Constructing a monthly indicator of fixed capital formation via high frequency interpolation: The state space approach

08

Floris van Ruth

Discussion paper (08009)

Statistics Netherlands

Voorburg/Heerlen, 2008

Explanation of symbols

.	= data not available
*	= 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
2005-2006	= 2005 to 2006 inclusive
2005/2006	= average of 2005 up to and including 2006
2005/'06	= crop year, financial year, school year etc. beginning in 2005 and ending in 2006
2003/'04–2005/'06	= crop year, financial year, etc. 2003/'04 to 2005/'06 inclusive

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

Publisher
Statistics Netherlands
Prinses Beatrixlaan 428
2273 XZ Voorburg

end of August 2008:
Henri Faasdreef 312
2492 JP The Hague

Prepress
Statistics Netherlands - Facility Services

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
http://www.cbs.nl

ISSN: 1572-0314

6008308009 X-10

Constructing a monthly indicator of fixed capital formation via high frequency interpolation. The state space 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 tests whether state space based approaches to high-frequency interpolation can provide a solution. The test case is creating a monthly index of private fixed capital formation. The results show that the techniques tested here can produce estimates of acceptable accuracy using only a limited number of readily available related indicators. 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

Contents

Executive Summary

There is a continuing demand for faster or more high-frequency (usually monthly) statistics. This is especially so for the economic and financial indicators important for policy making and financial markets. Unfortunately, this is at odds with other trends in official statistics; demands for increased accuracy, cost efficiency and the drive to reduce the administrative burden on the private sector. A consequence is that setting up new surveys might not be an option, whilst quality needs to be assured.

One response to these developments has been to look for methods to extract more information out of existing statistics and alternative data-sources. Usually this means trying to combine different sources into a new statistic. In recent years, attention has focused on using econometric techniques to construct new indicators and thus extract more information from existing data sources. These techniques are also able to process data from imperfect sources, improving accuracy. Developments in the past decade have made available a number of techniques which are well suited for these tasks. Using econometric techniques has additional advantages. After the development process, the actual production process can be very efficient, in the best cases completely automated. This means that the production process can be fast as well, which is very relevant for the production of monthly statistics. It is important to note that the techniques considered here will supplement rather than replace the traditional statistical process. Usually, a more traditionally produced low-frequency (quarterly, yearly) statistic is needed as reference series. In effect, these techniques build on the traditional statistical measurement process. The approach tested here is not only able to construct new monthly indicators, but can also be used to construct longer historical time series, so-called backcasting. If related high-frequency indicators are available, a backwards extrapolation of monthly and quarterly statistics is possible.

This study focuses testing the feasibility of using so-called high frequency interpolation methods in the state space framework. In the interpolation approach, monthly values are considered to be missing values between the observed, regular quarterly realisations. State space modelling is a very versatile and powerful approach, which has found uses in many areas of statistical and economic analysis. One of its great advantages is the ability to construct variables which are not directly observed, so-called latent variables. The test case in this study is a monthly indicator of the growth in private fixed capital formation. This is an unobserved quantity, as currently the highest frequency publication on fixed capital formation in the Netherlands stems from the quarterly National Accounts. Capital formation is one of the key macro-economic indicators, closely watched for reasons of policy making and business cycle analysis alike. Especially for business cycle analysis, a monthly indicator would be very useful. In the methodology tested here, a set of monthly related indicators is used to construct the missing monthly.

These related indicators were mainly the sources of fixed capital formation, as published in the yearly National Accounts. It concerns components of industrial production and imports of capital goods. Based on these data, it is shown that it is possible to use this approach to construct, without direct observation, credible monthly indicators of fixed capital formation. Using additional indicators of capacity utilization, long-term interest rates, and VAT data of the IT-sector strongly improved accuracy. The effect of adding these indicators can be compared to using information on other economic indicators and information on general economic conditions to refine the first estimates in the process of computing the National Accounts. The smallest in-sample errors were significantly below 1%-point. Real-time simulations however showed that this level of accuracy is not attainable in practice. The smallest real-time deviations were around 1.5%-points, using a new accuracy enhancing approach. This margin is still acceptable for a quick monthly indicator of such a volatile quantity as private fixed capital formation.

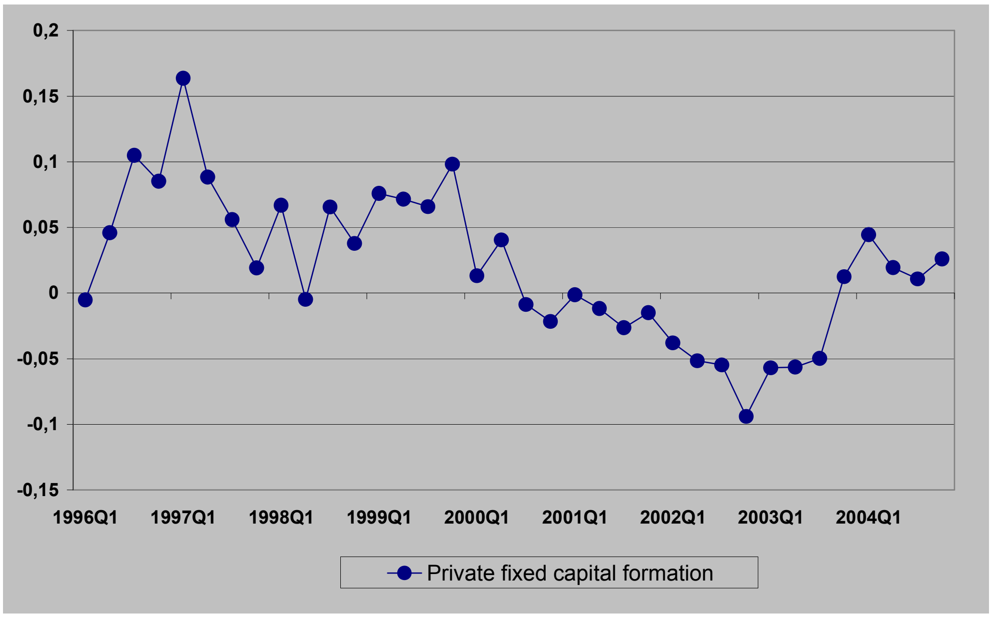
1. Introduction

Timely information is essential for reliable business cycle analysis, thus a demand for fast statistics exists. This means not only the quick publication of statistics, but preferably at a high frequency as well. A monthly publication on the developments in the main economic indicators is desired. Another important development in official statistics is the drive to increase efficiency and reduce the administrative burden on the private sector. This study has two aims; firstly the more specific one of investigating the feasibility of constructing a monthly indicator of private fixed capital formation. Furthermore, the more general methodological aim is to test whether advanced econometric techniques can be used to produce high-frequency statistics efficiently. A good way to do this is to make more use of existing data sources, whether official statistics or other databases.

The question is whether it is possible with these techniques to construct monthly or quarterly statistics without a using formal statistical framework, and whether this can be done with acceptable accuracy. Another aspect of this study is testing whether it is possible to derive a new statistic from existing and non-conventional data sources. Generally speaking, these will not have been designed for use in the estimation of the target statistic in question, and will therefore be imperfect data sources. It will be shown that this approach is able to deal with such data deficiencies. Another advantage of the econometric approach is that it offers the possibility of an extremely efficient and quick monthly production process. 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)].

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 making as well. This study focuses on the volume of private fixed capital formation, as this has the closest link with economic conditions.

Graph 1.1; Growth rate of volume of private fixed capital formation from quarterly National Accounts



Currently, 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.

This study concerns itself with the usefulness of state space based interpolation methods for the applications described above. 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. In the state space approach, the observed series is defined as being governed by an underlying dynamic process or unobserved variable which can then be efficiently modelled. One of its strengths is the ability to extract these unobserved processes or variables. This study is part of a set of two studies into the application of state space methods for the production of high-frequency statistics. In the companion study, Van Ruth (2006), a pure latent variable approach is tested. In that case, three separate unobserved monthly fixed capital formation variables are defined and computed. In-sample, the that approach worked well enough, but the results of the real-time simulations were unfortunately rather disappointing. An additional disadvantage was the fact that the computed monthly indicators exhibited relatively high volatility. Several lessons learned from that study were applied here though, and therefore these two studies are presented as a set. This also means that when reading both, there is a lot of duplication.

