omit the slope component ν_t or its disturbance term ζ_t. We assume that all series do have the level component μ_t. However, it could be that the level does not vary over time. In this case the level disturbance term ξ_t can be omitted.

So to find the best model for a series, we need to determine for the slope and seasonal component whether they should be included in the model or not and for all components that are included we need to determine whether they should be able to change over time (stochastic) or whether they are stable over time (deterministic).

To determine the correlation between series they need to be combined into one multivariate model. The multivariate model is very similar to the univariate model. In the multivariate model the y_t terms become vectors of observations. In a bivariate model:

$$y_t = \begin{pmatrix} y_t(1) \\ y_t(2) \end{pmatrix}$$

And the observation equation becomes:

$$\begin{pmatrix} y_t(1) \\ y_t(2) \end{pmatrix} = \begin{pmatrix} \mu_t(1) \\ \mu_t(2) \end{pmatrix} + \begin{pmatrix} \gamma_t(1) \\ \gamma_t(2) \end{pmatrix} + \begin{pmatrix} \varepsilon_t(1) \\ \varepsilon_t(2) \end{pmatrix}$$

For the level, slope and seasonal components equations can be formulated analogously. Also the disturbance terms become vectors and their variance is now described in a variance matrix. For the level disturbance vector in a bivariate model we get:

$$\xi_t = \begin{pmatrix} \xi_t(1) \\ \xi_t(2) \end{pmatrix}$$

The variance matrix for ξ_t has the following form:

$$\Sigma_\xi = \begin{pmatrix} \sigma_{\xi(1)}^2 & \rho_\xi \; \sigma_{\xi(1)} \, \sigma_{\xi(2)} \\ \rho_\xi \; \sigma_{\xi(1)} \, \sigma_{\xi(2)} & \sigma_{\xi(2)}^2 \end{pmatrix}$$

On the diagonal we see the level disturbance variance for the first and the second series $\sigma^2_{\xi(1)}$ and $\sigma^2_{\xi(2)}$. The $\rho_\xi \in [-1,1]$ is the correlation between the level of the two series. When ρ_ξ is close to one the two series have cointegrated levels and the level components of the two series will move up and down at the same time. When the two series have cointegrated levels only one stochastic component is needed to explain the variance in level for both series, so the variance matrix will be reduced to rank one. For a detailed description of the concept of common levels and correlation in level, see Commandeur and Koopman (2007), section 9.3.

In a multivariate model of n series the rank of the level variance matrix measures the total amount of variability in level within the n series. So the rank $r \leq n$ measures how many different stochastic components are needed to describe the variability in all level terms. If the matrix is less than full-rank ($r < n$) the series have common levels and we say that the levels are cointegrated.

Even if we do not find co-integration because ρ_ξ is not close enough to one, it still measures the co-movement between the two series. The closer ρ_ξ is to one, the more the level components of the two series move together in the same direction. So the ρ_ξ is a measure for correlation and has a similar interpretation as the correlation we found in Houweling and Zeelenberg (2017). Note that the ρ_ξ only measures correlation in level, whereas the correlation in Houweling and Zeelenberg (2017) measures correlation in the seasonally adjusted month-on-month changes.

So far, we have only discussed correlation in level. Using the same method, the correlation between the slope and season of two series can also be modelled. In this research we will work with correlation in level only, because models with a stochastic level turned out to have the best fit for the data. More details on the model-fit are given in section 4. Also note that in the model with a stochastic level and deterministic slope the level correlation is the same as the trend correlation. This is because the trend consist of level and slope and there is no variability in the slope.

Besides structural time series, we did consider other methods such as VAR (vector autoregressive) models. We chose the structural time series model because of the possibility to explicitly determine the correlation in trend between series. This is not possible in a VAR analysis. Another advantage of the multivariate structural time series model is the possibility to investigate whether the variance in level for the indicators can be explained by a limited number of common factors.

Data format

We choose to work with the log of the index series for this analysis. This is different from Houweling and Zeelenberg (2017), were we choose to work with the seasonally adjusted month-on-month changes. The reason we choose to work with index series in this research is that the structural time series models model can find the season itself and it is best to use the untransformed series. Taking the log of the series is a necessity, because the CPI is a multiplicative series.

Interest rate data format

The two interest rate series are different from the other series in the sense that they have no index series. They are not directly a price and therefore not directly transformable into an index series. Therefore we choose to include them in their primary form. Alternatives would have been to either create an index series by interpreting the interest rate as a price or to transform the other series to their first differences and compare these to the interest rate series. We did not choose the first alternative because there are negative values for the three months interest rate which gives a problem in taking the log of the index series. We did not choose the second alternative, because we prefer to use the index series for the other series.

4. Results

4.1 Introduction

As a first step, we will investigate univariate models for individual series to determine the optimal model per series. Next, we will derive a 'standard model'. This is a single model that is optimal for most series. For the series where it is not the optimal model, it is still a reasonable alternative. This model will form the basis for a multivariate extension of the analysis. Modelling the series in a multivariate setting is more convenient when the structure of the model is the same for all series in the model. When we have determined the standard model, we first use this model in order to analyse correlation factors between trends and optimal lags for indicators that are related to the CPI in bivariate models. Finally, we create one large multivariate model with all seventeen series included to find all correlation coefficients. The variance matrix of the level disturbance terms of this model provides information about the total amount of variance in all seventeen series together.

4.2 Univariate analysis

We start by examining some series individually to find the best fitting structural time series model. The series that are analysed in detail are the series for which we expect to find correlation with the CPI or a part of the CPI (see Houweling and Zeelenberg (2017)):

- CPI services and wages
- CPI energy and crude oil prices
- Price import industrial products and CPI
- Capital goods producer price and CPI
- Import price machinery and CPI

Within the framework presented in chapter 3, we test which of the components (level, slope and season) should be included in a univariate model and if they are included, whether they should be stochastic or deterministic. Based on the univariate results we will select one standard model for the multivariate analysis.

