A monthly indicator of private fixed capital formation; a model based approach

Floris van Ruth

Discussion paper (201020)

Statistics Netherlands

The Hague/Heerlen, 2010

Explanation of symbols

.	= data not available
*	= provisional figure
**	= revised provisional figure
x	= publication prohibited (confidential figure)
–	= nil or less than half of unit concerned
–	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2008–2009	= 2008 to 2009 inclusive
2008/2009	= average of 2008 up to and including 2009
2008/'09	= crop year, financial year, school year etc. beginning in 2008 and ending in 2009
2006/'07–2008/'09	= crop year, financial year, etc. 2006/'07 to 2008/'09 inclusive

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

Publisher
Statistics Netherlands
Henri Faasdreef 312
2492 JP The Hague

Prepress
Statistics Netherlands - Grafimedia

Cover
TelDesign, Rotterdam

Information
Telephone +31 88 570 70 70
Telefax +31 70 337 59 94
Via contact form: www.cbs.nl/information

Where to order
E-mail: verkoop@cbs.nl
Telefax +31 45 570 62 68

Internet
www.cbs.nl

ISSN: 1572-0314

600830201020 X-10

A monthly indicator of private fixed capital formation; a model based approach

Floris van Ruth

Summary: Having to produce accurate high-frequency economic statistics with limited data available is a common problem for statistical institutions. This study shows how state space based approaches to high-frequency interpolation can provide a solution. It is described how a monthly index of the development of private fixed capital formation can be computed, using monthly data on production and imports of capital goods. Two different methods are tested, an interpolation formulation and a cumulative formulation. Both approaches produce good monthly indicators, but the interpolation approach is to be preferred. An additional advantage is that this approach can be easily automated, resulting in a very fast and efficient production process.

Keywords: State Space models, Kalman filter, temporal disaggregation, high frequency interpolation, benchmarking, business cycle indicators, fixed capital formation, quarterly National Accounts

1. Introduction

This study describes two methods for producing a monthly indicator, a new statistic, using econometric methods and based on existing monthly data sources. The results in this study are interesting for two reasons. Firstly, the new statistic concerned is a monthly indicator for the development of private fixed capital formation, a very useful addition to the range of short-term economic statistics already produced by Statistics Netherlands. But also because it is shown how econometric methods can be used to produce new statistics efficiently, using existing data more fully, and thus reducing the need for (new) surveys. As more and more official and private sector data sources become available, the need for and utility of techniques like these can only increase. An enhancing factor is the ever increasing desire for faster information, which means not only the quick publication of statistics, but preferably at a high frequency as well. At the same time however, there is pressure on statistical agencies to reduce the administrative burden on the private sector.

Econometric techniques can be of great use here for several reasons. Production processes based on these techniques are readily automated, and therefore fast and efficient. Also, by their very nature econometric techniques can, when correctly applied, make the most of imperfect source data or optimize existing estimation processes. The interest in and use of econometric techniques in the production of high-frequency National Accounts statistics has been increasing [Liu and Hall (2001), Quilis (2005), Proietti and Moauro (2005)]. This study concerns itself with the usefulness of state space based methods for the production of high-frequency statistics. The state space framework is a very versatile method for describing and estimating the dynamics of economic and other variables. It makes it possible to estimate diverse and otherwise difficult to evaluate models. The application of state space based methods in statistics production is increasing. It is used for estimating unobserved monthly regional production indicators using observed regional annual indicators and related monthly indicators [Clar et al. (1998) and Israelevich and Kuttner (1993)], for repairing breaks in time series [Van de Brakel and Roels (2009)], improving the quality of existing statistics [Van de Brakel and Krieg (2009)] and backcasting and reparation of historical time series.

One of the strengths of the state space approach is the ability to extract unobserved processes or variables, like producing a new monthly statistic. This study is part of series of studies [Van Ruth (2006a,b)] into the application of state space methods for the production of high-frequency statistics, more specific a monthly indicator of fixed capital formation. One of the earlier studies [Van Ruth 2006b] showed that it is possible to use state space models to extract a credible monthly indicator from relevant monthly indicators. The question was whether it is possible with these techniques to construct monthly or quarterly statistics without using a formal statistical framework, and whether this can be done with acceptable accuracy. It showed that it is possible to derive a credible new statistic from existing data sources not specifically tailored to producing the desired statistic. It used an interpolation approach, where the desired high frequency indicators are defined as missing values between the observed (quarterly) realisations. Therefore, the high frequency indicator is unobserved, but assumed to exist. In this case, the observed variable is the quarterly growth rate of private fixed capital formation and the unobserved variables are the corresponding monthly growth rates. Quite a number of techniques have been developed for high frequency interpolation [Di Fonzo (2004)]. A number of European statistical institutes use one form or another of these techniques to estimate certain quarterly components of GDP using annual data and a form of state space modelling [Di Fonzo (2004), Proietti (2004, 2005)]. This can be a practical and efficient method for producing statistics for which it would be too costly or difficult to obtain source data at the relevant (monthly, quarterly) frequency. These techniques were initially developed for the retrospective deconstruction of low-frequency statistics into higher frequency ones. To use these techniques in the actual production of a high frequency statistic is difficult, as one is continually missing the last low-frequency observation. Using state space based models can compensate for this as these were constructed for making optimal predictions. This study presents an update of the interpolation approach for estimating the monthly indicator, and compares it to the outcomes of another variant, a cumulative state space formulation.

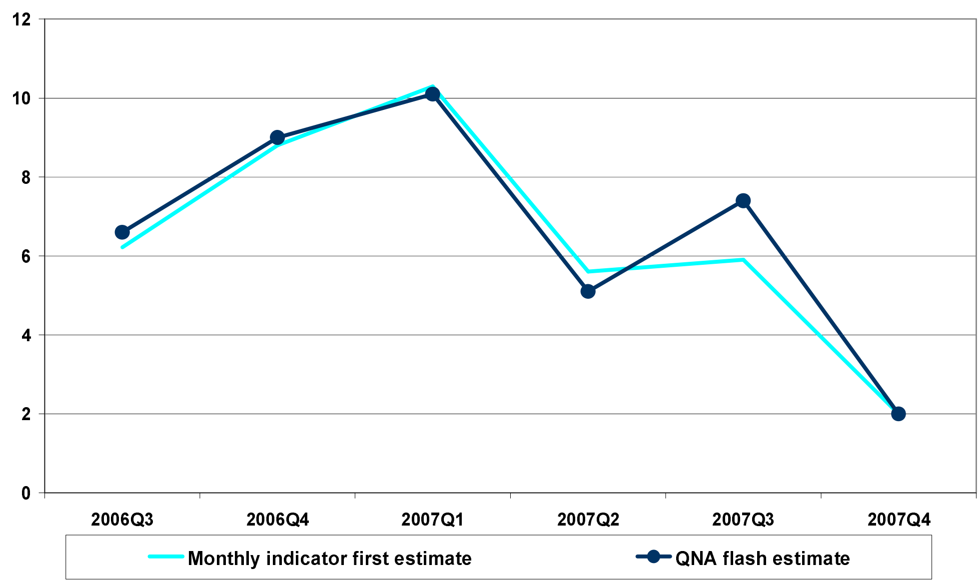
Fixed capital formation is one of the key macro-economic indicators, as capital expenditure plays a central role in business cycle dynamics, and the rate of capital formation is a key determinant of structural economic growth. Therefore, statistics on capital formation are not only of interest for financial markets but for policy makers as well. This study focuses on *private* fixed capital formation, as this has the closest link with current economic conditions. Traditionally, the highest frequency at which Statistics Netherlands publishes data on capital formation is quarterly. It is part of the quarterly National Accounts, of which the methodology is based on the annual system of National Accounts. There is no independent survey on fixed capital formation. It is a composite statistic, whose development is derived from other quantities in a system of supply and use tables. Replicating this methodology on a monthly basis might be difficult and time-consuming, if possible at all. In this study one of the alternatives is evaluated; using econometric techniques to produce high-frequency data from related indicators.

Using the econometric techniques described, the unobserved growth rates are estimated by using selected related monthly indicators. These are mainly components of industrial production and imports, which are the main sources of fixed capital. The lower frequency (quarterly) statistic is used as reference, and related monthly indicators are used to construct a monthly estimate of the target indicator. This is basically an indicator approach. This means that it is not attempted to measure the level of capital formation directly. Instead, indicators for the most important components of fixed capital formation are sought, and from their development the development of the target variable, fixed capital formation, is derived. Thus, monthly fixed capital formation is a kind of latent variable; not directly observed, but its development can be derived from the behaviour of related quantities.