In the interpolation approach tested in this study, the desired high frequency indicators are defined as missing values between the observed realisations. Therefore, the high frequency indicator is unobserved, but assumed to exist. In this study, 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. In this study, variants of the local linear trend model, and state space based autoregressive distributed lags and Litterman-Fernandez models are tested. As a final innovation, quarterly forecasts of private fixed capital formation (the target/signal variable) were introduced into the state space estimation to improve real-time results.

Using these techniques, the unobserved growth rates are estimated by using 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. These indicators can be existing official statistics, but alternative sources such as tax data can be useful as well [Buiten et al. (2006)]. 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.

The state space based models were able to extract credible monthly indicators from these data. This proves that it is possible to use the latent variable/interpolation approach to construct monthly indicators, without creating an explicit measurement system for the statistic in question. The combinations of industrial production and import indicators worked reasonably well, but the accuracy of the estimates was moderate. Adding additional indicators of a different nature strongly increased the quality of the monthly capital formation indicators. These were capacity utilization, long-term interest rates, and VAT-data on IT-services. Their effect can be explained by noticing that these indicators are either very closely related to general economic conditions, which are strong determinants of capital expenditure, or fill in gaps not covered by the traditional indicators. With the expanded set of indicators it was possible to construct monthly indicators of fixed capital formation at a good level of accuracy when compared to the final quarterly estimates from the National Accounts.

A final possible application of these techniques is the construction of high-frequency historic time series for new, higher frequency statistics. When a new monthly or quarterly statistic is introduced, its usefulness is often diminished by the lack of historical data. If high frequency, related indicators are available, these techniques can be used to quickly and relatively easily produce the desired long time series.

The next section of this report contains a brief description of how fixed capital formation is measured in the quarterly National Accounts. Then the background of the state space approach is described, including the specifications used in this study. This is followed by a description of the data used. The penultimate section gives the details of the set-ups tested and the corresponding results, and the final section contains discussion and conclusions.

2. Measurement of fixed capital formation in the Dutch quarterly National Accounts

A first important observation should be that fixed capital formation as measured via the Dutch approach is in fact an indirect statistic. There is no separate capital expenditure survey, and there is no direct measurement of fixed capital formation. In stead, fixed capital formation is measured by observing possible sources of fixed capital, e.g. industrial production and imports, and then accounting for their uses and destinations.

The standard practice of measuring fixed capital formation in the quarterly National Accounts at Statistics Netherlands is in effect a three-stage process. It begins by collecting the values of the relevant (monthly) indicators for the quarter, mainly construction, components of industrial production and imports. Corrections are made for exports, and the volumes are deflated with relevant price indices to obtain volume indicators. The growth rates thus obtained are then applied to the deflated quantities in the reference quarter of the components of fixed capital formation for which that particular indicator is relevant. This results in a rough first estimate of fixed capital formation in the most recent quarter.

So far, it is probably not very difficult to replicate this process at a monthly frequency. However, the next steps in the quarterly measurement process make this increasingly more complex. The statistics for fixed capital formation are not produced in isolation, but as part of an integrated system of macro-economic quantities in the National Account. In the second step, components or corrections are added for special projects or for investment not represented by the imports and industrial production statistics. This is largely based on supplementary research by the analyst and therefore relatively time-consuming to replicate. In the final step, the estimate of fixed capital formation is confronted with the estimates of the other macro-economic indicators for that quarter and general economic conditions in a statistical integration process. This ensures that the reported macro-economic statistics are consistent and that the individual statistics are benchmarked against independent information.

A first step in constructing the desired monthly indicator is to study the composition of fixed capital formation in the Netherlands. As fixed capital formation is a composite statistics, this analysis is necessary to ascertain which data are needed. For this, the supply and use tables from the National Accounts can be used. Among other things, these show the components of the major macro-economic aggregates. For this study, the table detailing how much the major product groups contribute to fixed capital formation in 2005 was used.

Table 2.1; Composition of private fixed capital formation according to source, from supply and use tables National Accounts 2005, preliminary estimates. Product categories with a share lower than 1% were omitted.

Supply classification	*Value (millions of euros)*	*Percentage of total private fixed capital formation*
Total	86497	
metal products	1332	1.54 %
Heavy machinery	6866	7.94 %
Office machines	4909	5.68 %
Medical, telecom and other electronic equipment	2657	3.07 %
Transport equipment	11648	13.47 %
Products other industries	2624	3.03 %
Production building industry	37773	43.67 %
Property services	1451	1.68 %
Commercial services	15109	17.47 %

Broadly speaking, private fixed capital formation consists of property construction, transport equipment, machinery and electronic equipment of all sorts, and of commercial services. The last component concerns things like software, licence- and patent agreements etc., what is termed intangible fixed capital. These are counted as fixed capital as they are integral parts of the production process and last for more than a year. For the first three components, good monthly indicators are available. The last component poses more problems as no direct monthly sources are available. This will have to be solved in the state space modelling process.

3. Theoretical foundations

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)].

In this study the technique will be used to perform temporal disaggregation via the interpolation approach. The technique is used to estimate unobserved high frequency (monthly) data from related high frequency indicators, based on a link with lower frequency data (quarterly) of the target indicator. The high frequency data are treated as missing values of the (low frequency) target variable. Then, the relationships at the quarterly level are used to interpolate the monthly "missing values". In essence, the high frequency data are produced by anchoring them on the normal, higher quality, low-frequency statistic. Thus, it is probably sufficient to set up a full production system for the annual or quarterly statistic, allowing respectively the quarterly or monthly statistic to be produced with little additional effort, and acceptable quality.

The methods used for temporal disaggregation described in the literature generally use two components to estimate the target variables; 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 a number of 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 type 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. It can take many forms, allowing for great flexibility which makes it possible to describe a broad range of processes by this approach. An example is the simple yet very powerful local linear trend model [Harvey (1989)];

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \eta_t + \nu_t \\ \eta_t &= \eta_{t-1} + \zeta_t \end{aligned} \tag{2}$$

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. 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.

$$\begin{aligned} y_t &= Z_t * \alpha_t + G_t * x_t + S_t * \xi_t \\ \alpha_t &= T_t * \alpha_{t-1} + R_t * \eta_t \end{aligned} \tag{3}$$

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:

$$\begin{aligned} y_t &= Z_t * \alpha_t + S_t * \xi_t \\ \alpha_t &= T_t * \alpha_{t-1} + G_t * x_t + R_t * \eta_t \end{aligned} \tag{4}$$

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.

3.2 High-frequency interpolation and the state space approach

This study concerns testing the interpolation approach for producing high-frequency statistics. The signal variable y_t then is the target variable formed by alternating observed quarterly realisations with unobserved monthly values.

$$y = \{.., I_t^{Q,m1}, I_t^{Q,m2}, I_t^{Q}, I_t^{Q,m3}, I_{t+1}^{Q,m1}, I_{t+1}^{Q,m2}, I_{t+1}^{Q}, I_{t+1}^{Q,m3}, ..\} \quad (5)$$

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

$I^{Q1}{}_t$ = the 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. The force of the state space approach lies in the ability to estimate these unobserved values and to produce optimal estimates for the periods at the end of the sample where no quarterly realisations is yet available. The accompanying study [Van Ruth (2006)] uses a more pure latent variable approach, defining a state vector consisting of three unobserved monthly capital formation rates. An approach similar to that of Clar et al. (1998) and Israelevich and Kuttner (1993), who estimate unobserved monthly regional production indicators using observed annual indicators and related monthly indicators. The exogenous related indicators are introduced in the state equation, with the unobserved regional production indices acting as the state variables. The approaches used in this study are based on the local linear trend model of Harvey and on two of the more prominent methods of temporal interpolation [See Proietti (2004), Di Fonzo (2003)]: The Litterman and Fernandez models and the autoregressive distributed lags models. Proietti (2004) shows how to cast these last two models in the state space form. The autoregressive distributed lags models (ADL) is a quite general formulation for performing temporal interpolation, using high frequency exogenous indicators and a general autoregressive process to estimate the missing values. An example is the broadest formulation of the ADL(1,1) model:

$$y_t = \phi y_{t-1} + m + g * t + G^0 * x_t + G^1 * x_{t-1} + \varepsilon_t \quad (6)$$

where: y_t = target variable

m = constant

g = time trend

x_t = vector of exogenous variables

In this study, no time trends or autoregressive lags greater than one were used. This leads to the following state space representation:

$$
\begin{aligned}
y_t &= \alpha_t \\
\alpha_t &= \phi * \alpha_{t-1} + G_t * x_t + \varepsilon_t
\end{aligned}
\tag{7}
$$