The model parameters are estimated using the Stamp (Structural Time Series Analyser Modeller and Predictor) module in the OxMetrics software package. The different models are compared based on model-fit, log-likelihood and number of parameters to be estimated. The model-fit is checked by looking at autocorrelation, homoscedasticity and normality of the innovations (see Koopman et al., 2009). For the models with a good model-fit, we select the ones with the highest log-likelihood. If there are multiple models with a high log-likelihood we select the model with the smallest number of parameters to be estimated.

Note that we found different strategies to select the best structural time series model for a set of data in the literature. Quite some authors impose a smooth trend to the model (deterministic level, stochastic slope) without any quantitative arguments (e.g. Koopman and Lee (2009), van den Brakel et al. (2016)). The rationale behind choosing the smooth trend model is that it gives a trend that is smooth (as the name already suggests) and stable, which corresponds with the intuition behind a trend. The idea of using a smooth trend model to impose a smooth trend is mentioned by e.g. Durbin and Koopman(2012). Also Nyblom and Harvey (2001) argue that the smooth trend model is often found to give a good fit to economic series. Furthermore, we see that the quantitative criteria to find the best fitting model differ. In the Stamp manual (Koopman et al. 2007) the prediction error variance (PEV) is presented to determine goodness of fit. This is different from Commandeur, Koopman (2007), who find the best model using the AIC. The method we chose for model selection is in line with Commandeur and Koopman (2007). So we did not impose a smooth trend model, but choose an approach based on model-fit. We do not follow the AIC as strictly as they do, because the AIC cannot be compared exactly between models with a different number of state variables. This is because the number of observations at the start of the series that have to be disregarded as part of the initialisation of the Kalman filter is not equal for models with a different number of state variables. However, this effect is relatively small due to the large number of 247 periods in our time series and thus the log-likelihood and AIC still give valuable information in the model selection process.

The selected models per series are printed in table 1.

	CPI services	**wages**	**CPI energy**	**Crude oil price**	**Capital goods**	**Import industrial goods**	**CPI**
Level	Stoch	stoch	Stoch	stoch	stoch	det	stoch
Slope	Det	stoch	det	-	det	stoch	det
Season	Stoch	stoch	det	-	det	-	stoch

Table 1 Optimal model - stoch = stochastic - det = deterministic

Standard model

To combine series in one multivariate structural time series model we need to select one univariate standard model that has the best fit to most of the series. Technically, it would be possible to have a multivariate model where components are included or excluded on an individual basis. Because of simplicity and limitations in Stamp we choose to use one standard model for all series.

A first observation is that the seasonal component needs to be included and needs to be stochastic, because otherwise the autocorrelation will become unacceptably large for some series. A second observation is that both level and slope will be included in the model, because they are included in almost all the individual best models. We see that in most optimal models only one of the two is stochastic. Also from a practical point of view it is preferred that only one of them is stochastic, because otherwise the model would give two correlation factors between state disturbances instead of one and the risk of over fitting the model becomes larger. Also, it is not clear how the two different correlation factors would need to be interpreted if we are interested in common trends between series. Because in most optimal models the level is stochastic and the slope is deterministic we select as **general model**: stochastic level, deterministic slope and stochastic season[4].

Note that the general model we chose is different from the smooth trend model (deterministic level, stochastic slope) that is often used in literature. For the purpose of this research -finding the correlation between the trends of series- we see that our standard model gives similar results[5] to a smooth trend model. Therefore we would not expect different conclusions if a model with a smooth trend was chosen.

4.3 Bivariate analysis

Before going to one large model with all series included, we first determine the correlation for the series that are expected to be correlated to the CPI or a part of the CPI. In this pairwise analysis we also investigate what is the optimal lag in the relationships. We do this by first performing a regression analysis[6] with lags 0 – 18 included. Then we select the lags with the highest t-value and coefficient to include in the multivariate analysis[7]. Also, we include the lag that was optimal according to the correlation in Houweling and Zeelenberg (2017)[8].

4.3.1 CPI services and wages

In the table below we see that the optimal level correlation occurs at lag 0. Wages lag 6 has a much lower correlation with CPI, while it had the highest correlation with CPI in the analysis of Houweling and Zeelenberg (2017).

	Level correlation with CPI services
wages lag 0	0.4815
wages lag 6*	0.2848
wages lag 12	0.3567
wages lag 15	0.2694

Table 2 correlation between wages and CPI services - *lag 6 has optimal correlation in Houweling and Zeelenberg (2017).

In the following figure the trend line of both series is shown.

[4] We checked the individual results for this model and in most cases this was ok.
[5] We performed some test to compare the results with the two different models.
[6] A regression analysis within the structural time series framework
[7] We did some checks to test this method and indeed we observed that the lags with the highest t-value had the highest correlation.
[8] Only if this was a positive lag and thus a lag where the series is leading to the CPI