2. Some history

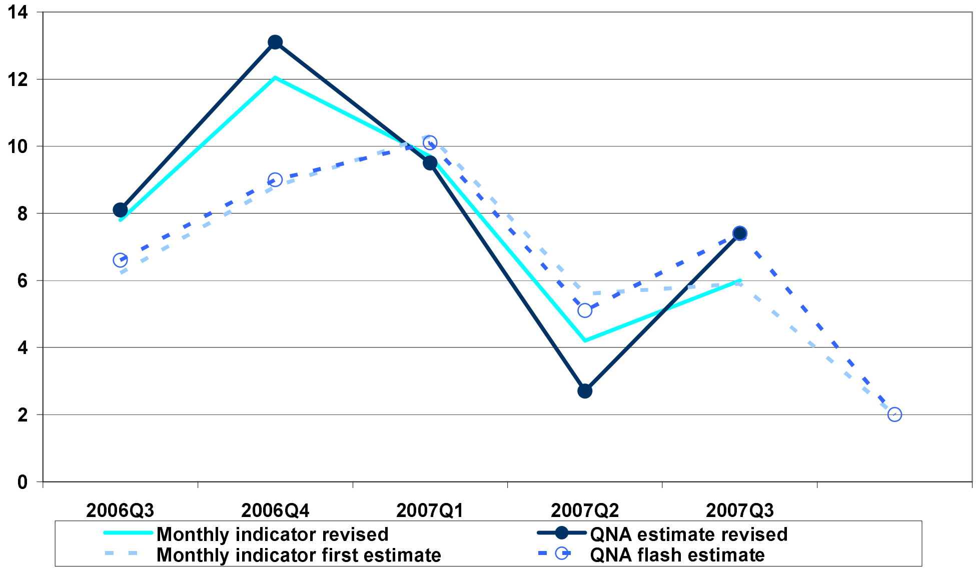
An earlier version of the interpolation method, based on the results form the study [Van Ruth 2006b], was tested over a trial period of six quarters from 2006Q3 to 2007Q4. During this period, each month an estimate of the monthly indicator of the development of private fixed capital formation was computed, based on the data that came available that month. The monthly indicators used were the production of building materials and the imports of IT-equipment. The results are shown in graph 2.1 compared with the Quarterly National Accounts Flash estimate of private fixed capital formation, which is the quarterly reference variable. The Flash estimate is the first, preliminary estimate produce by the Quarterly National Accounts, and the monthly indicator is a precursor to this Flash estimate.

Graph 2.1; Quarterly National Accounts Flash estimate for yoy growth rate of total private fixed capital formation, compared to quarterly average first estimates experimental monthly indicator.



The match between the monthly estimate and the corresponding quarterly flash estimate is surprisingly good, with an average mean absolute error of only 0.56 %-points. This is comparable to the rmse found in the out of sample exercises performed in the development stage. Of course the Flash estimate is only preliminary and subject to substantial revisions as more and more accurate data become available. But as graph 2.2 shows, the monthly estimation method is able to adapt and produce accurate revised estimates as the QNA realisations are revised.

Graph 2.2; Quarterly National Accounts estimates for yoy growth rate of total private fixed capital formation, both flash estimate and revised, compared to quarterly averages of first and revised estimates experimental monthly indicator.

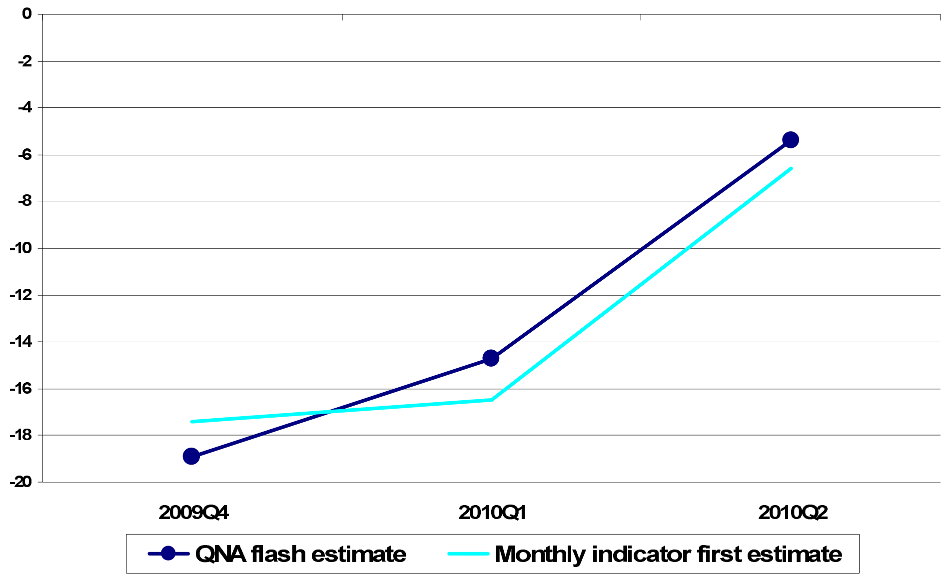


These actual production results show that producing statistics using this model based approach is absolutely viable. Results were outstanding and because the computations were automated, production time was negligible. Then a serious problem arose. An important input to the computations were data on imports of capital goods at the 2-digit level, i.e. broken down into categories like electrical equipment, IT-hardware, transport equipment etc. It turned out that the composition of these lower-level import aggregates is unstable from one year to the next. Mostly the changes are small, but sometimes they were significant. This means that the year-on-year growth rates are distorted, resulting in inaccurate estimates. The most practical solution to this problem was using higher level import aggregates, which are less precise but of stable composition. This spurred the research of which the results are reported in this paper.

The new computation module is again based on the interpolation approach described later in this paper. It now uses import data on single digit aggregation level, in this case the imports of machinery and transport equipemtn, with the other indicators mostly unchanged. Test production was resumed in the last quarter of 2009, which means that by now three quarters of actual production results are available. In graph 2.3, the outcomes are compared to the QNA flash estimates. When comparing graph 2.3. and 2.1, it becomes clear that the new system is not as accurate as the old one. The monthly estimates are still a reliable precursor of the QNA-flash estimates, but do not track them as closely as before. This is probably not only due to the less accurate import indicators, but also to the shorter time series now available (from

2000 Q1 instead of 1995 Q1). The mean average deviation is now 1.5 percentage points, which is acceptable for a rapid indicator.

Graph 2.3; Quarterly National Accounts Flash estimate for yoy growth rate of total private fixed capital formation, compared to quarterly average first estimates ***new version*** *experimental monthly indicator.*



3. Methodology

3.1 The state space framework

The state space framework is a method for reformulating the dynamics of economic and statistical models in a form which allows for analyses to be performed which would otherwise be very difficult. Used in combination with the Kalman filter, it offers computational efficiency and power. It is mainly used in seasonal adjustment, statistical analysis [Bikker et al. (2005)], business cycle research [Carvalho and Harvey (2003), Stock and Watson (1991), Valle e Azevedo and Koopman (2003), Van Ruth et al. (2005)] and temporal disaggregation of statistics [Clar et al. (1998), Di Fonzo (2003, 2004), Proietti (2003, 2004), Buiten et al. (2006)].

The methods used for temporal disaggregation described in the literature generally use two components to estimate the target variable: some form of autoregressive process and a process introducing exogenous, so-called related indicators. Both types of components are easily incorporated in the state space approach. Basically, in the state space framework there are two types of variables; observed indicators which are linked to the second type, the unobserved state variables which determine most of the dynamics of the system. In its most basic form a state space model can be described by the following equations [Harvey (1989)]:

$$
\begin{aligned}
y_t &= Z_t * \alpha_t + S_t * \xi_t \\
\alpha_t &= T_t * \alpha_{t-1} + R_t * \eta_t
\end{aligned}
\qquad (1)
$$

The first equation is the so-called signal or measurement equation, which describes how the directly observed variables y_t are related to the unobserved variables α_t, the so-called state variables. By elaborating on this basic structure, virtually all types of dynamics can be modelled. The state vectors α_t can be both scalar- and vector valued. The dynamics of the state vector are described by the second equation, the transition equation. Apart from autoregressive components, state space models can also incorporate information from exogenous variables, or related indicators. Basically, these can be entered in two ways; a straightforward approach is to introduce an additional term in the measurement equation:

$$y_t = Z_t * \alpha_t + G_t * x_t + S_t * \xi_t$$

$$\alpha_t = T_t * \alpha_{t-1} + R_t * \eta_t \tag{2}$$

Where x_t is a vector of exogenous variables, containing the relevant values at t, and G_t is a matrix of, usually fixed, coefficients. The other method is to incorporate the exogenous indicators in the state equation:

$$y_t = Z_t * \alpha_t + S_t * \xi_t$$

$$\alpha_t = T_t * \alpha_{t-1} + G_t * x_t + R_t * \eta_t \tag{3}$$

In the basic equation (1), Z_t and S_t, are constant coefficient vectors and T_t, and R_t are constant coefficient matrices. The disturbances ξ_t and η_t have mean zero and covariance matrix H_t and Q_t respectively, and are serially uncorrelated. The system parameters, or hyperparameters, Z_t, S_t, T_t, R_t, H_t and Q_t are unknown and need to be determined. For this, a powerful technique known as the Kalman Filter is used. This technique greatly simplifies estimating these models and facilitates the determination of the unobserved state vectors. A short discussion of the Kalman Filter can be found in Appendix I. The force of the state space approach lies here both in the ability to estimate the unobserved monthly values and to produce optimal estimates for the periods at the end of the sample where the quarterly realisation is not yet available.

3.2 The interpolation method in state space form

The first method considered in this study, and the one implemented in producing the monthly private fixed capital formation indicator, is a true interpolation approach for producing high-frequency statistics. In the approach used here, the signal variable y_t then is formed by alternating observed quarterly realisations of the target variable with unobserved monthly values.

$$y = \left\{ .., I_t^{Q,m1}, I_t^{Q,m2}, I_t^{Q}, I_t^{Q,m3}, I_{t+1}^{Q,m1}, I_{t+1}^{Q,m2}, I_{t+1}^{Q}, I_{t+1}^{Q,m3}, .. \right\} \tag{4}$$

Where: $I^{Q,m1}_t$ = the (unobserved) value of I in the first month of quarter t

I^{Q}_t = the (quarterly) realisation of I in quarter t

The unobserved monthly values are treated as missing values, which can be found by using related indicators and the state space formulation to interpolate the observed quarterly realisations. For variables in the form of year-on-year growth rates, the standard form in this study, this set-up works by assuming that the low frequency (quarterly) realisations are the average of the monthly values, and therefore lie in between them, both in time and in value. This data structure is of course an approximation, but as will be shown an effective one. The quarterly realisations are interpolated, with the related monthly indicators supplying information on the monthly pattern.