The Litterman and Fernandez models (LF) are a specialized sub-class of the ADL models. They combine exogenous indicators with a specific autoregressive process in the form of ARIMA (1,1,0) disturbances for the Litterman model is:

$$
\begin{aligned}
y_t &= u_t + G_t * x_t \\
\Delta u_t &= \phi * \Delta u_{t-1} + \varepsilon_t
\end{aligned}
\tag{8}
$$

In state space form:

$$
\begin{aligned}
y_t &= z * \alpha_t + G * x_t \\
\alpha_t &= T * \alpha_{t-1} + H * \varepsilon_t
\end{aligned}
\tag{9}
$$

$$
\alpha_t = \begin{bmatrix} u_{t-1} \\ \Delta u_t \end{bmatrix}, z = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, T = \begin{bmatrix} 1 & 1 \\ 0 & \phi \end{bmatrix}, H = \begin{bmatrix} 0 \\ 1 \end{bmatrix}
$$

The Fernandez model arises by setting φ=0, resulting in a random walk process. These resemble the local linear trend models (LLT), which here take the form:

$$
\begin{aligned}
y_t &= \mu_t + G * x_t + \varepsilon_t \\
\mu_t &= \mu_{t-1} + \eta_t + \nu_t \\
\eta_t &= \eta_{t-1} + \zeta_t
\end{aligned}
\tag{10}
$$

Which allows for different dynamic formulations by setting either ε_t, ν_t, or ξ_t to zero. Here ε_t, was set to zero by default.

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 leads directly to the variable of interest, monthly mutations in fixed capital formation..

4. Data

One aim of this study is to derive an indicator for the monthly 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 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 section 2, 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 entered by either using the production-index for the construction materials industry or by the index of the building industry, which closely track each other. For reasons of convenience, I chose to use the first. It should be noted that the dataset was compiled *before* the Dutch National Accounts revision of 2005.

The production indices are available in the form of volume indices, just as the monthly fixed capital formation indicator is desired to be. 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 combinations of imports and price indices used can be found in appendix III. Higher order aggregates used are imports of machines and transport equipment (**M7**), imports of other goods (**M8**) and the index of production in the manufacturing industry (**IPtot**). The dataset used consisted of data from the period 1996-2004.

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 DK)**	**various**
Heavy machinery	7.94%	**Manufacturing of machinery and equipment (IP DK)**	**Generators and motors (M71), specialized machinery (M72), Metal working equipment (M73), other heavy machinery (M74)**
Office equipment	5.68%	**Manufacturing of Electrical equipment (IP DL)**	**Office equipment (M75), professional equipment (M87)**
Medical, telecom and other electronic equipment	3.07%	**Manufacturing of Electrical equipment (IP DL)**	**Communication equipment (M76), electrical equipment (77), professional equipment (87)**
Transport equipment	13.47%	**Manufacturing of transport equipment (IP DM)**	**Transport equipment (M78, M79)**
Other industries	3.03%	**various**	**various**
Building industry	43.67%	**Manufacturing of building materials (IP DI)**	**-**
Property services	1.68%	**-**	**-**
Commercial services	17.47%	**-**	**-**

Several other data related issues have to be faced. For one, a significant part of Dutch imports and of industrial production is meant for the export. 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. Monthly data on re-exports are available, but unfortunately the time series is too short to be of use for this study. And 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 these imperfect indicators, and thus at least partly correct for the distortions introduced. Though it would of course be preferable to use corrected indicators, it is not certain that these could be produced with acceptable quality on a monthly basis. The approach used in this study has the advantages of speed and simplicity, and makes 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. But it is also possible to introduce related indicators, not directly part of capital formation, to add information. In an earlier study [Van Ruth (2004)] it was found that the changes in the rate of capacity utilization and in the long-term bond yield possessed a strong link with the rate of private fixed capital formation. The rate of capacity utilization is measured in the business survey of the manufacturing industry. It is only available on a quarterly basis, but this should not be a problem as it is introduced with a lag of one quarter. The long-term bond yield is represented by the yield on 10-year Dutch government bonds, on which monthly data are available from the Dutch central bank. This indicator is introduced with a two quarter lag. The formulations using these additional indicators were termed exogenous, in contrast to the ones only containing import and production data, which here were termed endogenous models

A related Statistics Netherlands study [Buiten et al. (2006)] found that VAT-data and business survey data of the services sector are valuable sources of information on monthly developments as well. VAT- and business survey data on the IT-sector are available. These are probably good sources of information on the commercial services component of fixed capital formation, on which short-term data were lacking. The VAT-data on IT-firms of size class 4 and larger will be used, as well as the IT-sector business survey questions on expected demand and employment in the coming three months. These alternative sources are often an imperfect match with the desired statistics, due to problems like population coverage. Econometric techniques can help in correcting these shortcomings. A different drawback of these data is that the timeseries only go back to the first quarter of 2000, which means that the sample will be rather small for this type of modelling. This is especially so for the more complex models, where more parameters need to be estimated. Finally, timely availability on a monthly basis of the VAT-data is unfortunately uncertain at this time.

5. Results and monthly indicators

This section consists of three parts, each containing the results of one of the interpolation techniques tested. The estimation results are given, and for the best formulations the computed monthly fixed capital formation indicator is shown. The first part gives the result of the more general autoregressive distributed lags (ADL) models, followed by the results of the models based on the Litterman approach. The final part concerns the combination of a local linear trend variant (LLT) and additional indicators. All these models are based on the combination of the information contained in related indicators and an autoregressive process. In general, this combination was found to be essential for producing an accurate and credible monthly indicator. Overall, the state space approach is shown to be an efficient method for extracting new information from existing, both traditional and alternative, data sources. If a reference statistic is present at a lower frequency, a high frequency statistic (monthly, quarterly) can be constructed by using related high-frequency indicators. If enough relevant, good quality indicators are available, no new surveys need to be introduced. On the other hand, state space methods can be used to produce acceptable estimates when direct source data do exist, but are flawed like the VAT-data. Using these techniques it should be possible to extract an optimal estimate, thus exploiting the information which is present in the imperfect data. Another advantage of this approach is the potential efficiency and speed of the production process. This is not unimportant for a monthly statistic. Almost all indicators used are readily available, and the estimation process can be easily automated. A last possible useful application of the techniques studied here is in extrapolation of new statistics backwards in time to create longer time series. If a new quarterly or monthly statistic is introduced, then these techniques in combination with related high-frequency indicators can be used to easily construct a historic time series.

The performance of the different set-ups and combinations of indicators will be evaluated according to a number of different measures. These are the standard log likelihood and AIC values to evaluate the fit, and the standardised residuals from the prediction step are used to evaluate model quality, testing for normality and residual correlation [see Proietti (2004)]. However, the foremost measures of model performance in this study are the root mean square errors of the estimates compared to the quarterly growth-rate realizations from the National Accounts, as these directly measure the accuracy of the outcomes. For this, a synthetic quarterly indicator is constructed by taking the average of three monthly (latent variable/state) estimates.

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

The Different state space formulations were scored both on in-sample (filtered) RMSE's and out-of-sample one's. The out-of-sample results were obtained by performing a rolling regression type procedure under simulated real-time circumstances, where the sample was lengthened a month at a time. A current state (monthly) estimate was then obtained, using the same information which would have been available in practice. These last results are representative for the estimates which would be obtained every month in a production environment. Therefore, these are very important in evaluating the different models. For several models the available time series were too short to perform a meaningful out-of-sample analysis (set here at three years). In those case, a n.a. is entered in the results table.

A final, somewhat more qualitative measure of quality was used as well. It was analysed, visually and via standard deviations, whether the resulting monthly indicators were not too volatile. It is possible to have a quite accurate quarterly average with wild monthly swings in growth rates. This would not be a credible monthly indicator, even though monthly indicators can be more volatile than lower frequency ones. Therefore, the smoothness of the indicator was a factor as well.

Starting values of coefficients and state space parameters are extremely important. This especially so for the use of this approach in a production environment. Therefore, a somewhat elaborate procedure was devised. Estimates of the coefficient values from an OLS estimation at the quarterly level were used as starting values for the state space coefficients. Final values for the coefficients were then estimated in the state space estimation Starting values for the state vector parameters were found by setting up an iterative state space procedure. First, an estimation without explicitly defined starting values was performed. The outcomes of this step were used to construct preliminary starting values for another round of the state space estimation. From this, the final starting values could be derived. Data on industrial production and imports are available at several levels of aggregation. Using low level aggregates offers detailed information and the possibility of selecting the most relevant aggregates. The companion study [Van Ruth (2006)] has shown that using low-level aggregates results in more accurate estimations. Furthermore, non-significant indicators were dropped in the modelling stage.