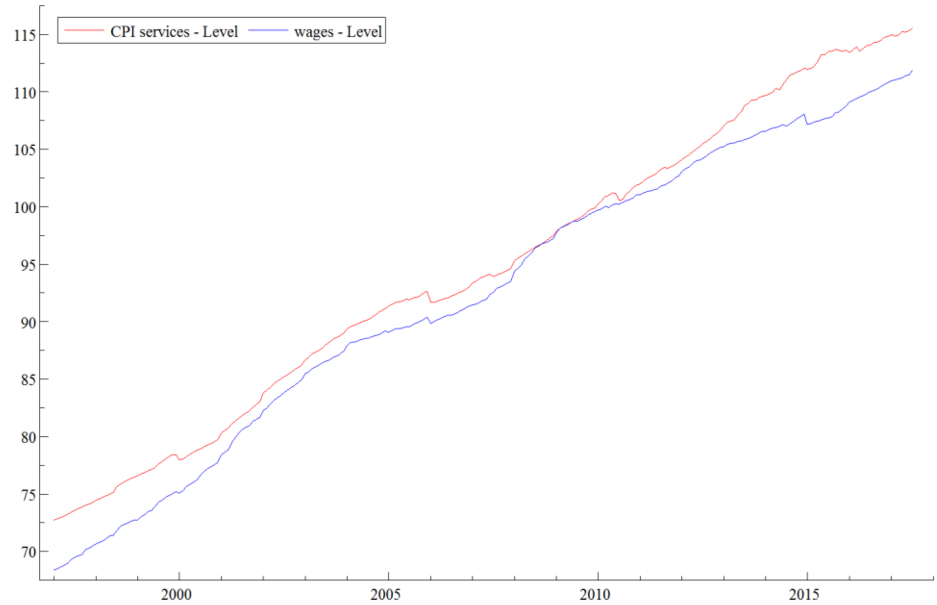

Table 3 Trend of CPI services and wages

4.3.2 CPI energy and crude oil prices

The highest level correlation is at lag 0. The lowest correlation at lag 2. Again, quite different results from the correlation analysis in Houweling and Zeelenberg (2017), where there was an optimal correlation at lag 2.

	Level correlation with CPI energy
Crude oil price lag 0	0.5868
Crude oil price lag 2*	-0.02419
Crude oil price lag 7	0.1862
Crude oil price lag 8	0.1419

Table 4 correlation between crude oil price and CPI energy - *lag 2 has optimal correlation in Houweling and Zeelenberg (2017).

4.3.3 Capital goods producer price and CPI

This is the first time that lag 0 does not have the highest correlation. We see that the correlation is much lower than in the previous two models. In Houweling and Zeelenberg (2017) the optimal lead time to CPI was negative and is therefore not included in the table below.

	Level correlation with CPI
Capital goods lag 0	0.09165
Capital goods lag 12	0.1434
Capital goods lag 13	0.1444

Table 5 correlation between capital goods producer price and CPI

4.3.4 Price import industrial products and CPI

Again, an optimal correlation at lag 0.

	Level correlation with CPI
Import industrial products lag 0	0.3243
Import industrial products lag 1*	0.09783
Import industrial products lag 13	0.1236

Table 6 correlation between price import industrial products and CPI - *lag 1 has optimal correlation in Houweling and Zeelenberg (2017)

4.3.5 Import price machinery and CPI

Here we see the interesting result that the optimal correlation is a negative correlation. This is not totally unexpected, because also in the first part of this discussion paper we saw that the optimal correlation was negative[9]. Still, we would have expected this relation to be positive because higher import prices for producers are expected to result in higher consumer prices.

	Level correlation with CPI
Import price machinery lag 0	-0.02741
Import price machinery lag 4	0.1177
Import price machinery lag 10	-0.1098
Import price machinery lag 15	-0.1970

Table 7 correlation between import price machinery and CPI

4.3.6 Summary bivariate analysis

We observe that the optimal lead time to the CPI in the structural time series model is quite different from the optimal lead time we found in Houweling and Zeelenberg (2017). There is not one series where we find the same lead time. In most cases the lead time with the highest correlation to the CPI is zero. Only for capital goods producer price we find an optimal lag of thirteen months, though this correlation of 0.14 is not very high. For import price machinery we find negative correlations which we would not expect. We conclude that including lags of series brings only very little added value when we want to determine the optimal correlation between series. Therefore we continue this analysis without taking lags of series.

Another conclusion is that there is no co-integration between the CPI or a part of the CPI and any of the other series. The highest obtained correlation is between crude oil price and CPI energy and has value 0.59. This does not come close to true co-integration where this value should be almost one.

Note that we did experiment with including interventions to adjust for outliers and level shifts in the time series. Stamp offers the possibility to automatically include intervention parameters. Adding the interventions found by Stamp will always increase the log-likelihood of the model and in most cases also the AIC. It is however doubtful whether they should be included without knowing the reason for the intervention. We investigated the effect of the interventions on the correlation

[9] Optimal correlation between import price machinery and CPI was -0.34 at lag -15.

coefficient. We noticed that for some pairs it increases and for some pairs it decreases the correlation. The increases and decreases are similar in size and not very large in absolute terms. The observation that adding interventions can either increase or decrease the correlation in trend can be explained by the following reasoning. An extreme value at the same moment in two series will increase their correlation. When the extreme value is removed from the trend by adding an intervention, this positive effect to trend correlation is removed. This is how adding interventions can decrease the trend correlation. In the case that two series do not have an extreme value at the same time, adding interventions will increase the correlation between the series, because the effect of the extreme value on the trend of one series is removed. In this case adding an intervention increases the trend correlation. Because we did not want to remove possible causes for correlation we choose to not include any interventions.

4.4 Multivariate analysis

As a last step in the CPI analysis we will include all series related to the CPI into one multivariate model. Instead of including the subsets of the CPI like CPI services and CPI energy, we now include only the total CPI. This results in one six-dimensional level disturbance variance matrix as shown below. In the top right the correlation factors between the components are shown. In the bottom left the variance terms $\rho_\xi\ \sigma_{\xi(1)}\sigma_{\xi(2)}$ are shown. The correlation factors are very similar to the correlation factors we found in the bivariate models. The reason that they can vary is that all parameters are now estimated as part of this single model. The advantage of including all series into one model is that we now also have the correlations between all series and not just their correlation with CPI.

	CPI	Wages	Crude oil	Capital goods	Industrial products	Machinery
CPI	3.9E-06	**0.39**	**0.40**	**0.16**	**0.32**	**-0.20**
Wages	1.4E-06	3.2E-06	**0.03**	**0.17**	**-0.03**	**-0.07**
Crude oil	7.3E-05	5.6E-06	0.01	**0.03**	**0.55**	**-0.03**
Capital goods	8.1E-07	7.7E-07	7.4E-06	6.2E-06	**-0.13**	**0.01**
Industrial products	3.9E-06	-3.8E-07	0.00	-2.0E-06	4.0E-05	**0.07**
Machinery	-1.3E-06	-4.0E-07	-7.4E-06	9.3E-08	1.4E-06	1.0E-05

Table 8 Variance/correlation matrix for level

Another advantage of combining all series into one model is that we can investigate whether the variance in level for the indicators can be explained by a limited number of common factors. If the variance matrix is less than full rank it means that the variance in level for one series can be expressed as a combination of variances in level for other series and thus a common level. This is not the case for our data. We find that the variance matrix has full rank and thus there are no common levels.