In an earlier study [Van Ruth (2006b)], different formulations of the interpolation model were tested, based on state space forms of some of the most popular benchmarking models. The first method described in the present study is an update of the most successful model resulting from that earlier study. The formulation used is based on the local linear trend model [Harvey (1989)],

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t \\
\mu_t &= \mu_{t-1} + \eta_t + v_t \\
\eta_t &= \eta_{t-1} + \zeta_t
\end{aligned}
\tag{5}
$$

Where y_t is the observed variable and μ_t is the so-called level component and η_t the slope component. These last two form the state variable, and the last two equations are the state transition equations. This very flexible model is able, by changing the properties of the disturbances ε_t, υ_t and ζ_t, to capture the dynamics of many different time series. This is one example of how state space models can incorporate autoregressive components. In the model formulation used here for the interpolation, the exogenous monthly indicators X_t are introduced in the observation equation;

$$
\begin{aligned}
y_t &= \mu_t + G * X_t + \varepsilon_t \\
\mu_t &= \mu_{t-1} + \eta_t + v_t \\
\eta_t &= \eta_{t-1} + \zeta_t
\end{aligned}
\tag{6}
$$

This allows for different dynamic formulations by setting either ε_t, v_t, or ξ_t to zero. Here ε_t, was set to zero by default, meaning that all the disturbances enter via the state equation.

3.3 The cumulative method in state space form

The other model formulation tested here is a more pure form of temporal disaggregation. It is based on the approach suggested by Harvey [Harvey 1989] as described by Proietti [Proietti 2004]. In this study, it is termed the cumulative approach, as the low frequency reference variable is seen as the accumulation of the monthly indicator to be estimate. In this specific case, concerning growth rates, the quarterly realisation is defined as the average of three monthly observations:

$$I_t^Q = \left(I_t^{m3} + I_{t-1}^{m2} + I_{t-2}^{m1} \Big/ 3 \right) \tag{7}$$

The dataset is structured with a monthly frequency, which means that the quarterly realisation of the year on year growth rate I^Q_t has an observation in the final month t of a quarter, and missing values in the preceding two months of the quarter. I^{m1}_{t-2} is the monthly realisation to be estimated for the first month of the quarter, which is situated at t-2 months. To implement this in an estimation procedure, an observational constraint is necessary. Here, is has the form:

$$I_t^Q = \psi_t * \left(I_t^{m} + I_{t-1}^{m} + I_{t-2}^{m} \Big/ 3 \right) \tag{8}$$

$\Psi_t = 1$ if t=n*3 , with n=1, 2, 3,… and 0 otherwise

The actual model performs the interpolation by combining an autoregressive element with the information contained in related monthly indicators. The state space form of this formulation is:

$$I_t^Q = \psi_t * Z * \alpha_t \tag{9}$$

$$\alpha_t = T * \alpha_{t-1} + G * X_t + e_t \tag{10}$$

$$\alpha_t = \begin{pmatrix} I_t^m \\ I_{t-1}^m \\ I_{t-2}^m \end{pmatrix}, \quad Z = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \quad T = \begin{bmatrix} c & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad X_t = \begin{pmatrix} x_t^1 \\ x_t^2 \end{pmatrix}, \quad G = \begin{bmatrix} \beta^1 & \beta^2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

$$e_t = \begin{pmatrix} \varepsilon_t \\ 0 \\ 0 \end{pmatrix}$$

Where the first equation (9) is the measurement equation, relating to the observed quarterly realisations. The second equation (10) is the state equation describing the evolution of the state vector α_t, which contains the monthly indicator to be estimated. It was chosen to limit the autoregressive element to an AR(1) formulation, captured by the first element c of the transition matrix T. The exogenous related indicators x^i_t are entered via the vector X_t, with the matrix G containing the respective coefficients. The set-up of matrix G assures that the X_t are only related to the first element I^m_t of α_t.

4. Data

This study presents two methods for computing a monthly indicator of the development of private fixed capital formation, the target variable. Fixed capital is defined as consisting of both tangible and intangible assets which are used in the production process, and last for more than a year. The private sector is defined as containing private enterprises, non-profit and not-for-profit organisations and market enterprises run by the state. Furthermore, the desired quantity is the real, or volume development of fixed capital formation, i.e. corrected for price changes. For short-term statistics, the interest lies mainly in the development of the variable, i.e. the growth rate. Actual levels are less relevant for business cycle analysis, and often difficult to measure on a monthly basis with acceptable accuracy. Therefore, the variable of interest is the monthly relative year-on-year change in the volume of private fixed capital formation. All other variables in this study are in the form of relative year-on-year growth rates as well. A welcome side-effect is that this deals with issues of stationarity and scale as well.

As mentioned in the introduction, fixed capital formation mainly takes place via investment in construction, transport equipment, machinery, electronic equipment, and via investments in intangible fixed capital. Tangible fixed capital has to be either produced or imported in the form of goods. Both industrial production and imports are well-covered by detailed monthly statistics. Most goods categories in table 4.1 can be linked to specific industries, for which monthly production-indices are available. Construction can be represented by either using the production-index for the construction materials industry or by using data on the building industry itself.

The production indices are available in the form of volume indices, as desired. Unfortunately, this is not so for the data on imports. It was necessary to construct volume quantities by deflating the data with appropriate indicators of price development. For this, producer price indices (supply) were used, matched as good as possible with the import categories. The dataset used consisted of data from the period 2000 M1-2009 M3. The period 2007M1-2009M3 was used for quasi real-time simulation exercises.

Table 4.1; Composition of private fixed capital formation according to source industry and possible industrial production and import indicators.

Source classification (use from supply-use tables)	*Percentage of total private fixed capital formation*	*Industrial Production indices*	*Import categories*
metal products	1.54%	**Manufacturing of machinery and equipment (IP 24-30)**	**various**
Heavy machinery	7.94%	**Manufacturing of machinery and equipment (IP 24-30)**	**Generators and motors (M71), specialized machinery (M72), Metal working equipment (M73), other heavy machinery (M74)**
Office equipment	5.68%	**Manufacturing of Electrical equipment (IP 26-28)**	**Office equipment (M75), professional equipment (M87)**
Medical, telecom and other electronic equipment	3.07%	**Manufacturing of Electrical equipment (IP 26-28)**	**Communication equipment (M76), electrical equipment (77), professional equipment (87)**
Transport equipment	13.47%	**Manufacturing of transport equipment (IP 29-30)**	**Transport equipment (M78, M79)**
Other industries	3.03%	**various**	**various**
Construction	43.67%	**Manufacturing of building materials (IP 16+23), Turnover Construction industry**	**-**
Property services	1.68%	**-**	**-**
Commercial services	17.47%	**-**	**-**

Several other data related issues have to be faced. A very important one is that as mentioned earier, it was found that the composition of the 2-digit import aggregates was not stable from year to year. Changes could be small or large. It was therefore necessary to resort to less precise, but higher order single digit import aggregates. These are imports of machines and transport equipment (**M7**) and imports of other manufactured goods (**M8**). As most changes constitute shifts between 2-digit aggregates inside a single one-digit aggregate, this would eliminate this problem. Another problem related to trade data is that a significant part of Dutch imports and of industrial production is meant for exports. Far from all of it will end up as fixed capital in the Netherlands. Simply using data on imports and industrial production thus introduces a distortion. And simply subtracting exports from imports is not an option either. From 2002 onwards, monthly data on re-exports are available, allowing the import data to be corrected for capital goods not destined for final use in The Netherlands. This takes two years from the already somewhat limited sample available. Therefore, both corrected and uncorrected import indicators were tested. It was found that the improvement in accuracy gained by the re-exports correction outweighed the data loss. Therefore, the monthly indicators of fixed capital formation will be produced using the corrected import data.

It might seem that it would still be necessary to find a correction for exported industrial production. The combination of the indicator approach and state space models means that this is not a major problem. Imports and industrial production development are not expected to lead directly to fixed capital formation development. Instead, they are indicators from which this development is derived. The estimation procedure will be calibrated on deriving the fixed capital formation data from imperfect indicators, and thus at least partly correct for the distortions introduced. The approach used in this study has the advantages of speed and simplicity, and can make the most of imperfect source statistics. This last advantage of using a state space approach is important for another reason as well. For several components of fixed capital formation no direct or imperfect monthly data of any sort is available, foremost concerning intangible capital formation. This lack of data will be partially compensated by the state space models.

5. Results

The aim of this study was to find a way to produce a credible monthly indicator of fixed capital formation. The relevant results are therefore the monthly series produced and how well these describe the development of fixed capital formation. The derived monthly indicators were scored on two counts. Firstly the credibility of the development of the monthly series itself, i.e. its volatility and the monthly indicators on which it is based. A monthly indicator can be either too volatile or too smooth, this will be elaborated later. Furthermore, a model formulation which is based on monthly indicators with a strong direct link with fixed capital formation is more credible. The second criterion is less qualitative, it is the match between the realisations from the Quarterly National Accounts and the quarterly average of the estimated monthly growth rates:

$$RMSE = \sqrt{\left(\sum_{t=1}^{T} \left(I_t^Q - \left(I_t^{Q,m1} + I_t^{Q,m2} + I_t^{Q,m3} \Big/ 3 \right) \right)^2 \Big/ N \right)}$$

As mentioned earlier, the target fixed capital formation growth rates are here the signal variables y_t . The unobserved monthly and quarterly realisations are defined as relative year-on-year changes. The related monthly indicators are entered in the same form. This overcomes possible stationary problems, and simplifies the model as aggregation of monthly flows is not required. Also, it directly leads to the variable of interest, monthly growth rates for fixed capital formation.