5.1 Autoregressive distributed lags models

A number of models were created based on the very flexible autoregressive distributed lags structure. The main difference between the models is which indicators were included, and how strict the indicator selection was. The first model contains only “real” related indicators, i.e. industrial production and imports. The others contain the informative indicators; capacity utilization and bond yield, and in the final model VAT-data of the IT-sector as well. For model 4 variable selection criteria were applied very strictly, to come to a very parsimonious model. In other cases the selection process was less stringent in order to test whether otherwise information is wasted.. The composition of the best performing models is described in the table below.

Table 5.1; Indicators entered in autoregressive distributed lags models.

State Space Model	*Industrial production indicators (lag, months)*	*Imports indicators (lag, months)*	*Other indicators (lag, months)*
ADL 1	Construction, Electronics(-1), Machines(-1), transport equipment	M74, M75, M77, M77(-1)	
ADL 2	Construction, Electronics, Machines(-1), transport equipment	M74, M75, M76, M77	Capacity Utilization (-3), 10-year bond yield (-6)
ADL 3	Construction, Electronics, transport equipment	M74, M75, M77	Capacity Utilization (-3), 10-year bond yield (-6)
ADL 4	Construction, Electronics	-	Capacity Utilization (-3), 10-year bond yield (-6)
ADL 5	Construction, Electronics, transport equipment	M74, M75, M77	Capacity Utilization (-3), 10-year bond yield (-6), VAT IT-services (-3)

These models were estimated, and the resulting monthly capital formation growth rates constructed. The usual test statistics can be found in table 5.2, along with the in-sample (filtered) and real-time quarterly deviations from the reference quarterly growth rate realisations.

Table 5.2; Estimation results for the autoregressive distributed lags mdoels. Reference is quarterly realisations of growth rate of private fixed capital formation.

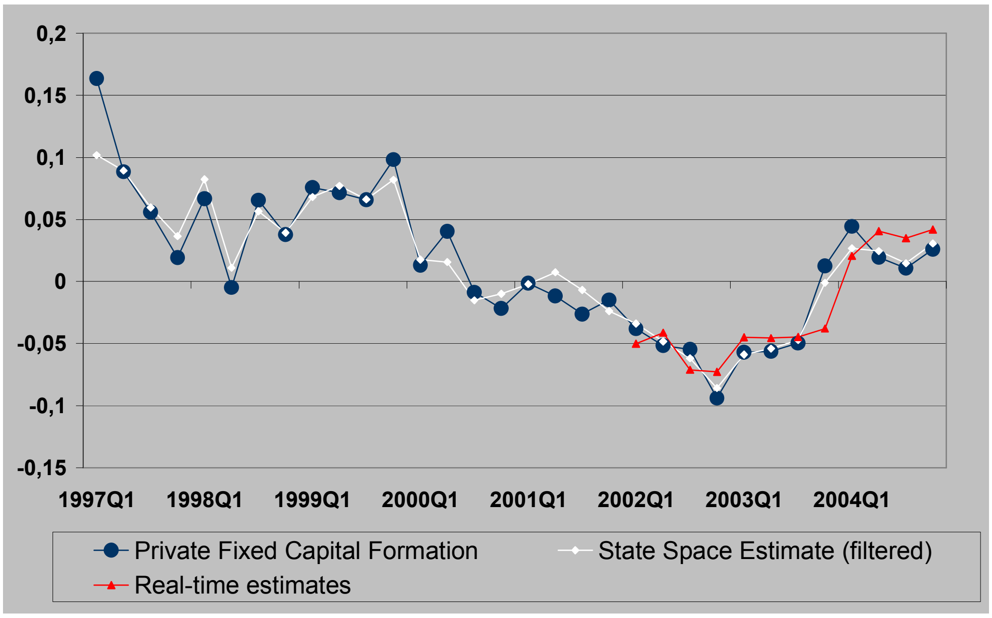
Model	*RMSE filtered (%-points)*	*RMSE real-time (%-points)*	*Log Likeli-hood*	*AIC*	*Q-prob 4 lags*	*Jaque-Berra probability*
ADL 1	**4.9%**	**3.8%**	69.1	-3.63	0.601	0.752
ADL 2	**1.3%**	**4.2%**	70.3	-3.64	0.980	0.284
ADL 3	**1.9%**	**2.4%**	75.3	-4.02	0.344	0.650
ADL 4	**1.1%**	**2.2%**	77.0	-4.12	0.015	0.582
ADL 5	**0.3%**	**n.a.**	59.7	-5.02	0.480	0.249

These results show that it is possible to measure monthly capital formation via state space-based interpolation. This is an important result. It shows that by using readily available monthly indicators and with a relatively simple state space model based on the quarterly National Accounts realisations, it is possible to construct a monthly statistic. This approach allows very fast production of statistics, without the need of an elaborate production process such as in the quarterly National Accounts. However, this approach in effect builds on the analytical work done at the quarterly level.

Accuracy varies significantly between the models. One conclusion that can be drawn immediately is that the additional indicators, bond yield and capacity utilization are crucial for an accurate measurement of capital formation. The information contained in these indicators probably represents the influence of the state of the economy on capital formation. The in-sample (filtered) results are very good, an average error of around 1%-point is quite acceptable for a statistic as volatile as fixed capital formation. Almost all the residual statistics are good as well, indicating no misspecification. The results for model ADL 5, containing VAT-data, cannot be easily compared, as the estimation sample was much shorter due to the limited availability of the VAT-data. This is also the reason that the real-time simulation could not be performed. A close analysis of the model indicates that its superior performance probably will not survive in practice. But the VAT-data remain an promising data source.

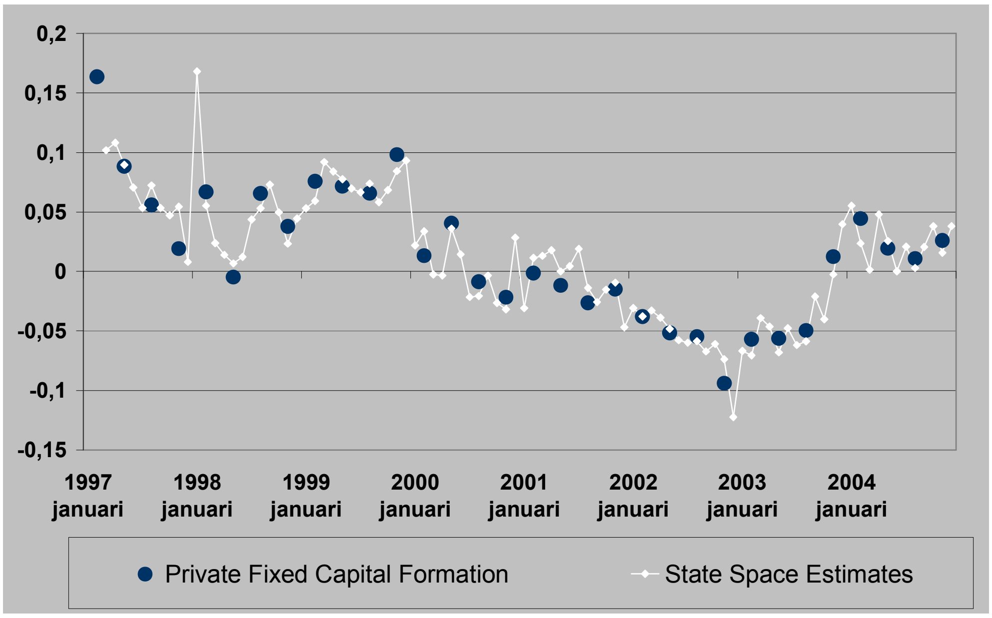
The best results, especially from the real-time analysis, are for model ADL 4. It is, probably not by chance, the most parsimonious model, containing relatively few indicators. The calculated quarterly averages of its constructed real-time and filtered monthly indicators are shown next to the actual quarterly realisations of capital formation in graph 5.1.

Graph 5.1; Synthetic quarterly year-on-year growth rates, both filtered and real-time data, for model ADL 4 compared to actual realisations of the change in private fixed capital formation from the National Accounts.