4.4.1 Multivariate analysis all series

As a final step we combine all seventeen series in one multivariate model. Due to extremely large calculation durations, we decided to extract the season beforehand[10]. The resulting model only has a stochastic level and deterministic slope.

[10] The season is determined by using a structural time series model using the standard model. We did notice a decline in model-fit for some series using this approach.

Correlation table

	CPI	CPI energy	CPI industrial goods	CPI food	CPI services	AEX	3-months interest	Gold price	10-year interest
CPI		0.63	0.48	0.40	0.59	0.04	0.04	0.06	0.17
CPI energy			0.06	-0.09	0.04	0.08	0.18	0.08	0.24
CPI industrial goods				0.20	0.09	-0.06	-0.06	-0.11	0.04
CPI food					0.14	-0.21	-0.03	0.02	-0.05
CPI services						-0.07	-0.16	0.02	0.00
AEX							0.11	0.00	0.21
3-months interest								-0.02	0.26
Gold price									-0.12
10-year interest									
Import industrial products									
Wages									
Crude oil price									
Output price industry									
Privately owned houses									
New build houses									
Capital goods									
Import price machinery									

	Import industrial products	Wages	Crude oil price	Output price industry	Privately owned houses	New build houses	Capital goods producer price	Import price machinery
CPI	0.27	0.36	0.35	0.37	0.05	0.20	0.11	-0.11
CPI energy	0.38	0.09	0.57	0.53	0.10	0.02	0.26	-0.09
CPI industrial goods	0.01	0.18	0.09	0.05	0.02	0.15	0.04	-0.17
CPI food	-0.08	0.14	-0.14	-0.16	0.07	0.06	-0.12	-0.02
CPI services	-0.01	0.44	0.05	0.02	-0.04	0.23	-0.10	-0.02
AEX	0.16	-0.12	0.17	0.19	-0.00	-0.10	0.02	0.00
3-months interest	0.30	-0.09	0.17	0.38	0.22	0.08	0.16	-0.05
Gold price	0.27	-0.06	0.16	0.12	0.00	0.03	0.09	0.35
10-year interest	0.22	0.04	0.26	0.31	0.12	0.05	0.05	0.05
Import industrial products		-0.05	0.54	0.75	0.02	-0.04	0.30	0.40
Wages			0.06	0.01	0.20	0.23	-0.13	-0.21
Crude oil price				0.75	0.03	-0.00	0.05	0.06
Output price industry					0.03	0.01	0.23	-0.02
Privately owned houses						0.25	0.05	-0.11
New build houses							0.16	-0.06
Capital goods								0.41
Import price machinery								

Table 9 correlation in level between the indices of the price dashboard

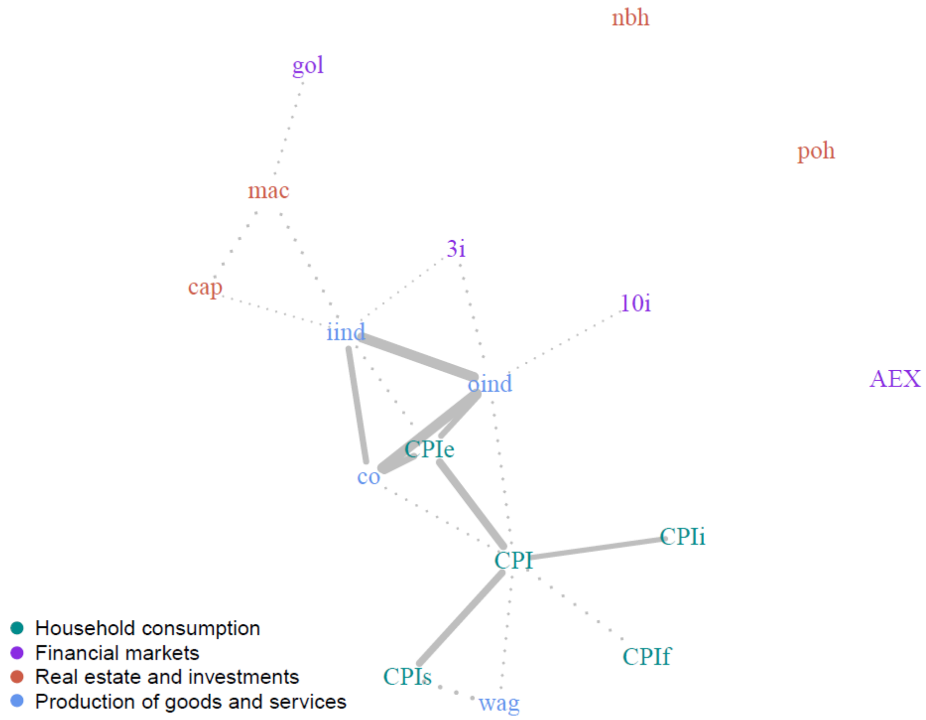

Figure 1 Correlation between indices. Includes only connections that have a higher correlation than 0.3. Dotted lines for correlations with value between 0.3 and 0.45. Colours represent price dashboard categories.

Short name	Long name
CPI	CPI
CPIe	CPI energy
CPIi	CPI industrial goods
CPIf	CPI food
CPIs	CPI services
AEX	AEX
3i	3-months interest
gol	Gold price
10i	10-year interest
iind	Import industrial products
wag	Wages
co	Crude oil price
oind	Output price industry
poh	Privately owned houses
nbh	New build houses
cap	Capital goods
mac	Import price machinery

Table 10 Short and long names for indices

A graphical representation of the relationships between the seventeen indicators is given in figure 1. It shows only correlations of 0.3 and higher. Correlations with a value between 0.3 and 0.45 are shown by a dotted line. Correlations with a value higher than 0.45 are shown by a straight line. The width of the line represents the strength of the correlation. A first observation is that three indicators are not related to any of the other indicators by a level correlation of 0.3 or higher; AEX share price index, price index of new build houses and price index for privately owned houses. For the AEX share price index this is according to expectations and similar to the results in Houweling and Zeelenberg (2017). For the house prices we were expecting a negative correlation with interest rates, but observe a small positive relation similar to findings in Houweling and Zeelenberg (2017). We expect that a shock in house prices is responsible for this. From 2008 until 2015 the house prices have been decreasing as well as the interest rates.

