A latent variable approach to constructing monthly indicators; fixed capital formation

08

Floris van Ruth

Discussion paper (08010)

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

6008308010 X-10

A latent variable approach to constructing monthly indicators; fixed capital formation

Floris van Ruth

Summary: Econometric techniques can offer alternative methods for constructing statistical indicators. It was tested whether it is possible to use a state space-based latent variable approach to construct a monthly indicator of private fixed capital formation. This means that it is not necessary to explicitly define the statistic which is to be measured. By using related monthly indicators and the quarterly realisations of fixed capital formation, credible monthly indicators could be constructed. Simulations however showed that the method would be too inaccurate in real-time situations for it to be used in practice. Specifications with an autoregressive component performed best, and warrant further research.

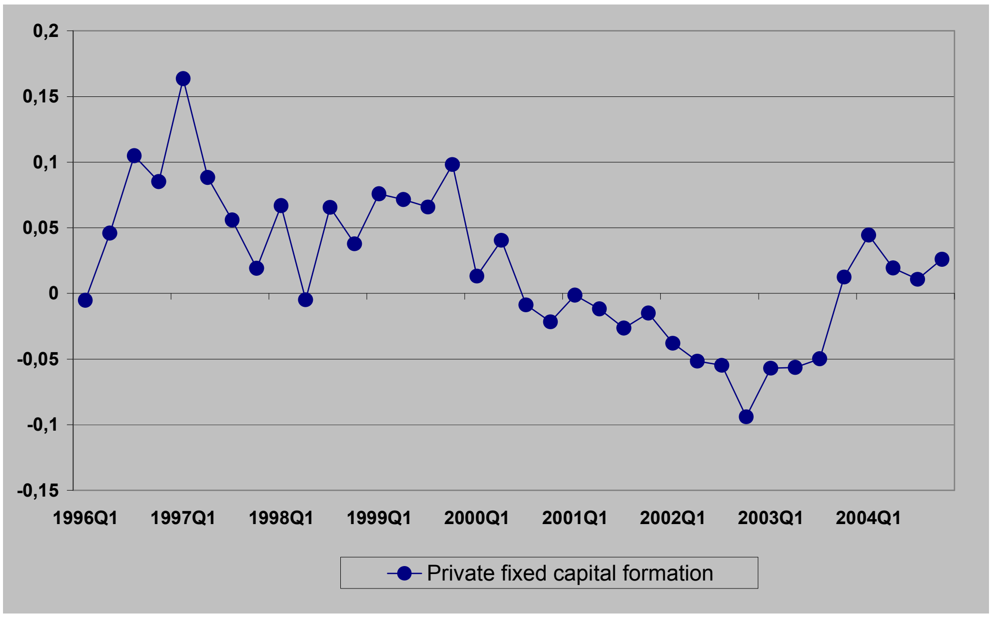
Keywords: State Space, Kalman filter, latent variables, temporal disaggregation, business cycle indicators, fixed capital formation, quarterly National Accounts

1. Introduction

This study was performed to test the viability of using a pure latent variable approach to construct statistical indicators. The target variable is a monthly indicator of the development of private fixed capital formation. Here, a latent variable approach means that a separate monthly indicator is assumed to exist, but is not measured independently. Instead, it is derived from the development in related monthly indicators and the development of the quarterly indicator. This study aims to test and evaluate the use of one of the more advanced econometric techniques in the production of high-frequency statistics, and at the same time whether it is thus possible to derive a new statistic from existing and non-conventional data sources. 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 is the precursor to Van Ruth (2006), where a different state space approach to this problem is tested. When reading both, some overlap will be evident.

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 formulation as well.

Graph 1.1; Year-on-year growth rates of the volume of private fixed capital formation from quarterly National Accounts



This study focuses on the volume of private fixed capital formation, as this has the closest link with economic conditions. Currently, the highest frequency at which data on capital formation are published by Statistics Netherlands is quarterly. It is part of the quarterly National Accounts, of which the methodology is based on the annual system of National Accounts.

This study is based on the state space framework, a versatile method for describing the dynamics of economic and other variables. The way it is constructed allows for the estimation of very diverse and otherwise difficult to evaluate models. It defines the observed series 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. In this study, the observed variable, the quarterly growth rate of private capital formation, is defined as consisting of three monthly growth rates, the unobserved or latent variables. The unobserved monthly growth rates are estimated by using related monthly indicators. These can be existing official statistics, such as imports and industrial production, but alternative sources of data can be useful as well [Buiten et al. (2006)].

Formulations using different combinations of indicators at different levels of aggregation were tested. The state space models were able to extract credible monthly growth rates from these indicators. In itself, this is quite impressive, because it means that it is thus possible to construct statistics without creating an explicit statistical framework for the statistic in question. Using many low aggregation level indicators worked best, probably because it enables the relative importance of each class of goods to be incorporated explicitly. 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 business survey 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. Overall, the outcomes are rather ambiguous. On the one hand it was possible with the expanded set of indicators to construct monthly indicators of the volume fixed capital formation at a good level of accuracy when compared to the final quarterly estimates from the National Accounts. However, the evolution of the monthly indicators was somewhat erratic, and more seriously in a real-time simulation the accuracy decreased dramatically. It seems that this latent variable approach possesses much better in-sample properties than for actual production.