Overall, the estimated indicators trace the actual realised growth rate quite well. But especially the real-time results show occasional large deviations. The corresponding monthly index is of private fixed capital formation is also quite well behaved, see graph 5.2. This shows what the monthly capital formation statistic could look like.

Graph 5.2; Synthetic monthly year-on-year growth rates (filtered) of model ADL 4 compared to quarterly realisations of the change in private fixed capital formation from the National Accounts.



The monthly indicator may seem somewhat more volatile than the quarterly one, but this is purely due to the different frequency as becomes clear when comparing the standard deviations in appendix II. There, means and standard deviations of quarterly private fixed capital formation, all indicators used and all estimated synthetic indicators can be found. This makes it also possible to compare volatilities of different set-ups in order to find those which most similar to the reference series. It is interesting to see that the high volatility of the source statistics is not translated into the computed monthly capital formation growth rates. This did happen in the latent variable state space approach studied in Van Ruth (2006). The interpolation techniques are apparently able to compensate for this. Overall, state space models in the autoregressive dynamic lags formulation are able to produce a credible monthly index of private fixed capital formation. Only two production indicators and two additional indicators of the state of the economy are needed, though the whole of course builds on the standard quarterly National Accounts.

5.2 Litterman and Fernandez models

The Litterman and Fernandez models are a constrained variant of the ADL-models, with a specific dynamic structure. This resulted in rather smaller models after indicator selection, see table 5.3.

Table 5.3; Indicators entered in Litterman-Fernadez models.

State Space Model	*Industrial production indicators*	*Imports indicators*	*Other indicators (lag, months)*
LF 1	Construction	-	-
LF 2	Construction	-	10-year bond yield (-6)
LF 3	Construction, Electronics, transport equipment	M74, M75, M77	Capacity Utilization (-3), 10-year bond yield (-6), VAT IT-services (-3)

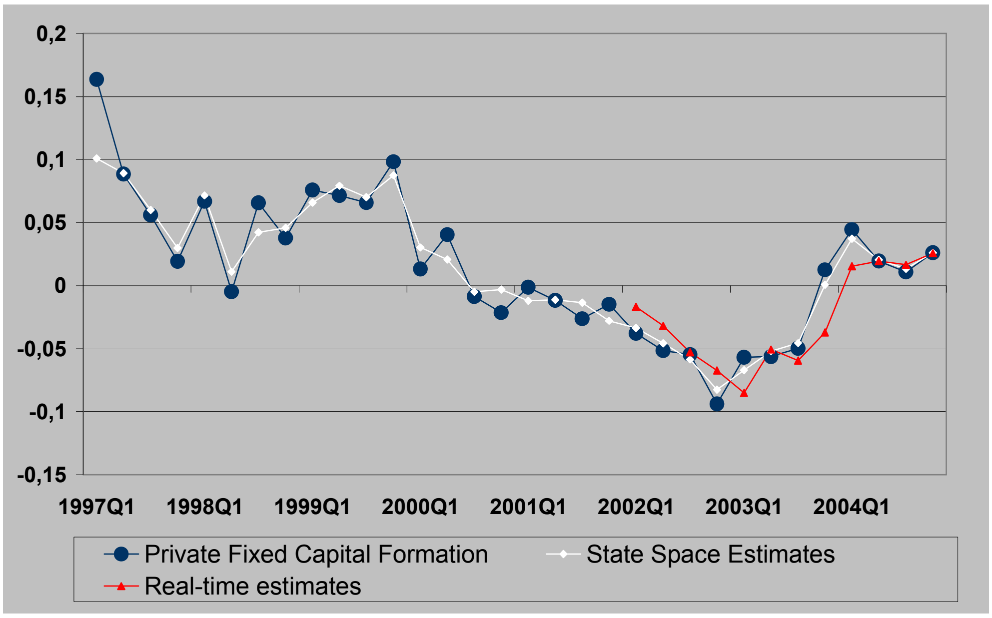
The results in the table below show that the parsimonious nature of these models does not reduce their accuracy. Apparently the autoregressive component is able to capture a large part of the dynamics by itself. The additional information contained in the construction and bond yield indicators is sufficient to construct reliable monthly indices, see table 5.4 for the modelling results.

Table 5.4; Estimation results for Litterman-Fernandez models. Reference is quarterly realisations of growth rate of private fixed capital formation.

Model	*RMSE filtered (%-points)*	*RMSE real-time (%-points)*	*Log Likeli-hood*	*AIC*	*Q-prob 4 lags*	*Jaque-Berra probability*
LF 1	**0.9%**	**2.4%**	60.2	-3.12	0.057	0.323
LF 2	**1.1%**	**2.2%**	60.0	-3.24	0.167	0.225
LF 3	**0.25%**	**n.a.**	48.6	-3.85	0.506	0.193

Again, the short time span makes it difficult to evaluate the good performance of the model containing IT-services VAT-data (LF 3). For now, the conclusion remains that it is doubtful whether this performance can be replicated in practice. More testing with longer time-series is needed. The very parsimonious Litterman-Fernandez models are as accurate as the somewhat more elaborate autoregressive distributed lags models, both in the in-sample and real-time analysis. The minimum expected real-time deviation is that of model LF 2 at 2.2%-points, see also graph 5.3. Residual statistics do not indicate any specification problems.

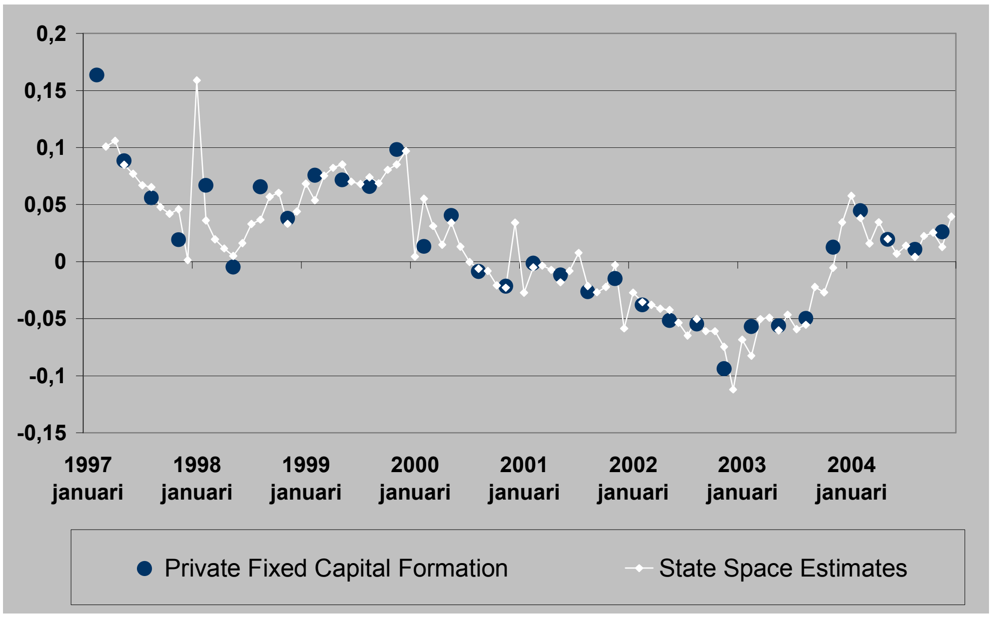
Graph 5.3; Synthetic quarterly year-on-year growth rates of Litterman-Fernandez models compared to actual realisations of the change in private fixed capital formation from the National Accounts.



As can be seen in graph 5.2, the synthetic growth rates from model LF 2 track the quarterly realisations quite well. The real-time estimates again exhibit a few large deviations in individual quarters, but overall the resulting growth rates are credible as well. When considering the margin of error, one has also to keep in mind that in practice these monthly statistics will be produced next to the regular quarterly measurement process. This means that from the second month of a quarter, it will probably be possible to start adjusting the outcomes to the quarterly national accounts.

In graph 5.4 the corresponding monthly estimates are shown, next to the quarterly realisations. Apart from a few problematic points at the beginning, this monthly indicator seems to track the realisations better than the monthly ADL estimates in the previous section, as the evolution of the series is smoother here.

Graph 5.4; Synthetic monthly year-on-year growth rates of Litterman-Fernandez models compared to quarterly realisations of the change in private fixed capital formation from the National Accounts.