We see that CPI food and CPI industrial goods are only related to the CPI itself. For CPI food this is expected, though for CPI industrial goods we would expect a positive relation with import price of industrial products. We find no evidence of any relation at all because of the small correlation factor of 0.01. Also in Houweling and Zeelenberg (2017) the correlation between CPI industrial goods and import price of industrial products is relatively small (-0.28).

As expected we see relationships between wages, CPI services and CPI. Wages has the strongest relationship with CPI services and a smaller relationship to the CPI. This is also what we found in Houweling and Zeelenberg (2017).

We see a very strong relationship between the CPI and CPI energy. This is expected, because CPI energy is part of the CPI. CPI energy is part of a small group of indicators that is closely related, consisting of crude oil import price, import price of industrial products and output price of industrial products. This is the subset of indicators where most correlation is found. The strong correlation between import and output price of industrial goods can be explained by a global price level of industrial goods. There is even an overlap in the two groups. Industrial goods that are both imported and produced in the Netherlands are part of both groups. It is not unexpected that there is a high correlation between the import price of industrial products and crude oil. Some of the industrial products will be made of crude oil and both import prices are dependent on exchange rates. Also the import price of machinery is dependent on exchange rates and this might be one of the reasons that it is connected to import industrial products. The import price of machinery is connected to producer price capital goods and gold. The relationship between import price of industrial goods and gold is remarkable, but small. Houweling and Zeelenberg (2017) did not find a clear relationship between gold and any other indicator. The relationship between capital goods producer price and import price of machinery is expected, because partly there is an overlap between the groups. Some of the machinery that is imported is similar to machinery built in the Netherlands. If the producer price of such machinery increases, this will visible in both indicators.

The two interest rates indicators are only loosely connected to output and input price of industrial products. We do not expect a causal relationship, but it could be that there is a third variable outside the price dashboard that is related to both.
The level disturbance variance matrix is full-rank. This means that there are no common levels in the series in the price dashboard.

An alternative way to visualize the results is presented in figure 2. We used the hierarchical clustering to change the order of the indices such that indices with a high correlation are close to each other. This visualization includes all correlation coefficients.

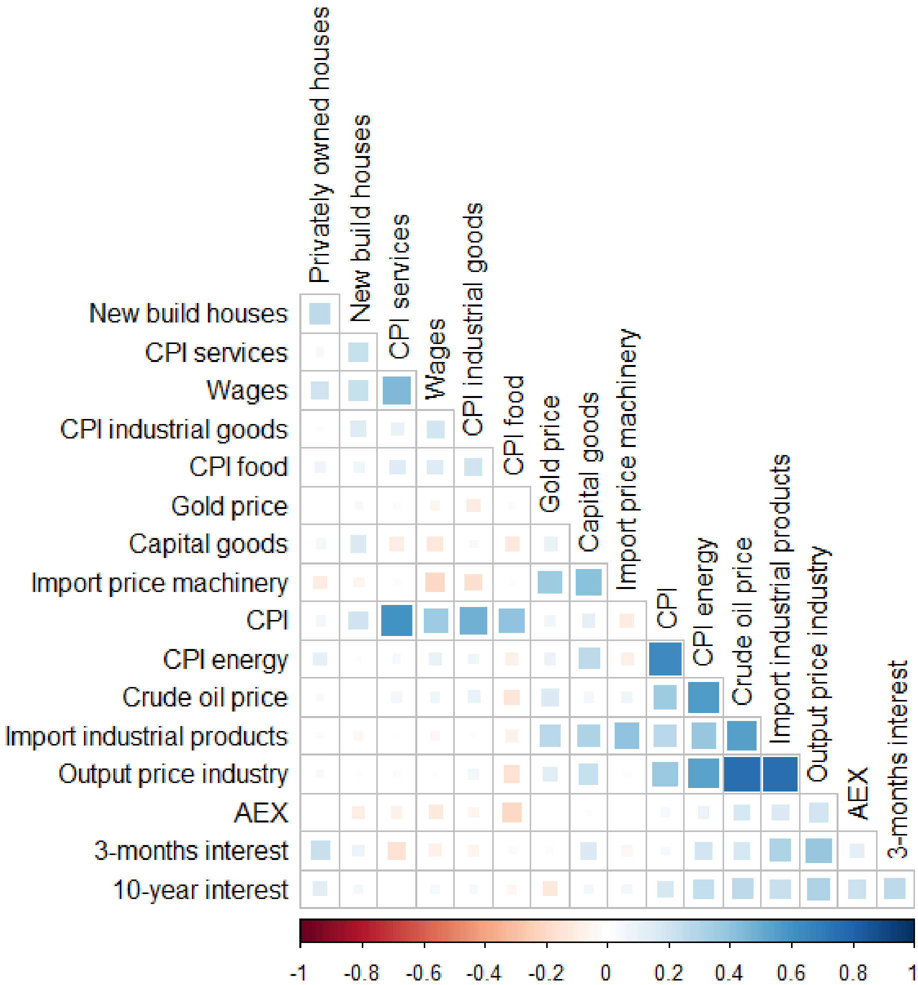

Figure 2 visualisation of correlation of indices using hierarchical clustering.

4.4.2 Difference between results

When we compare the results we found in this research to the results in Houweling and Zeelenberg (2017) we see certain similarities. Most of the series that have a large correlation in Houweling and Zeelenberg (2017) also have a large correlation in this paper. However, there are some differences.