Only the outcomes for the most successful models are reported here. In the studies [Van Ruth (2006a, 2006b)] the performance of other types of formulations can be found. As the aim is to produce a monthly estimate in actual practice, a type of real-time simulation is necessary to assess the performance of the different methods. State space methods tend to produce excellent in-sample fits, which usually cannot be reproduced in real-time. Therefore, a quasi-real time exercise was performed, where monthly estimates were produced for a certain time span by estimating the model, extracting the monthly estimate, lengthening the data set with one month and repeating the exercise. This simulates the actual production of a monthly indicator (ignoring data vintage issues). These simulation results are termed forecasts, and were an important selection criterion.

The results for two variants of fixed capital formation are shown, total private fixed capital formation and total private fixed capital formation excluding investment in intangible assets. The first aggregate is the headline number as reported by QNA, the second aggregate is the one for which Statistics Netherlands plans to publish a monthly indicator. First the monthly indicators derived using the interpolation method will be presented, and then the ones resulting from the cumulative approach. In the final part of this section these two sets of outcomes will be compared to each other and to the realisations from the QNA.

5.1 Results interpolation method

As also described in the earlier study [Van Ruth 2006b], different types and combinations of related indicators were tested. Only the final selected ones will be reported here. Unlike in the earlier studies, it was found that for this time period indirect indicators, like capacity utilization or interest rates, added little or nothing to formulations consisting of materially related indicators. The selected interpolation model for total private fixed capital formation was:

$$\mathbf{I^{Q,m}_t = \mu_t - 1.13 + 0.065{*}PI^{building}_t + 0.037{*}\, PI^{elec}_t + 0.147{*}M^{mt}_t + 0.040{*}M^{other}_t}$$

$$\mu_t = \mu_{t-1} + \eta_t + \nu_t$$

$$\eta_t = \eta_{t-1} + \xi_t$$

$\mathbf{PI^{building}_t}$ = yoy volume growth of building materials industry

$\mathbf{PI^{elec}_t}$ = yoy volume growth of electrical equipment industry

$\mathbf{M^{mt}_t}$ = yoy volume growth imports of machinery and transport equipment

$\mathbf{M^{other}_t}$ = yoy volume growth imports of other manufactured goods

Log likelihood = 51.8,

$\sigma_\nu \approx 0$, $\sigma_\xi = 0.0055$

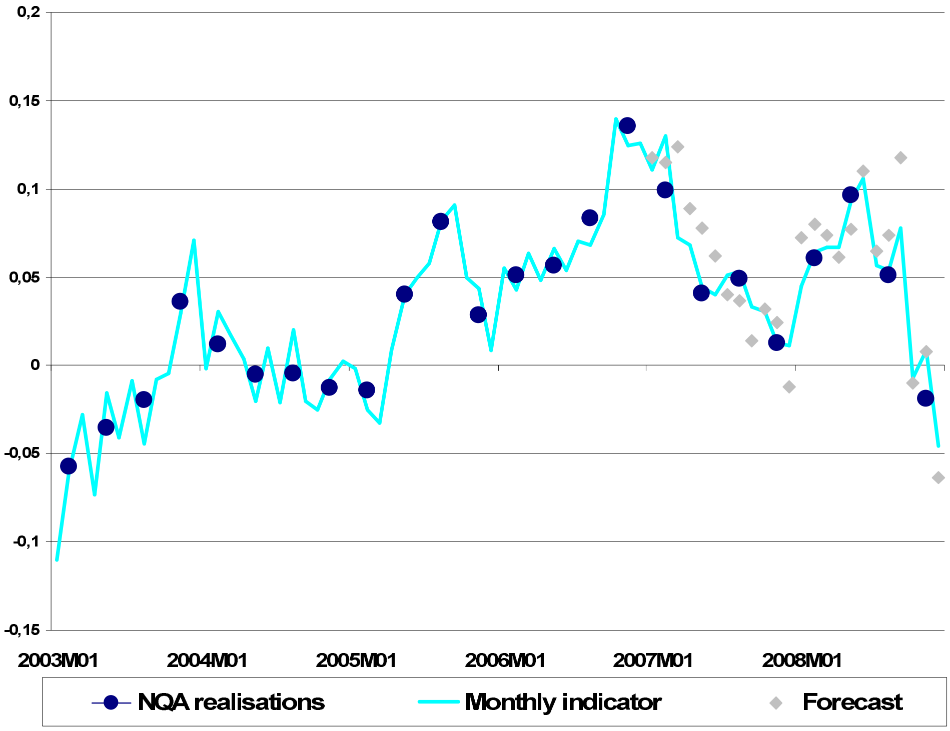
RMSE in-sample = 0.6 %-point

RMSE forecast = 2.1 %-point

Standardised prediction residuals (probabilities in parentheses): Jarque_Bera = 0.726(0.696), Q-stat 4 lags = 6.95(0.139)

The indicators used to derive the monthly values of fixed capital formation represent the most important components of fixed capital, both locally produced and imported: Investment in buildings, production of electrical equipment and imports of all types of machinery and transport equipment. The advantages of this model based approach also become apparent; industrial production is represented by an index, not corrected for exports, whilst imports are measured as a value quantity, though deflated. By formulating the model in growth rates and applying a benchmarking/state space framework, an optimal estimate can be derived from these otherwise difficult to combine source data. The resulting monthly indicator exhibits a credible development, given the development of the quarterly estimate and the observed monthly volatility. A monthly short term-economic indicator can be expected to be more volatile than a quarterly one. Residual diagnostics are in order as well. Unfortunately the out-of-sample estimates are not as accurate as the in sample results, see graph 5.1.

Graph 5.1; Total private fixed capital formation (volume, year on year growth rate) Quarterly National Accounts realisations compared to monthly estimates resulting from interpolation method. Forecast from quasi-real time rolling regression exercise, in-sample results are smoothed estimates.



An additional model was constructed for private fixed capital formation excluding investment in intangible fixed assets. This is a highly volatile, difficult to measure quantity, and therefore the expectation is that excluding this aggregate will make estimation easier and more accurate. The optimal indicator set differs somewhat from the one used to estimate total private fixed capital formation:

$$\mathbf{I^{Q,m}_t = \mu_t -2.146+0.23*Construction_t+0.10* PI^{elec}_t+0.15*M^{mt}_t}$$

$$\mathbf{\mu_t = \mu_{t-1} + \nu_t}$$

$\mathbf{Construction_t}$ = yoy turnover growth in the construction industry

$\mathbf{PI^{elec}_t}$ = yoy volume growth of electrical equipment industry

$\mathbf{M^{mt}_t}$ = yoy volume growth imports of machinery and transport equipment

Log likelihood = 67.0

σ_v = 0.012

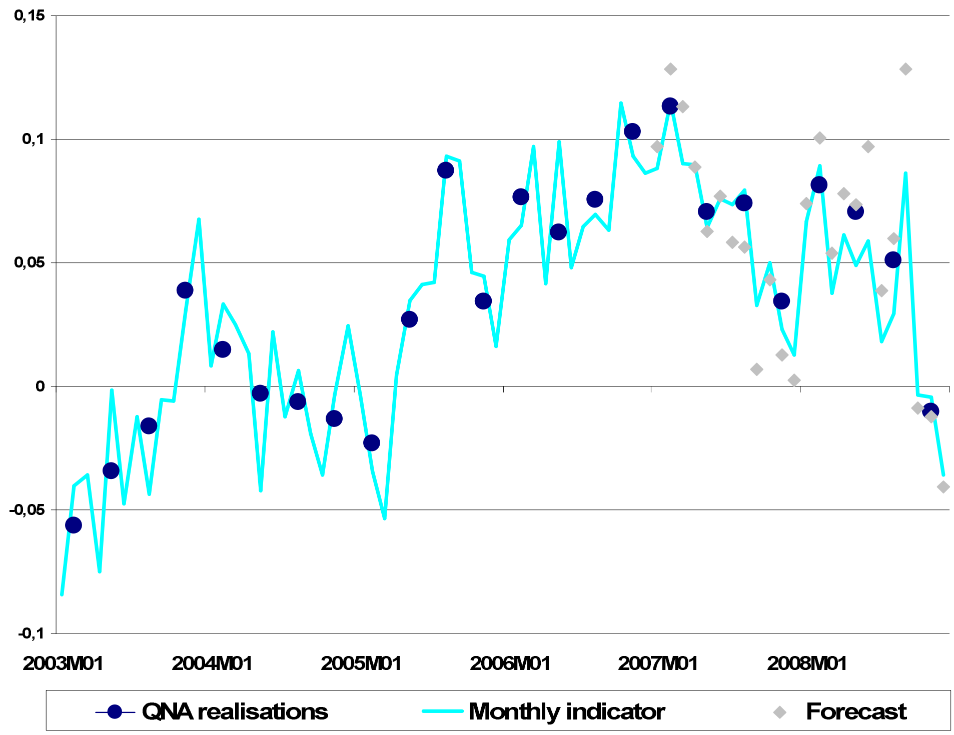
RMSE in-sample = 0.8 %-point

RMSE forecast = 1.6 %-point

Standardised prediction residuals (probabilities in parentheses): Jarque_Bera = 0.474(0.789), Q-stat 4 lags = 3.45(0.485)

The main difference is that instead of the production of building materials, now turnover growth in the construction industry itself is found to be most relevant for estimating the construction component of private fixed capital formation. As expected, omitting intangible fixed assets means that the adjusted monthly aggregate can be measured more accurately. See graph 5.2 for the estimated monthly indicator and the out-of-sample forecast.

Graph 5.2 ; Private fixed capital formation excl. intangible assets (volume, year on year growth rate) Quarterly National Accounts realisations compared to monthly estimates resulting from interpolation method. Forecast from quasi-real time rolling regression exercise, in-sample results are smoothed estimates.