Overall, the results are quite promising. Estimating monthly private fixed capital formation growth rates via a state space based interpolation approach is shown to be feasible. This means that is possible to construct monthly indicators for fixed capital formation without explicitly defining them, using a state space model to extract the necessary information from related monthly indicators. These are mainly production and imports statistics selected via an analysis of the yearly National Accounts input-output tables. But information from additional, indirectly related indicators is shown to be essential for accurate measurement. Earlier research [Van Ruth (2004)] has shown that changes in the long-term bond yield and in the rate of capacity utilization have a strong, leading link with capital formation. These indicators reflect changes in business conditions, and thus contain additional information relevant for fixed capital formation. Including these significantly enhanced accuracy. Other Statistics Netherlands research [Buiten et all. (2006)] indicated that VAT- and business survey data of the IT-industry contain information on developments in the IT-services industry. This industry is an important source of intangible fixed capital formation, thus these data give information on a component of fixed capital formation not covered by other indicators. The best set-ups produce synthetic quarterly indicators with a good accuracy when compared to the reference quarterly National Accounts growth rates, especially for a first monthly estimate. There is no clear winner among the three types of state space interpolation models tested here. The best autoregressive distributed lags models, Litterman models and local linear trend variants all possess similar accuracy.

5.3 Local Linear Trend models

In the literature, many models are built around some form of autoregressive process, usually a variant of the local linear trend model [Harvey (1989), (2000)]. The results in those studies indicate that this is a quite powerful approach. Here, quite a number of different variants were tested, but in the end the simplest form was the most successful one. The trend component has the form of a simple random walk, complemented by several additional indicators, which incidentally equals the Fernandez model:

$$
\begin{aligned}
y_t &= \mu_t + G * x_t + \varepsilon_t \\
\mu_t &= \mu_{t-1} + \eta_t + \nu_t \qquad () \\
\eta_t &= \eta_{t-1}
\end{aligned}
$$

A number of different sets of additional indicators (x_t) were tested, see table 5.5.

Table 5.5; Indicator sets added to local linear trend model.

State Space Model	*Industrial production indicators*	*Imports indicators*	*Other indicators (lag, months)*
LLT 1	Construction, Electronics	-	Capacity Utilization (-3), 10-year bond yield (-6)
LLT 2	Construction, Machines ,	M71, M74, M75, M76, M77	Capacity Utilization (-3), 10-year bond yield (-6)
LLT 3 **Exouit2difalt**	Construction, Electronics	M71	Capacity Utilization (-3), 10-year bond yield (-6)
LLT 4	Construction, Machines ,	M71, M74, M75, M76, M77	Capacity Utilization (-3), 10-year bond yield (-6), VAT IT-services (-3)
LLT 5	Construction	M74, M75, M76	-

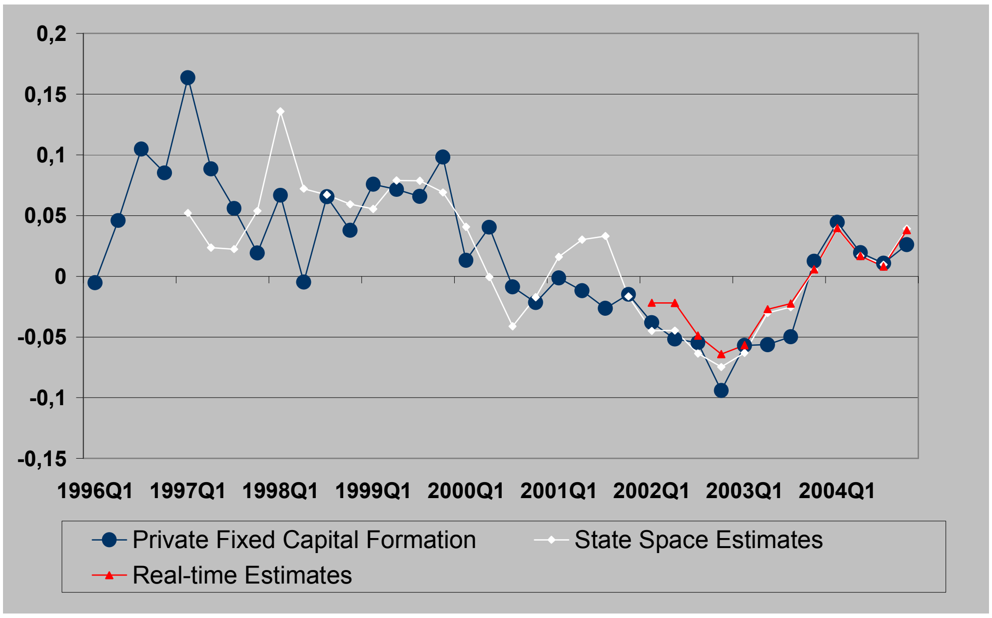
The resulting models seem to perform quite well, again with differing in-sample accuracies, but comparable simulated real-time errors. Overall, real time performance seems to be anything between 0.1%-points and 2.5%-points worse than the in-sample results.

Table 5.6;. Estimation results for Local Linear trend models. Reference is quarterly realisations of growth rate of private fixed capital formation.

Model	*RMSE filtered (%-points)*	*RMSE real-time (%-points)*	*Log Likeli-hood*	*AIC*	*Q-prob 4 lags*	*Jaque-Berra probability*
LLT 1	**2.3%**	**1.8%**	62.6	-3.27	0.042	0.663
LLT 2	**0.7%**	**3.2%**	60.9	-3.18	0.215	0.828
LLT 3	**0.6%**	**2.0%**	70.7	-4.04	0.062	0.939
LLT 4	**1.5%**	**n.a.**	36.3	-2.67	0.487	0.844
LLT 5	**1.9%**	**2.0%**	52.1	-2.82	0.004	0.956

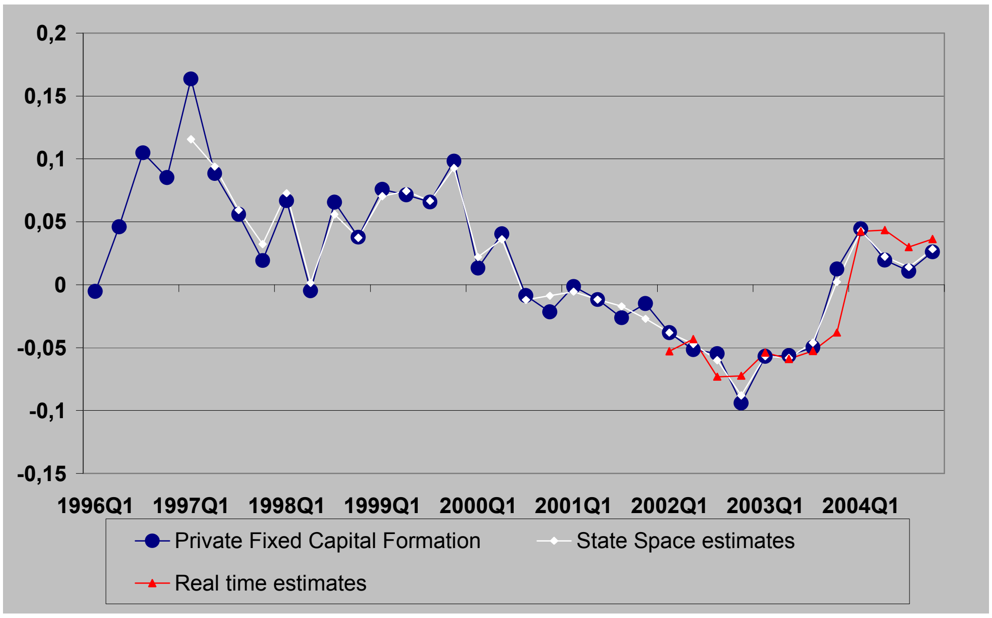
The most accurate model according to the real-time accuracy is LLT 1, with an error of 1.8%-points. However, this formulation has the rather strange problem that in-sample results are worse those of the out-of-sample test results. This is due to the fact that the model performs badly in the first part of the sample, but much better in the last few years, the out-of-sample test span. Real-time estimates are therefore not more accurate than in-sample ones, see graph 5.5. Another factor which casts some doubt on this specific model is the presence of autocorrelations in the residuals, as measured by the Q-statistic. The same problem is present in LLT 5. This indicates that these specifications are suboptimal.

Graph 5.5; Synthetic quarterly year-on-year growth rates of model LLT 1 compared to actual realisations of the change in private fixed capital formation from the National Accounts.