One of the reasons for the differences could be the different data format that is used. In Houweling and Zeelenberg (2017) we work with month on month changes, whereas in this research we work with the index series. To objectively compare both methods, we would need to use the same data. However, instead of doing a comparison between both methods our goal was to get the best results for each method; that is why we choose a different data format for both analyses (see section 3).

Another reason for differences would be the different ways to handle lags. In the first study we included lagged series to find the highest correlation. In the current study we did not, because we found that often the highest correlation was at lag zero.

In figure 3 we see the differences in correlation coefficients between the two studies. If both methods were equivalent, all points should be close to the regression line. The points with a high distance to the regression line are numbered and the related indices are given in table 11.

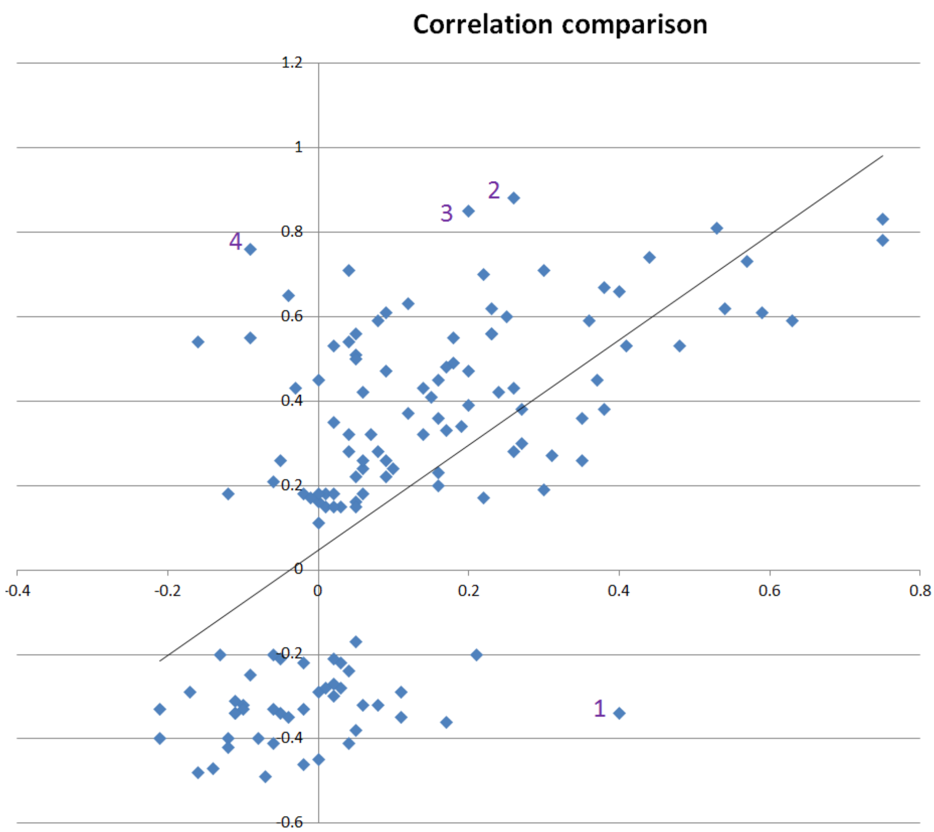

Figure 3 Correlation comparison. On x-axis results from this study, on y-axis results Houweling and Zeelenberg (2017) with a regression line added.

Number in figure 2		
1	Import price of industrial products	Import price of machinery
2	10 year interest rate	3 months interest rate
3	Wages	Price of privately owned houses
4	3 months interest rate	Wages

Table 11 Explanation of numbers in figure 2

For the correlation between Import industrial products and import machinery (point 1) we see a positive correlation of 0.4 in this study, whereas it has a negative correlation of -0.34 in Houweling and Zeelenberg (2017). One of the reasons could be the lag of 10 months in import price machinery that gave the highest correlation in Houweling and Zeelenberg (2017).

For the correlation between 3 months and 10 years interest rate (point 2) we see that the correlation of 0.88 in Houweling and Zeelenberg (2017) is clearly higher than the correlation of 0.26 we found in this study. This could not be due to a difference in lag, because the correlation of 0.88 was found at lag 0. It could also not be because of a difference in data format, because in both studies we used the interest rates in their primary form. The two interest rate indicators are expected to be correlated, so this is the only case where the results from Houweling and Zeelenberg (2017) are more in line with our expectations.

For the correlation between wages and privately owned houses (point 3) we see in Houweling and Zeelenberg (2017) a correlation of 0.85 when privately owned houses has a lag of 18 months, whereas the correlation is only 0.2 in this study. We would not expect such a high correlation between these two variables, because we do not expect a strong causal relation between the two variables. They could both be dependent on the growing economy.

For the correlation between wages and 3 months interest rate (point 4) we see in Houweling and Zeelenberg (2017) a correlation of 0.76 when 3 months interest rate has a lag of 3 months, whereas the correlation is only -0.09 in this study. Also in this case, we would not expect such a high correlation between these two variables, because we do not expect a strong causal relation between the two variables.

In general we could say that the results from the current studies are more according to what we would expect. For most pairs of series with a large difference in correlation we see a high correlation in Houweling and Zeelenberg (2017) and a low correlation in the current study. In most cases we would not expect a high correlation, so the current study seems to give better results. The one exception is the pair of interest rate series.

The graphical representation of the relationships between the seventeen indicators for the results of Houweling and Zeelenberg (2017) is shown in Figure 4. We choose the boundary values of when to include a (dotted) line such that the number of (dotted) lines is almost equal to figure 1[11]. This will make sure that we are not influenced by the amount of lines in the graph. This graph seems less structured, because there aren't as many clear groups of correlated indices. The relationships

[11] The correlation results have a different distribution for both methods. In Houweling and Zeelenberg (2017) there are more pairs with a relatively high correlation than in the current study. We have tried different ways to visualise this, but adding extra correlations to figure 3 mainly increases its complexity and only strengthens the point of view that it has less structure.

that stand out are not always the ones where we would expect the highest correlation, e.g. between wages and 3-months interest rates.