5.2 Results cumulative approach

The interpolation approach reported in the previous section was able to produce credible and reliable monthly indicators. However, it is based on an approximation so it might be preferable to switch to a cumulative approach. The results of testing this are reported here. To facilitate the comparison, the same exogenous indicators are used in the cumulative models as in the interpolation ones. Other compositions were tested as well, but did not perform materially better. The estimated equation of the cumulative formulation for the monthly indicator of the development of total private fixed capital formation is:

$$\mathbf{I^m_t = 0.77*I^m_{t-1} - 0.002 + 0.135*PI^{building}_t + 0.119*\ PI^{elec}_t + 0.002*M^{mt}_t + 0.013*M^{other}_t + \varepsilon_t}$$

$\mathbf{PI^{building}_t}$ = yoy volume growth of building materials industry

$\mathbf{PI^{elec}_t}$ = yoy volume growth of electrical equipment industry

$\mathbf{M^{mt}_t}$ = yoy volume growth imports of machinery and transport equipment

$\mathbf{M^{other}_t}$ = yoy volume growth imports of other manufactured goods

Log likelihood = 54.7

$\sigma_\varepsilon = 0.024$

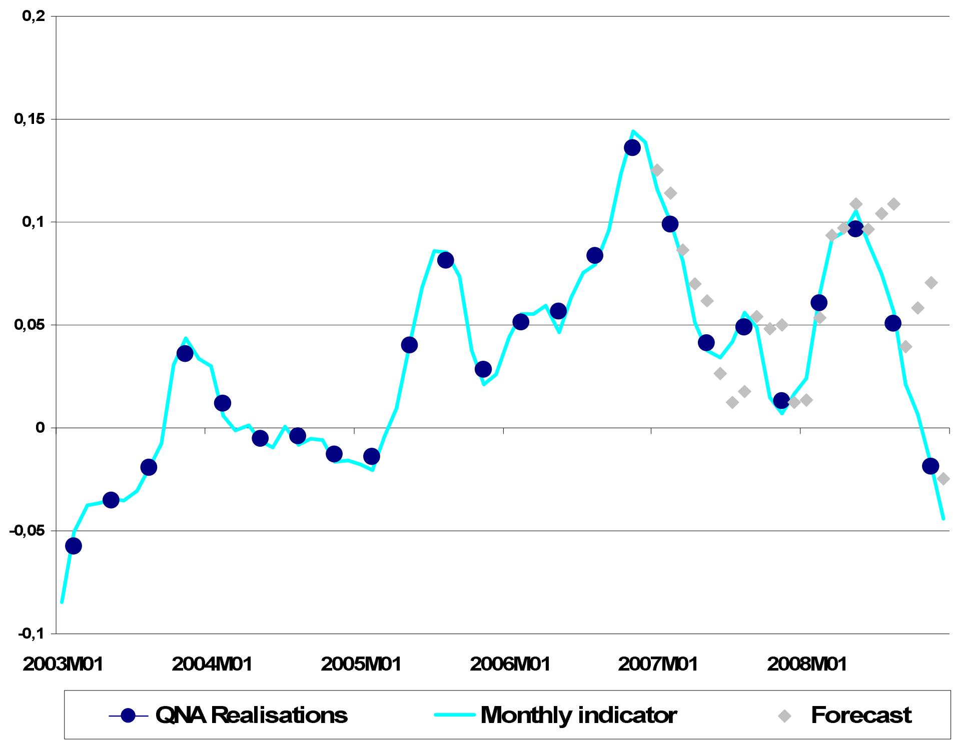
RMSE in-sample ≈ 0 %-point (smoothed)

RMSE forecast = 3.0 %-point

Standardised prediction residuals (probabilities in parentheses): Jarque_Bera = 1.171(0.557), Q-stat 4 lags = 4.88(0.300)

The in-sample fit is perfect, but this is often the case for smoothed estimates of state space models. Therefore, the out-of-sample forecast results become even more important. These are less accurate than those of the interpolation approach. The computed monthly indicator develops more smoothly than in the interpolation case (graph 5.3), which is not necessarily desirable. These two points will be explored further in the next section, where the outcomes of the cumulative and interpolation approach are compared in greater detail.

Graph 5.3; Total private fixed capital formation (volume, year on year growth rate) Quarterly National Accounts realisations compared to monthly estimates resulting from cumulative method. Forecast from quasi-real time rolling regression exercise, in-sample results are smoothed estimates.



The same conclusions hold for the estimated cumulative model for private fixed capital formation excluding intangible fixed assets.

$$\mathbf{I^m{}_t = 0.77*I^m{}_{t-1} - 0.011 + 0.24*Construction_t + 0.164*\ PI^{elec}{}_t + 0.006*M^{mt}{}_t + \varepsilon_t}$$

Construction$_t$ = yoy turnover growth in the construction industry

PI$^{elec}{}_t$ = yoy volume growth of electrical equipment industry

M$^{mt}{}_t$ = yoy volume growth imports of machinery and transport equipment

Log likelihood = 62.7

$\sigma_\varepsilon = 0.018$

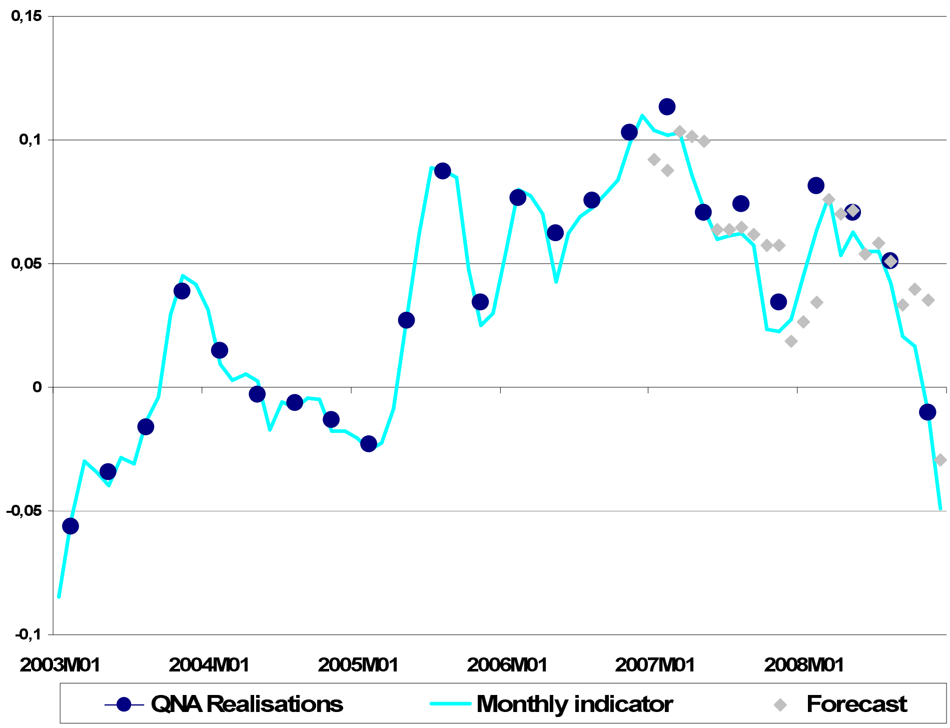
RMSE in-sample ≈ 0 %-point (smoothed)

RMSE forecast = 2.0 %-point

Standardised prediction residuals (probabilities in parentheses): Jarque_Bera = 1.159(0.560), Q-stat 4 lags = 7.98(0.092)

The in-sample estimate of the monthly development of this investment aggregate also develops pleasantly smooth, in itself a welcome property for a business cycle indicator. See graph below.

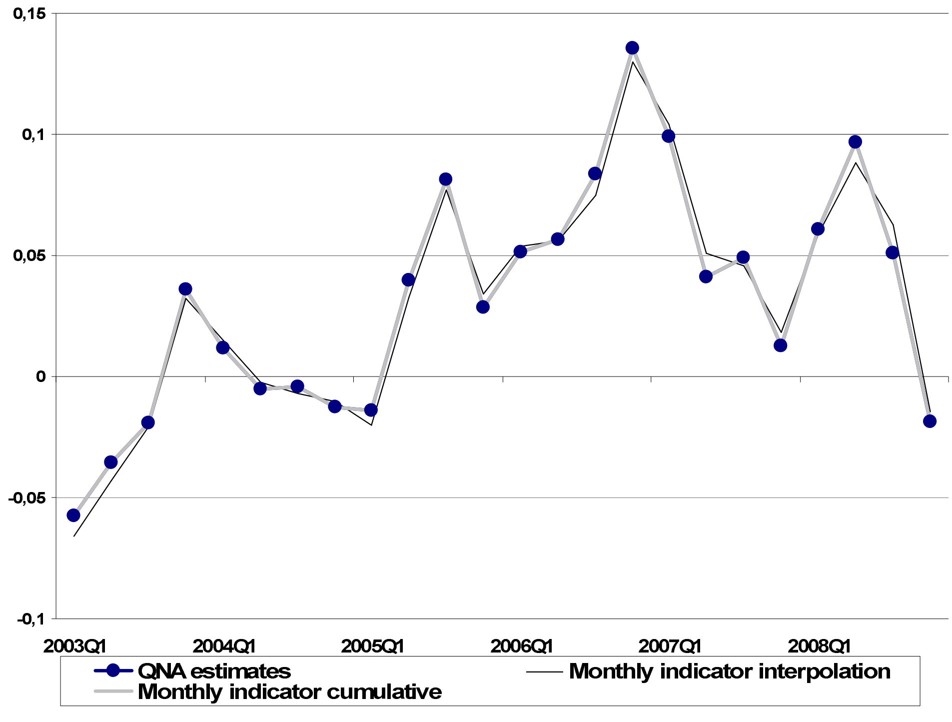
Graph 5.4 ; Private fixed capital formation ex. intangible assets (volume, year on year growth rate) Quarterly National Accounts realisations compared to monthly estimates resulting from cumulative method. Forecast from quasi-real time rolling regression exercise, in-sample results are smoothed estimates.