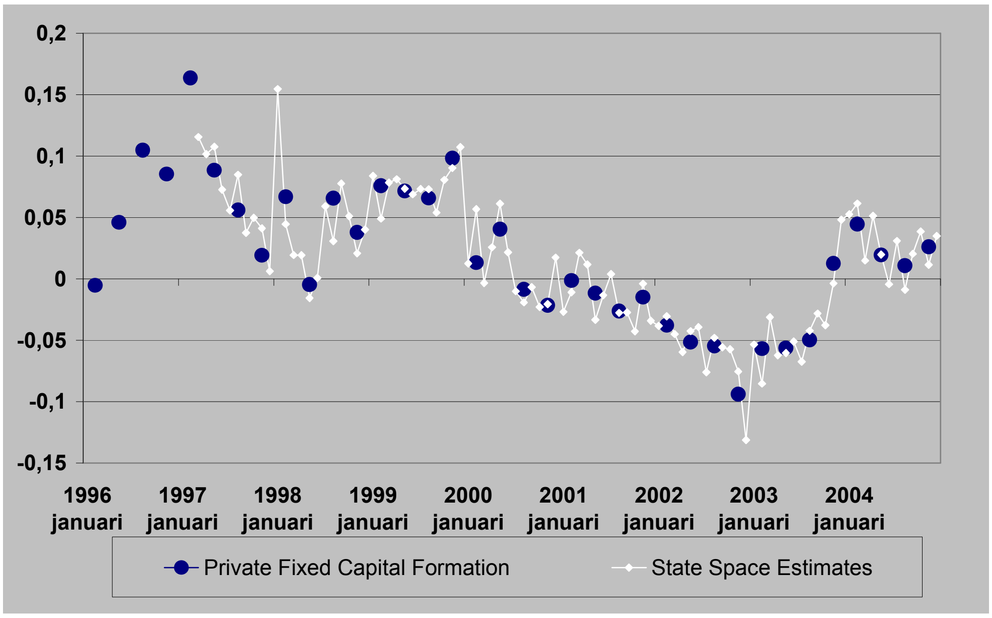
These results cast doubt on the usefulness of model LLT 1. It seems safer to consider the next best performing one, model LLT 3. It is similar to LLT 1, but contains an additional indicator and is constructed in a slightly different manner (the monthly indicators are entered in the signal equation instead of in the state equation). Its residual statistics are in order and with a real time error of 2%-points, it is still reasonably accurate. Both the real time and in-sample results track the reference quarterly realisations quite well.

Graph 5.6; Synthetic quarterly year-on-year growth rates of model LLT 3 compared to actual realisations of the change in private fixed capital formation from the National Accounts.



The corresponding monthly indicator is a little more volatile than the ones from sections 5.1 and 5.2, but still shows a credible evolution. As mentioned before, a monthly indicator is bound to be more volatile than a quarterly one.

Graph 5.7; Synthetic monthly year-on-year growth rates of model LLT 3 compared to quarterly realisations of the change in private fixed capital formation from the National Accounts.



5.4 A new approach for improving accuracy

Obviously, the main reason for the inferior performance of the models in real-time simulations is the lack of current data of the reference capital formation series. This series is the signal variable in the state space model. Accuracy of state space models improves greatly when more signal data are available. This offers an opportunity to enhance the accuracy of the models tested in this study. A quarterly time series model can be used to produce a forecast for the current quarterly value of the growth rate of fixed capital formation. This forecast can then be substituted into the signal series instead of the yet unobserved realisation. If the forecast is accurate enough, this should improve the determination of the corresponding monthly growth rates.

Several ARMA, ARMAX and quarterly state space models were tested, both in-sample and via rolling regression forecasting exercises. The following model performed best:

I = -0.083 -0.329***construction**(-1) +0.819***machines**(-1)+0.259***import75**
(0.0000) (0.0024) (0.0000) (0.0000)
-2.644***capacity utilization**(-2) – 3.196***10year-bond yield**(-3)
(0.0003) (0.0005)
-0.541***AR**(2) -0.992***MA**(3)
(0.0233) (0.0000)

R^2 = 0.95, **AIC**= -5.58, **Q-statistic** (4 lags) = 0.239,
Jarque-Berra = 0.283, **LM-statistic** = 0.196,
RMSE in-sample= 1.09%-points,
rolling regression forecast RMSE = 0.62%-points

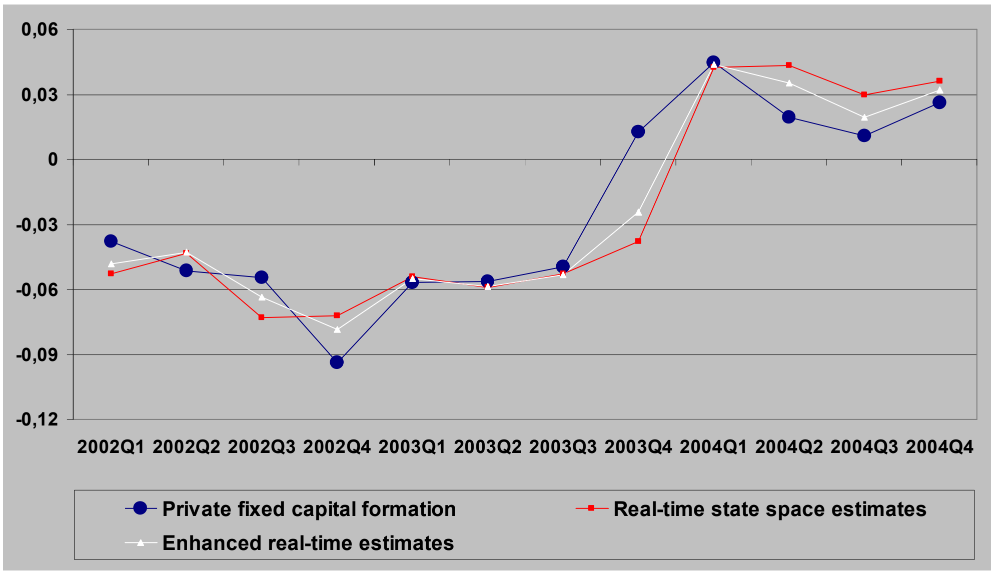
It should be noted that this model is optimal in this sample, in actual production circumstances the model used should be updated regularly. In order to maintain the real-time character of the test results, forecasts were produced from the quarterly model in a rolling regression forecast exercise. These simulated real-time forecasts were then substituted for the current value in the reference/signal series of the state space models. This strongly improved results, with far lower errors in the real-time simulation, see table 5.7

Table 5.7;. Comparison of accuracy of standard real-time results and of real-time results of state space models enhanced with quarterly forecast of private fixed capital formation. Reference is quarterly realisations of growth rate of private fixed capital formation.

Model	*RMSE in-sample (%-points)*	*RMSE real-time simulation (%-points)*	*RMSE real-time simulation enhanced with quarterly forecasts (%-points)*
ADL 4	**1.1%**	**2.2%**	**1.6%**
LF 2	**1.1%**	**2.2%**	**1.6%**
LLT 1	**2.3%**	**1.8%**	**1.6%**
LLT 3	**0.6%**	**2.0%**	**1.4%**

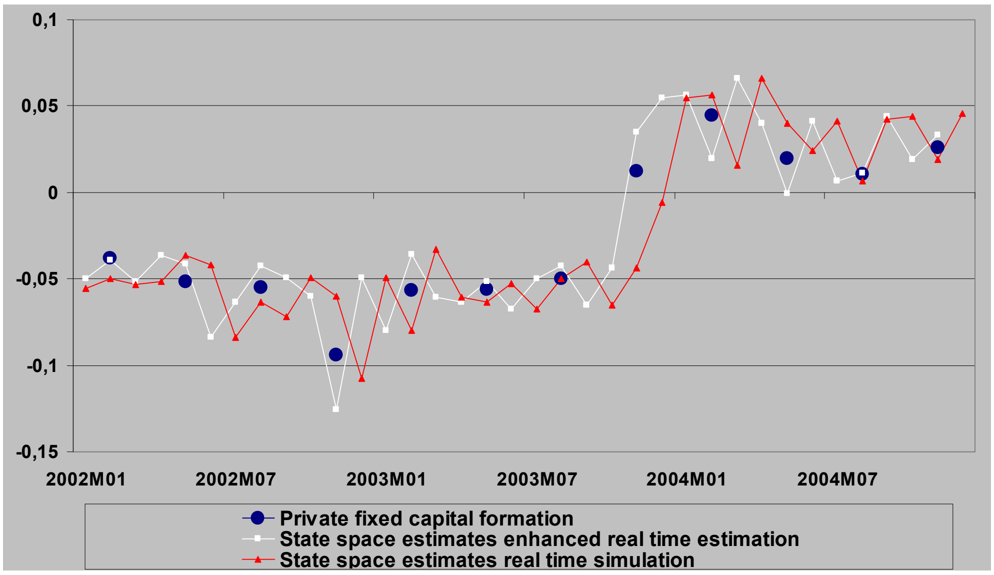
It seems that the availability of even a rough approximation of the relevant quarterly realisation of the target series strongly improves the estimation process. The real-time accuracy of the constructed monthly indicatores is now much more acceptable. As can be seen in graph 5.8, the result is that each quarter, the average of the monthly indicator is somewhat, but consistenly, closer to the reference quarterly realisations.