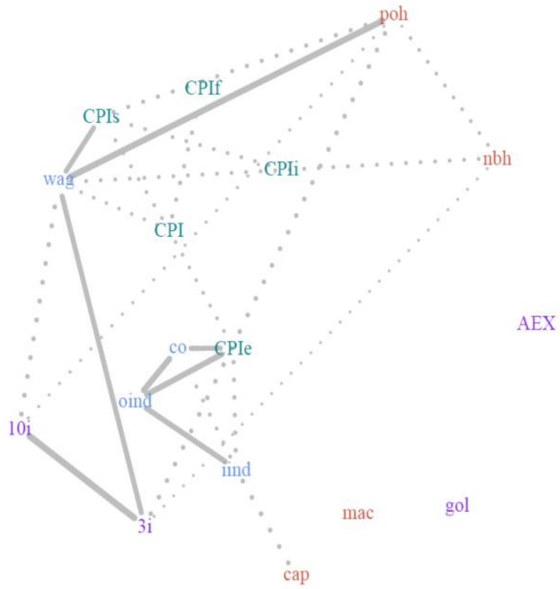

Figure 4 Correlation based on Houweling and Zeelenberg (2017). Only correlations of 0.59 and higher are shown. Correlations between 0.59 and 0.73 are dotted. Abbreviations are explained in Table 10.

We see that especially for the interest rate series there is a large difference with the current study. In figure 1 the two interest rates indices are not related to any other point, whereas in figure 4 they both have three relationships. We expect that this is due to the different data representation in both analyses. The interest rate series are a bit different from the other series in the price dashboard because the other series are all index series, whereas the interest series are not. For interest rates we chose in both studies to include them in their primary form. However, the other series were not the same in both studies. In Houweling and Zeelenberg (2017) the seasonally adjusted, smoothed month-month changes were used and in this research we used the index series.

In general we see that in Houweling and Zeelenberg (2017) there were a lot of pairs of indicators with a high correlation that we could not always explain. In this analysis we mainly see one group of highly correlated indicators that we can explain (crude oil import price, CPI energy and import and output price industrial products). This could be an indication that the method used in this research is less likely to find spurious relationships between indices than the method in Houweling and Zeelenberg (2017).

We see the method used in this research as a more fundamental method, because it looks at correlation in the trend of the series. The results of the current analysis is also more in line with what we would expect. In Houweling and Zeelenberg (2017) there were quite some high correlation results that we would not expect and could not explain based on economic theory. In this analysis we mainly see one group of highly correlated indicators that we can explain (crude oil import price, CPI energy and import and output price industrial products). This could be an indication that the correlation analysis in Houweling and Zeelenberg (2017) is more likely to give false positive signals of correlation (spurious relations) than the current analysis. All in all we would suggest to value the current results higher than the results in Houweling and Zeelenberg (2017).

4.5 Conclusion

We have investigated the degree of co-movement between indicators within the price dashboard using correlation between the levels of series in structural time series models. The results are visualised in one graphic, showing the strength of all relations that were found.

We have found no true co-integration between any pair of series. The highest correlation was found between the group of import price of crude oil, CPI energy and import and output price of industrial products. The correlation in level using the structural time series seems to give more plausible results than the correlation in Houweling and Zeelenberg (2017), because there are fewer unexplainable correlations found. One topic for further research could be how much this is due to a different data format; i.e. analysing index series instead of relative differences. A remarkable result is that for most pairs of indices that were investigated the lag with the highest correlation was zero. We would expect to find some indicators that are leading the CPI with a few months.

The indices that are barely related to other indices in the price dashboard are AEX share price index and house prices for privately owned and new houses. Indices that are weakly related to other indices are gold price and interest rates for three months and ten year loans.

References

Artis, M. J., Bladen-Hovell R. C., Osborn, D. R., Smith, G and Zhang, W. (1995). Predicting Turning Points in the UK Inflation Cycle, The Economic Journal, Vol. 105, No. 432 (Sep., 1995), pp. 1145-1164

Kazemier, B., Zeelenberg, C. and Walschots, J. (2017) - Monitoring inflation, internal report/discussion paper nr…., Statistics Netherlands,

Commandeur, J.F. and Koopman, S.J. (2007), An introduction to state space time series analysis, Oxford University Press.

Durbin, J. and Koopman, S. J. (2012), Time series analysis by state space methods, Oxford University Press.

ECB (2011), The Monetary Policy of the ECB, 3rd edition, European Central Bank, Frankfurt am Main.

Harvey, A.C.(1989), Forecasting, structural time series models and the Kalman filter, Cambridge university press.

Houweling, S. and Zeelenberg, C.(2017), Relationships between inflation indicators, discussion paper, Statistics Netherlands,

Koopman, S.J., Harvey, A.C., Doornik, .J.A., Shephard, N. (2009). STAMP 8.2: Structural Time Series Analyser, Modeler, and Predictor, Timberlake Consultants, London.

Koopman, S.J. and Lee, K.M. (2009), Seasonality with Trend and Cycle Interactions in Unobserved Components Models, Tinbergen institute discussion paper.

Nyblom, J. and Harvey, A.C. (2001), Testing against Smooth Stochastic Trends. Journal of Applied Econometrics, Vol. 16, No. 3, Special Issue in Memory of John Denis Sargan, 1924-1996: Studies in Empirical Macroeconometrics (May - Jun., 2001), pp. 415-429.

Seitz (1998) - An Index of Leading Indicators on Inflationary Trends and the Business Cycle

van den Brakel, J, Söhler, E., Daas, P., Buelens, B. (2016), Social media as a data source for official statistics; the Dutch Consumer Confidence Index

Appendix

Correlation table: In the top right the results from this study; correlation between the levels of series for a multivariate model with 17 series. In the bottom left the results from Houweling and Zeelenberg (2017); optimal correlation with between brackets the optimal lag.