5.3 Comparison

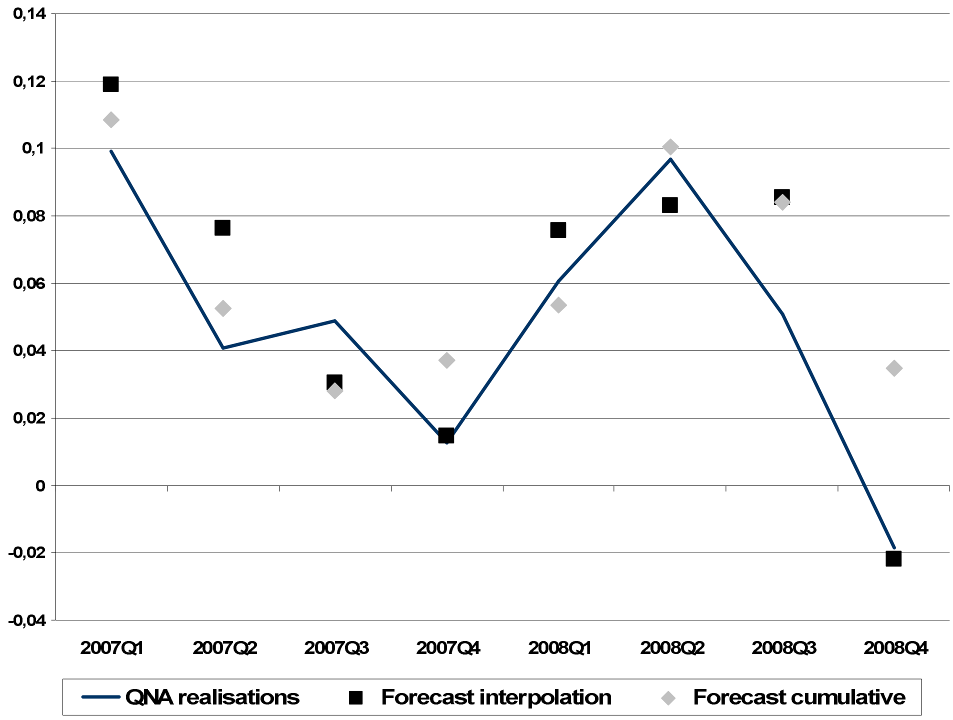
The question which approach is to be preferred is a difficult one. The in-sample fit of the cumulative approach is perfect, but this is somewhat of an artefact of the smoothed state space estimates. As figure 5.5 shows, the quarterly averages of the monthly interpolation approach estimates are quite close to the Quarterly National Accounts realisations.

Graph 5.5; Total private fixed capital formation (volume, year on year growth rate) Quarterly National Accounts realisations compared to quarterly averages of monthly estimates resulting from cumulative and interpolation method. All monthly indicator values are smoothed estimates.



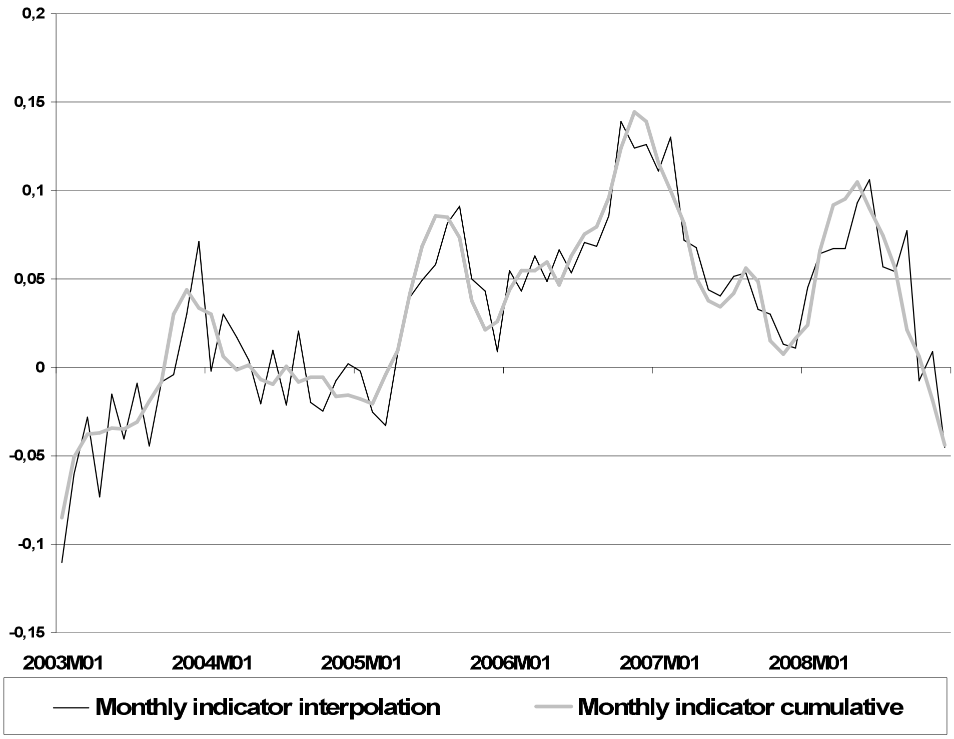
Also, the real-time simulation results from the interpolation approach are closer to the QNA estimates than the cumulative estimates, by about a percentage point, see graph 5.6. This quite an important criterion for choosing a technique which is meant to be used in practice, producing monthly estimates.

Graph 5.6; Total private fixed capital formation (volume, year on year growth rate) Quarterly National Accounts realisations compared to quarterly averages of monthly forecast from quasi-real time rolling regression exercise, both cumulative and interpolation method.



As observed earlier, the cumulative monthly indicator does evolve more smoothly than the interpolation variant, see graph 5.7.

Graph 5.7; Total private fixed capital formation (volume, year on year growth rate) monthly indicator computed via the cumulative approach compared to the monthly indicator computed by the interpolation approach. Both in-sample, smoothed estimates.



The difference in dynamics is mostly relevant on a month-to-month basis, the longer term evolution is quite comparable, as is reflected in the similarity of the quarterly averages. In itself, smoothness is a desirable property in a short-term economic indictor, as it facilitates identifying trends and turning points. It will be argued that here this is not the case. The relevant measure of volatility here is the month-on-month change in the computed and observed growth rates, as this reflects the "noisiness" of the series. In table 5.1 the average absolute month-on-month changes in the computed monthly indicators of fixed capital formation (excluding intangible assets) and those of the monthly indicators used as inputs are shown.

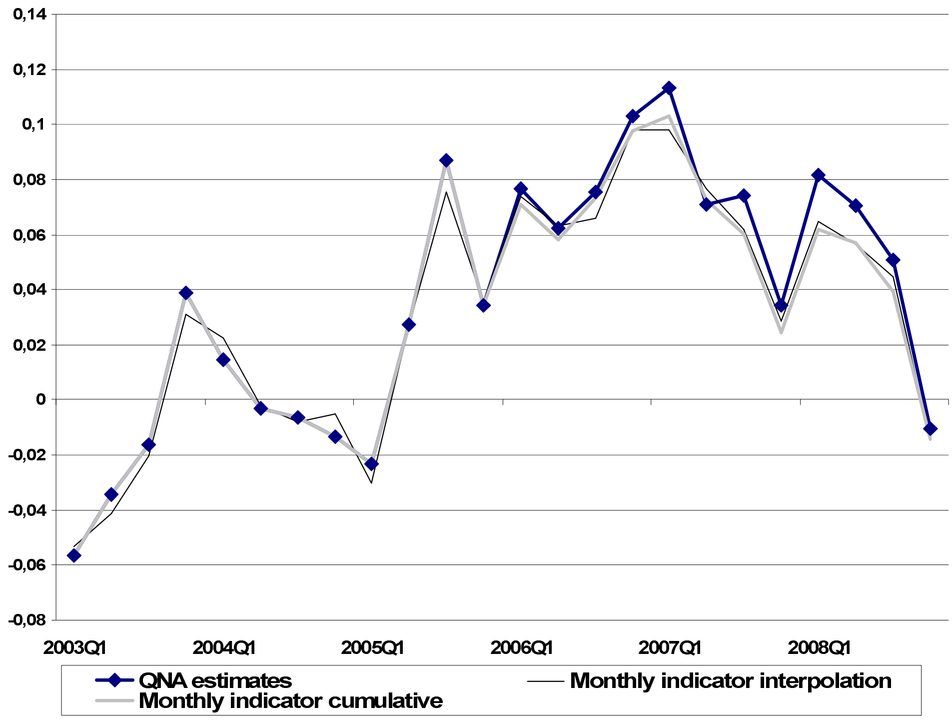
Table 5.1; Average absolute month-on-month changes in the growth rates of monthly indicators (cumulative and interpolation) of private fixed capital formation (ex. Intangible) and monthly input indicators.