Graph 5.8; Synthetic quarterly year-on-year growth rates of model LLT 3 from real time simulation and from real time simulation enhanced with quarterly forecasts of private fixed capital formation, compared to actual realisations of the change in private fixed capital formation from the National Accounts.



The beneficial effects of this approach are even more visible when comparing the monthly real-time estimates with and without the enhancement with the quarterly forecasts. The enhanced monthly series is more credible, tracking the quarterly values

Graph 5.9; Synthetic monthly year-on-year growth rates of model LLT 3 from real time simulation and from real time simulation enhanced with quarterly forecasts of private fixed capital formation. Reference series are the quarterly realisations of the change in private fixed capital formation from the National Accounts.



Conclusions

The first question which this study set out to answer is whether it is possible to use state space techniques to construct high-frequency statistics without explicitly defining them. This means not constructing a formal statistical framework to measure a quantity, but to use econometric techniques to derive the desired information from related indicators and the low-frequency reference series. Specifically, the aim was to construct a monthly indicator for private fixed capital formation.

In this study, the interpolation approach to producing high frequency statistics was tested. This means that the to be estimated monthly indicator values are considered to be missing values between the quarterly realisations. Via econometric techniques these are then interpolated. This is a traditional approach, but applied here in a more modern fashion by using the state space framework. This allows for greater flexibility and computational efficiency. The state space approach is very well suited for identifying quantities which are not directly observed, like the monthly growth rate of fixed capital formation in this case. Three types of interpolation techniques were tested; the autoregressive distributed lags models, the Litterman-Fernandez models and local linear trend variants. All three contain some form of autoregressive process, combined with new information from monthly related indicators. From this, the evolution of fixed capital formation was derived. This approach proved to be very successful. In all formulations tested, at least a credible monthly indicator of fixed capital formation development was produced. Accuracy varied, but the basic approach clearly is viable. It was found that for accurate and consistent estimation, it is important to use appropriate starting values for the coefficients and hyperparameters. This was achieved by using the coefficient values from an ordinary regression at the quarterly level as starting values for the state space coefficients, and by setting up an iterative state space procedure for estimating good starting values for the state vector parameters.

The elementary form of the models tested here contained just import and industrial production indicators, at the two-digit level. These were selected based on information from the supply and use tables from the National Accounts. The resulting monthly indicators varied in accuracy, but were generally not the best ones possible. These estimates lacked two components; information on components of fixed capital formation not covered by the import and production statistics, such as investment in intangibles, and information on general economic conditions which influence fixed capital formation. Using non-traditional indicators which make up for these deficiencies improved accuracy. Business survey data and bond yields, which reflect economic conditions were yielded significant additional information. VAT-data on the IT-sector tended to increase accuracy as well, though the short time series available made it difficult to assess the quality of these models. Overall, it was shown that if the right set of related monthly indicators is available, good accuracy is possible. Interesting is the fact that the best models needed only a few related indicators, typically around four. This can be caused by the fact that state space models tend to perform best when not too many parameters are used.

In-sample, the best results possessed an average deviation of the computed quarterly growth rate from the reference quarterly realisations of less than one percentage point. There was no real difference in accuracy between the different methods tested. Important is that the monthly indicators were credible, developing relatively smoothly around the quarterly realisations. This is probably due to the autoregressive component present in all techniques tested. The source statistics can be very volatile, but the models were able to cope with this. Another reason for the success of this approach is probably the fact that it was possible to use monthly indicators with a strong direct link with capital formation, which facilitates estimation. Real-time simulations indicated however that in practice an accuracy of around 2%-points in the growht rate would be achieved. This is still quite acceptable for a monthly statistic which is meant to give a first impression of developments. But a clear improvement of accuracy can be achieved by enhancing the estimation procedure with forecasts of the growth rate of private fixed capital formation likely to be realised in the current quarter. Average errors came down to about 1.5%-points when using this approach. The signal variable (here fixed capital formation) is very important in state space estimation. By providing a rough estimate of the missing value of the current quarter, accuracy can be improvedBy coordinating the monthly production process with the quarterly National Accounts, further improvements in accuracy should be possible. As far as the real-time results are concered, again no state space formulation was clearly superior to the others.

Literature

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

Buiten, G, El Bouchehati, M, Jansen, M, Mikulic, B, Ras, P, Van Ruth, F, Schellings, R. (2006) “A feasibility, study for a monthly turnover indicator in monthly services”. Statistics Netherlands report.

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.(2004) “Investeringen; een onderzoek naar belangrijke indicatoren en mogelijkheden voor nowcasting” (in Dutch). Statistics Netherlands report.

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. (2006), "A latent variable approach to constructing monthly indicators; fixed capital formation". Statistics Netherlands Discussion paper 2006.

Appendix I; 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]:

$$\begin{aligned} y_t &= Z_t * \alpha_t + S_t * \xi_t \\ \alpha_t &= T_t * \alpha_{t-1} + R_t * \eta_t \end{aligned} \tag{A1}$$

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;

$$a_{t/t-1} = T_t * \alpha_{t-1} \tag{A2}$$

With error:

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

Which gives for the covariance matrix of the estimation error:

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

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:

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

which has variance:

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

The prediction error ν_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.

Appendix II; Means and volatilities indicators and monthly estimates

Indicator	*Mean of relative growth rate **(1997-2004)***	*Standard deviation of relative growth rate **(1997-2004)***
Fixed capital formation	0.015	0.056
IP tot dy	0.001	0.0287
IP DIbouw dy	0.011	0.098
IP DKmach dy	0.018	0.054
IP DLelek dy	-0.004	0.075
IP DMtrans dy	0.025	0.065
M7	0.061	0.108
M8	0.035	0.084
M71	0.019	0.198
M72	0.031	0.085
M73	-0.018	0.266
M74	0.020	0.077
M75	0.091	0.159
M76	0.134	0.241
M77	0.080	0.157
M78	0.022	0.085
M79	0.064	0.734
M87	0.090	0.116
Model ADL 1	0.006	0.052
Model ADL 2	0.011	0.054
Model ADL 3	0.011	0.050
Model ADL 4	0.012	0.052
Model ADL 5	0.026	0.040
Model LF 1	0.012	0.051
Model LF 2	0.012	0.051
Model LLT 1	0.017	0.054
Model LLT 2	0.015	0.059
Model LLT 3	0.012	0.054
Model LLT 5	0.013	0.048

Indicator	*Mean of relative growth rate **(2000-2004)***	*Standard deviation of relative growth rate **(2000-2004)***
Fixed capital formation	-0.021	0.037
Model ADL 5	-0.019	0.047
Model LF 3	-0.019	0.048
Model LLT 4	-0.018	0.043

Appendix III; Producer price indices used for deflating import data.

Import category		*Producer price index used for deflating*
Machines and transport equipment	**M7**	DK machines and tools
Other goods	**M8**	DK machines and tools
Generators and motors	**M71**	DK machines and tools
specialized machinery	**M72**	DK machines and tools
Metal working equipment	**M73**	DK machines and tools
other heavy machinery	**M74**	DK machines and tools
Office equipment	**M75**	US bureau of labour statistics Producer price index electronic computer manufacturing
Communication equipment	**M76**	DL electrical and optical equipment
electrical equipment	**M77**	DL electrical and optical equipment
Road transport equipment	**M78**	DM transport equipment
Other transport equipment	**M79**	DM transport equipment
professional equipment	**M87**	US bureau of labour statistics Producer price index electronic computer manufacturing