	CPI	CPI energy	CPI ind. goods	CPI food	CPI services	AEX	3-months interest	Gold price	10-year interest
CPI	1 (0)	0.63	0.48	0.40	0.59	0.04	0.04	0.06	0.17
CPI energy	0.59 (-3)	1 (0)	0.06	-0.09	0.04	0.08	0.18	0.08	0.24
CPI industrial goods	0.53 (1)	0.26 (-12)	1 (0)	0.20	0.09	-0.06	-0.06	-0.11	0.04
CPI food	0.66 (3)	0.55 (9)	0.39 (-4)	1 (0)	0.14	-0.21	-0.03	0.02	-0.05
CPI services	0.61 (16)	0.32 (20)	0.61 (13)	0.43 (15)	1 (0)	-0.07	-0.16	0.02	0.00
AEX	-0.41 (9)	-0.32 (-13)	-0.33 (20)	-0.33 (2)	-0.49 (-1)	1 (0)	0.11	0.00	0.21
3-months interest	0.54 (4)	0.49 (11)	0.21 (20)	0.43 (-1)	0.54 (-14)	-0.29 (-3)	1 (0)	-0.02	0.26
Gold price	0.18 (-9)	0.28 (4)	-0.31 (4)	-0.21 (20)	-0.3 (10)	0.11 (8)	0.18 (0)	1 (0)	-0.12
10-year interest	0.48 (11)	0.42 (16)	0.28 (20)	0.26 (12)	0.45 (-6)	-0.2 (-17)	0.88 (0)	0.18 (-19)	1 (0)
Import industrial products	0.38 (-1)	0.67 (2)	-0.28 (4)	-0.4 (9)	0.17 (-20)	0.45 (4)	0.19 (-12)	0.3 (3)	0.17 (-14)
Wages	0.59 (12)	0.26 (-5)	0.55 (12)	0.32 (8)	0.74 (-6)	-0.4 (-2)	0.76 (3)	-0.2 (13)	0.71 (0)
Crude oil price	0.36 (-16)	0.73 (-2)	0.22 (-19)	-0.47 (2)	0.15 (11)	-0.36 (11)	0.33 (-15)	0.36 (-8)	0.28 (20)
Output price industry	0.45 (-3)	0.81 (0)	-0.17 (2)	-0.48 (6)	0.15 (-20)	0.34 (2)	0.38 (-12)	0.37 (-6)	0.27 (-16)
Privately owned houses	0.5 (-17)	0.24 (-13)	0.53 (-11)	0.32 (-18)	0.65 (-20)	-0.29 (-18)	0.7 (-16)	-0.45 (20)	0.63 (-20)
New build houses	0.47 (-2)	0.18 (-8)	0.41 (13)	0.42 (-2)	0.56 (-17)	-0.32 (-12)	0.59 (-9)	0.15 (-16)	0.56 (-12)
Capital goods	-0.35 (13)	0.43 (3)	-0.24 (6)	-0.42 (10)	-0.33 (13)	0.35 (9)	0.23 (-4)	0.47 (3)	0.16 (-15)
Import price machinery	-0.34 (15)	-0.25 (-3)	-0.29 (16)	-0.22 (15)	-0.46 (14)	0.18 (-6)	-0.34 (-9)	0.26 (3)	-0.38 (-20)

	Import industrial products	Wages	Crude oil price	Output price industry	Privately owned houses	New build houses	Capital goods producer price	Import price machinery
CPI	0.27	0.36	0.35	0.37	0.05	0.20	0.11	-0.11
CPI energy	0.38	0.09	0.57	0.53	0.10	0.02	0.26	-0.09
CPI industrial goods	0.01	0.18	0.09	0.05	0.02	0.15	0.04	-0.17
CPI food	-0.08	0.14	-0.14	-0.16	0.07	0.06	-0.12	-0.02
CPI services	-0.01	0.44	0.05	0.02	-0.04	0.23	-0.10	-0.02
AEX	0.16	-0.12	0.17	0.19	-0.00	-0.10	0.02	0.00
3-months interest	0.30	-0.09	0.17	0.38	0.22	0.08	0.16	-0.05
Gold price	0.27	-0.06	0.16	0.12	0.00	0.03	0.09	0.35
10-year interest	0.22	0.04	0.26	0.31	0.12	0.05	0.05	0.05
Import industrial products	1 (0)	-0.05	0.54	0.75	0.02	-0.04	0.30	0.40
Wages	-0.21 (-10)	1 (0)	0.06	0.01	0.20	0.23	-0.13	-0.21
Crude oil price	0.62 (-4)	0.24 (20)	1 (0)	0.75	0.03	-0.00	0.05	0.06
Output price industry	0.78 (-1)	0.18 (20)	0.83 (2)	1 (0)	0.03	0.01	0.23	-0.02
Privately owned houses	-0.27 (-11)	0.85 (-18)	-0.22 (20)	-0.28 (17)	1 (0)	0.25	0.05	-0.11
New build houses	-0.35 (-15)	0.62 (-7)	0.16 (-20)	0.15 (10)	0.6 (7)	1 (0)	0.16	-0.06
Capital goods	0.71 (1)	-0.2 (15)	0.51 (7)	0.56 (4)	0.22 (3)	0.2 (-2)	1 (0)	0.41
Import price machinery	-0.34 (-10)	-0.4 (20)	-0.32 (0)	-0.33 (-1)	-0.34 (-4)	-0.41 (-9)	0.53 (2)	1 (0)

Explantion of symbols

Empty cell	Figure not applicable
.	Figure is unknown, insufficiently reliable or confidential
*	Provisional figure
**	Revised provisional figure
2017–2018	2017 to 2018 inclusive
2017/2018	Average for 2017 to 2018 inclusive
2017/’18	Crop year, financial year, school year, etc., beginning in 2017 and ending in 2018
2015/’16–2017/’18	Crop year, financial year, etc., 2015/’16 to 2017/’18 inclusive

Due to rounding, some totals may not correspond to the sum of the separate figures.

Colophon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Statistics Netherlands, CCN Creation and visualisation

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contactform: www.cbsl.nl/information