	Average absolute month-on-month change
$\mathbf{I^{m}_{cumulative}}$	0.015
$\mathbf{I^{m}_{interpolation}}$	0.031
$\mathbf{PI^{building}}$	0.037
$\mathbf{PI^{electric}}$	0.038
$\mathbf{M^{m}}$	0.067
$\mathbf{M^{other}}$	0.068
Construction	0.054

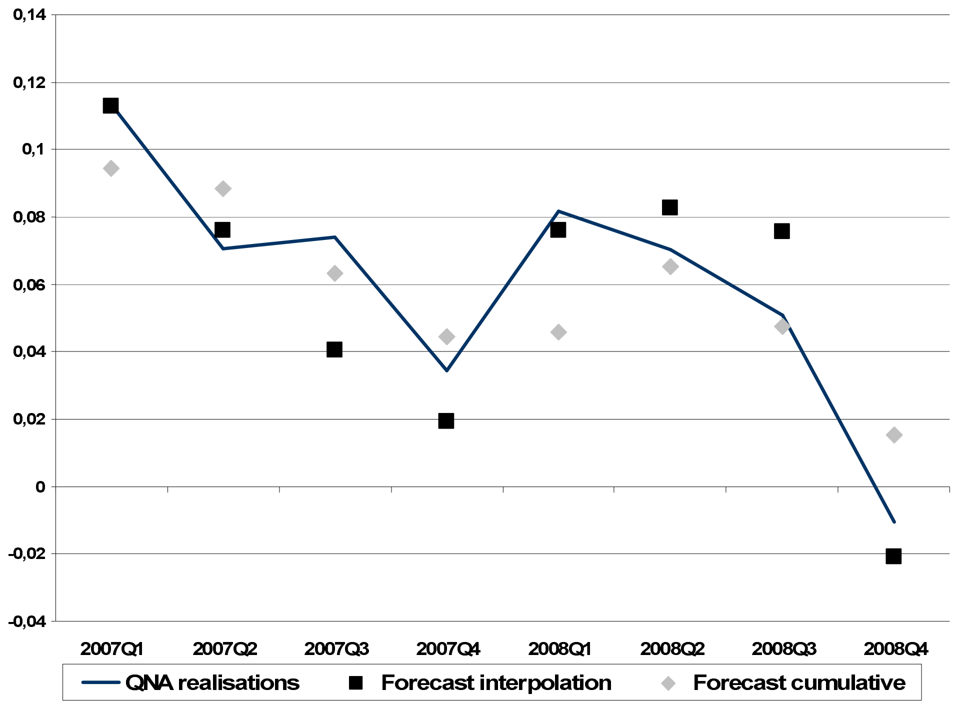
From these data, it seems that if anything both indicators of fixed capital formation are too smooth. The "volatility" of all the input indicators is higher, sometimes much higher. And as these monthly input series are supposed to supply the information on the monthly development pattern, the volatility of the computed fixed capital formation indicators should reflect that of the of the monthly source data. On this ground it is concluded that the interpolation approach is to be preferred, as it seems to yield outcomes which reflect monthly developments more realistically.

The same conclusions hold for the computed monthly indicators for the growth rate of private fixed capital formation excluding intangible fixed assets. On a quarterly basis the fit with QNA realisations is good, both in sample (graph 5.8) and out-of-sample (graph 5.9).

Graph 5.8; Private fixed capital formation ex. intangible assets (volume, year on year growth rate) Quarterly National Accounts realisations compared to quarterly averages of monthly estimates resulting from cumulative and interpolation method.



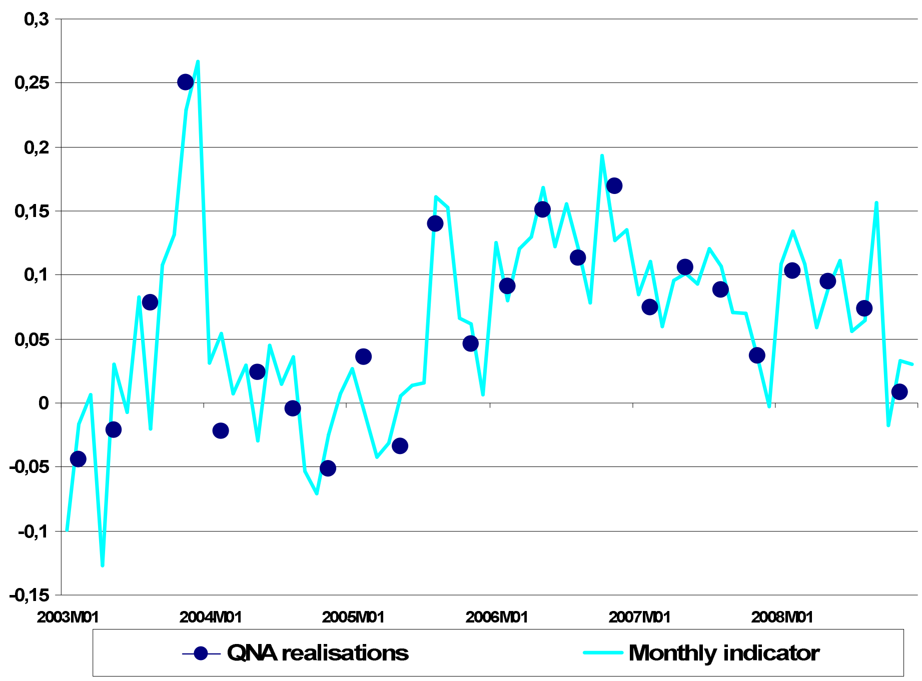
Graph 5.9; Private fixed capital formation ex. intangible assets (volume, year on year growth rate) Quarterly National Accounts realisations compared to quarterly averages of monthly forecast from quasi-real time rolling regression exercise, both cumulative and interpolation method.



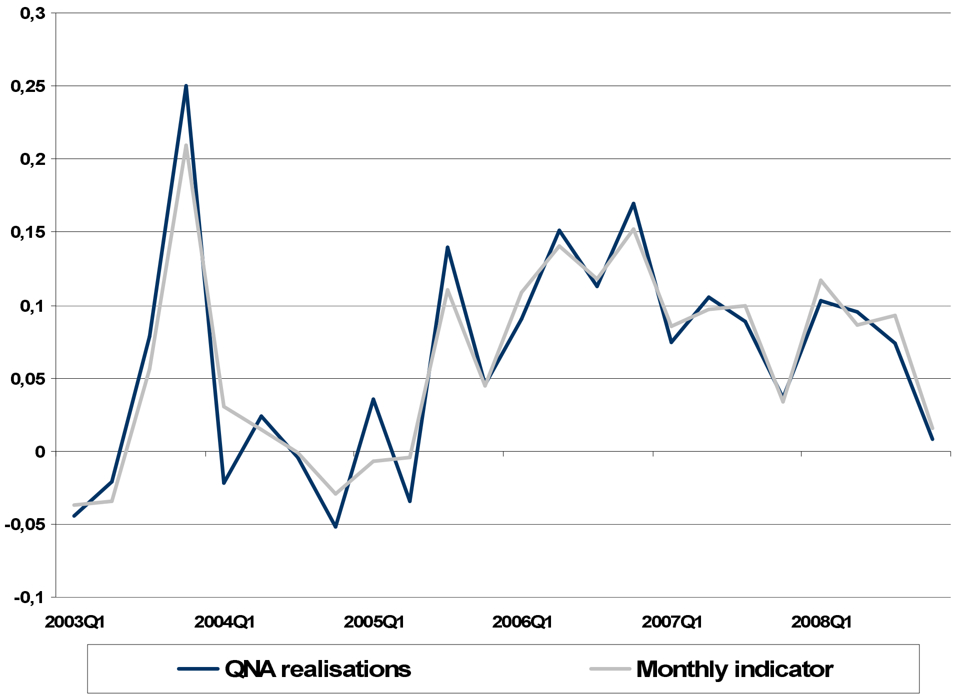
5.4 Indicators for sub-aggregates

Fixed private capital formation is a rather broad concept, ranging from the building of houses and offices via lorries to software. It is therefore potentially useful to have short-term information on the development of more homogenous aggregates. The most important distinction is between real estate construction and other fixed capital formation. This last group is most relevant for the developments in the production capacity, and is therefore termed production-related investment. A monthly indicator was developed for this aggregate, together with three additional indicators for investment in machinery, transport equipment and IT-hardware. The monthly indicators for these last three low-level aggregates proved to be rather unreliable. This is caused both by the inherent volatility of these aggregates and the fact that the import-data at the single digit level are just to aggregated to be useful indicators for these more specific aggregates. The results for the production related-investment are shown in graphs 5.10 and 5.11. Real-time simulations give an rmse of 2.0 percentage-points, but in practice the indicator seems to be somewhat more unreliable than this. One issue is that the QNA-series itself is more unpredictable than the QNA series for total private fixed capital formation.

Graph 5.10; Investment in production related fixed assets (volume, year on year growth rate) Quarterly National Accounts realisations compared to monthly estimates resulting from interpolation method.



Graph 5.11; Investment in production related fixed assets (volume, year on year growth rate) Quarterly National Accounts realisations compared to quarterly averages of monthly estimates resulting from interpolation method.



6. Discussion and conclusions

Short-term information on economic developments is in ever more demand. At the same time, statistical agencies suffer from constraints due to budgetary considerations and complaints from respondents regarding administrative burdens. Under these circumstances, econometric techniques can be of great help, either by costless improving the quality of existing statistics, or by producing entirely new statistics. Both uses are based on maximizing the information contained in data already available, by exploiting coherency and cross-relations. This study concerns itself with the second application: using existing data and model-based techniques to produce a new statistical product. It describes the methodology used to produce the Statistics Netherlands monthly indicator of the volume growth of private fixed capital formation (excluding intangible fixed assets).

This study shows how it is possible to produce a monthly indicator of the development of private fixed capital formation, a key business cycle statistic, using related monthly indicators and high-frequency interpolation techniques. The fundamental idea is to find a way to "interpolate" between the existing Quarterly National Accounts estimates. These serve as reference and as structural backbone of the estimation system. Only for the method to be useful in practice, i.e. for producing a monthly statistic, it must be able not only to interpolate between the quarterly realisations, but also extrapolate for the months in the current quarter. Because of this, it was chosen to use a state space based approach, which is not only a powerful estimation technique in itself, but also uniquely suited to extrapolation. The resulting models were rigorously tested, not only by performing real-time simulations in the development phase, but also in two separate periods of actual monthly trial-production. For private fixed capital formation, the starting point for producing a monthly indicator is actually quite good, as some excellent monthly indicators are available. These are indicators of the production and imports of capital goods, the main components of fixed capital formation. So there is quite a lot of information on a monthly basis on the supply of fixed capital formation. Combined with appropriate modelling techniques, this can be exploited to produce the desired monthly indicator of the volume growth of private fixed capital formation.

Two different estimation approaches were tested, both set in a state space formulation. The first is termed the interpolation approach, where the approximation is used that the quarterly realisations lie between the monthly ones, which are to be estimated. Using the monthly indicators, the missing monthly data are "filled in". The second approach is a cumulative one, in which the monthly estimates are linked to the quarterly realisations by requiring that the average of three monthly estimates is equal to the realisation of the corresponding quarter. Both approaches were tested using the same selection of monthly indicators, i.e. indicators of imports and industrial production, appropriately transformed. Both approaches were able to successfully produce monthly indicators, both in-sample and in production simulating out-of-sample exercises. The derived monthly indicators look credible, with excellent in-sample fits, the root mean square errors being lower than one percentage point. As expected, the out-of-sample rmse's were somewhat higher, around 1.5%-points for the interpolation approach and around 2.5%-points for the cumulative approach. But these are quite acceptable for a rapid monthly economic indicator. The cumulative approach monthly indicator did evolve more smoothly on a monthly basis than the interpolation approach indicator, and it was argued that the interpolation approach gave a more realistic picture of monthly developments. On these two criteria the interpolation approach seems to be preferable to the cumulative approach.

There have already been two periods of successful actual trial production on a monthly basis, yielding actual real-time results. The first period of six quarters used a model formulation based on earlier research and yielded excellent results, with on average less than 0.5%-points between the first monthly estimates and the QNA flash estimate. Using the models described in this study, monthly trial estimates were produced during a period of three quarters, with an average deviation from QNA flash of 1.5%-points. The models used in the first period of trial production made use of more detailed and therefore more precise import data, than those used in the second trial. The interpolation method used in the second trial period and described in this study is also the method chosen for producing Statistics Netherlands monthly indicator of private fixed capital formation. It was unfortunately impossible to keep on using the more detailed import data, as there were issues with the composition of the required import aggregates. This does show that the accuracy of these methods depends for a large part on the availability of good, relevant monthly source data. The method described here is not only quite successful, all things taken into consideration, but also very efficient. The measurement is completely based on existing data, and because econometric methods lend themselves very well to automation, the production process is also very fast and cheap.

Literature

Bikker, R., Hijman, R., Kee, P. and Leeuwen, G., (2005). "Labour force survey data on unemployment: Identifying outliers". Statistics Netherlands discussion paper 05009

Van de Brakel, J and Krieg, S, (2009) "Structured time series modelling of the monthly unemployment rate in a rotating panel" Statistics Netherlands discussion paper 09031

Van de Brakel, J. and Roels, J. (2009) "Intervention analysis with state-space models to estimate discontinuities due to a survey redisgn" Statistics Netherlands discussion paper 09039

Carvalho, V.M. and Harvey, A.C., (2003) "Convergence and cycles in the Euro-zone" Eurostat working paper, theme 1 general statistics.

Clar, M., Ramos, R., Surinach, J. (1998) "A latent variable model to measure regional manufacturing production in Spain" Workshop on Regional Economic Indicators, University of Minho.

Di Fonzo, T. (2003) "Constrained retropolation of high-frequency data using related series: A simple dynamic model approach" Eurostat working paper, theme 1 general statistics

Di Fonzo, T. (2004) "Temporal disaggregation of economic time series: towards a dynamic extension" Eurostat working paper, theme 1 general statistics

Harvey, A.C., (1989): "Forecasting Structural Time Series Models and the Kalman Filter" Cambridge University press, Cambridge

Harvey, A.C., Koopman, S.J. (2000), "Signal extraction and the formulation of unobserved components models", Econometrics Journal 3, 84–107.

Israelevich, P., and Kuttner, K.(1993) "A mixed frequency model of regional output", Journal of Regional studies, 33, p321-342

Liu, H and Hall, S.G. (2001) "Creating high frequency National Accounts with state space modelling: a Monte Carlo experiment" Journal of forecasting vol. 20 (6), 441-449

Proietti, T., (2004) "Temporal disaggregation by state space methods: Dynamic regression methods revisited" Eurostat working paper

Proietti, T and Moauro, F (2005), "Temporal disaggregation and seasonal adjustment" Eurostat working paper

Quilis, E.M. .(2005) "Benchmarking techniques in the Spanish quarterly National Accounts" Eurostat working paper

Stock, J.H., and Watson, M.W. (1991). "A probability model of the coincident economic indicators" in "Leading Economic Indicators", Lahiri, K. and Moore, G. (eds.); Cambridge University Press, New York.

Valle e Azevedo, J. and Koopman, S (2003). "Tracking the business cycle of the Euro area: A multivariate model-based band-pass filter" Eurostat working paper, theme 1 general statistics

Van Ruth, F., Schouten, B., Wekker, R., (2005) “The Statistics Netherlands’ Business Cycle Tracer. Methodological aspects; concept, cycle computation and indicator selection.” Statistics Netherlands report 2005-MIC-44.

Van Ruth, F. (2006a), “A latent variable approach to constructing monthly indicators; fixed capital formation”. Statistics Netherlands Discussion paper 2006.

Van Ruth, F. (2006b), “Constructing a monthly indicator of fixed capital formation via high frequency interpolation. The state space approach”. Statistics Netherlands Discussion paper 2006.

Appendix A; The Kalman Filter

In order to estimate the hyperparameters and latent variables of the state space model described in section X, it is necessary to use the Kalman Filter. This is an iterative procedure based on prediction and efficient incorporation of new information. This construction results in an easy way to evaluate the log-likelihood in each step.

For purposes of convenience, I will first restate the basic state space model, which is described by the following two equations [Harvey]:

$$y_t = Z_t * \alpha_t + S_t * \xi_t$$

$$\alpha_t = T_t * \alpha_{t-1} + R_t * \eta_t \qquad \text{(A1)}$$

The first equation is the signal or measurement equation, which describes how the observed variables y_t are related to the state variables α_t. Z_t and S_t, are constant coefficient vectors and T_t, and R_t are constant coefficient matrices. The disturbances ξ_t and η_t have mean zero and covariance matrix H_t and Q_t respectively, and are serially uncorrelated. The system parameters, or hyperparameters, Z_t, S_t, T_t, R_t, H_t and Q_t are unknown and need to be determined

This is were the Kalman filter comes in. It is in essence a two step procedure, consisting of a prediction step followed by an updating step. In the first step an optimal prediction, *given all available information at that time*, of the signal and state variables in the next period is generated. This yields the prediction $a_{t/t-1}$ of α_t;

$$a_{t/t-1} = T_t * \alpha_{t-1} \qquad \text{(A2)}$$

With error:

$$a_{t/t-1} - \alpha_t = T_t * (a_{t-1} - \alpha_{t-1}) - R\eta_t \qquad \text{(A3)}$$

Which gives for the covariance matrix of the estimation error:

$$E[(a_{t/t-1} - \alpha_t)(a_{t/t-1} - \alpha_t)'] = TP_{t-1}T' + RQR' = P_{t/t-1} \qquad \text{(A4)}$$

The matrix P_{t-1} is important as it is the covariance matrix of the optimal estimate of α_{t-1} and plays a crucial role in the computations. $P_{t/t-1}$ is therefore the covariance matrix of the expected error in the estimate at t-1 of α_t. Given the optimal prediction $a_{t/t-1}$ of α_t, the optimal prediction of y_t at t-1 is:

$$\tilde{y}_{t/t-1} = Z_t a_{t/t-1} \tag{A5}$$

with prediction error:

$$v_t = y_t - \tilde{y}_{t/t-1} = Z_t'(\alpha_t - a_{t/t-1}) + \xi_t \tag{A6}$$

which has variance:

$$\mathrm{var}(v_t) = Z_t P_{t/t-1} Z'_t + H_t = F_t \tag{A7}$$

The prediction error v_t and its variance F_t are crucial factors in the Kalman filter procedure. The prediction errors are used to evaluate the log likelihood of the model, which allows the model to be estimated. The prediction errors and variance are also needed in the next step of the procedure, the *updating*. In this step, the observation at t y_t, or possibly other new information, is used to update the prediction $a_{t/t-1}$ and thus refine the estimate of α_t. An augmented model can be formulated:

$$\begin{pmatrix} a_{t/t-1} \\ y_t \end{pmatrix} = \begin{pmatrix} I \\ Z_t' \end{pmatrix} \alpha_t + \begin{pmatrix} a_{t/t-1} - \alpha_t \\ \xi_t \end{pmatrix} \tag{A8}$$

The updating process is then:

$$a_{t/t} = a_{t/t-1} + P_{t/t-1} Z_t' F_t^{-1} (y_t - Z_t' a_{t/t-1}) \tag{A9}$$

$$P_t = P_{t/t-1} - P_{t/t-1} Z_t' F_t^{-1} Z_t P_{t/t-1} \tag{A10}$$

$$F_t = Z_t' P_{t/t-1} Z_t + H_t \tag{A11}$$

This process allows the efficient incorporation of new information and optimal estimation of the state vectors and the parameters of the state space model. Casting a system in the state space form and using the Kalman Filter makes it possible to evaluate the likelihood relatively easily via the prediction errors. The optimal estimates of the (hyper)parameters are then obtained by maximizing the likelihood in an iterative loop. After obtaining initial values, the model can be estimated. A final option of the Kalman filter is to obtain so-called smoothed estimates. These are a further refinement of the filtered estimates, which are obtained by adding to the one step ahead predictions the information of period t. Smoothing goes one step further, using all information in the sample to obtain optimal estimates at every period. This means that for the estimates of the state vectors at period t in the middle of the sample, all data from before and after period t are used. This allows for very good state vector estimates, but is of course less relevant for the evaluation of the real-time properties of the models.