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. State Space approach

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)]. For most of these applications, the crucial aspect of the state space framework is the ability to estimate unobserved components, or latent variables. These are quantities which can not be independently observed, but which are nonetheless present in the dynamics of observed time series. For example, in business cycle research it can be used to find the cyclical component in economic growth, which cannot be observed directly. The same goes for seasonal components. Another application can be found at a number of European statistical institutes, which estimate certain quarterly components of GDP using annual data and a form of state space modelling [Di Fonzo (2004), Proietti (2003, 2004)]. 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.

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. The approach used in this study leans heavily on these 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 or latent, 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)]:

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

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 vectors. By elaborating on this basic structure, virtually all type of dynamics can be modelled. For most of the formulations tested in this study Z_t and S_t, are constant coefficient vectors of dimension 3. T_t, and R_t are constant coefficient matrices of dimension 3x3. The disturbances ξ_t and η_t have mean zero and time-independent covariance matrix H and Q respectively, and are serially uncorrelated. The system parameters, or hyperparameters, Z, S, T, R, H and Q are unknown and need to be determined. The models were estimated using the eviews-package. It uses the Kalman filter to evaluate the log-likelihood via one-step ahead prediction errors. The Kalman filter is also used to construct the state estimates, which are this study the latent monthly realisations of fixed capital formation. For this, a powerful technique known as the Kalman Filter is used. A short discussion of the Kalman Filter can be found in Appendix I.

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, 2003)];

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

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. 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.

In this study, the y_t will be one or more directly observed low frequency target variables. I followed 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. In their studies, the monthly regional production indicators are anchored on the annual regional and monthly national production indicators, and are estimated using monthly labour volumes and electricity production statistics (the related indicators). Essential is that the exogenous related indicators are introduced in the state equation, with the unobserved regional production indices acting as the state variables.

It is then a small step from observed annual indices and unobserved monthly indices to observed quarterly indices and unobserved monthly of fixed capital formation. In the model, the unobserved monthly growth rates will be anchored to the observed quarterly mutations by requiring the average of the three monthly growth rates to equal the quarterly growth rate. Industrial production, import data, and several additional indicators will be used as related indicators. The results should be even better than those of Clar et al. (1998), as there is a stronger direct link between these related indicators and the target variable, fixed capital formation. The general form of the state space model used here is given below. The monthly indicators are entered in the form of relative year-on-year changes. 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. Two slightly different state space formulations will be tested, the first is termed *variant A*.

State Space formulation Variant A

$$I_t^Q = \left(\tfrac{1}{3}\ \tfrac{1}{3}\ \tfrac{1}{3}\right) * \begin{pmatrix} I_t^{m1} \\ I_t^{m2} \\ I_t^{m3} \end{pmatrix} + (\varepsilon_t)$$

$$\begin{pmatrix} I_t^{m1} \\ I_t^{m2} \\ I_t^{m3} \end{pmatrix} = \begin{pmatrix} 0...0...0 \\ 0...0...0 \\ 0...0...0 \end{pmatrix} * \begin{pmatrix} I_{t-1}^{m1} \\ I_{t-1}^{m2} \\ I_{t-1}^{m3} \end{pmatrix} + (c) * \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \beta_{IP}..\beta_m..0..0..0..0 \\ 0..0..\beta_{IP}..\beta_m..0..0 \\ 0..0..0..0..\beta_{IP}..\beta_m \end{pmatrix} * \begin{pmatrix} IP_t^{m1} \\ M_t^{m1} \\ IP_t^{m2} \\ M_t^{m2} \\ IP_t^{m3} \\ M_t^{m3} \end{pmatrix} + \begin{pmatrix} \eta_t^{m1} \\ \eta_t^{m2} \\ \eta_t^{m3} \end{pmatrix}$$

Where: I_t^Q = change in fixed capital formation in quarter t

I_t^{m1} = change in fixed capital formation in the first month of quarter t

I_t^{m2} = change in fixed capital formation in the second month of quarter t

I_t^{m3} = change in fixed capital formation in the third month of quarter t

$IP^{m1}{}_t$ = industrial production indicator growth rate in the first month of quarter t

$M^{m1}{}_t$ = imports indicator growth rate in the first month of quarter t

C = constant

β_{IP} = coefficient of industrial production indicator

β_M = coefficient of imports indicator

ε_t = disturbance term at t in measurement equation

η_t^{m1} = disturbance in state variable of month 1 of quarter t

In the variant A formulation, the index t refers to the quarter being decomposed, and m1 is the value in the first month of that quarter. This is a true latent variable approach, the monthly fixed capital formation indicators are purely a product of the state space formulation. The link between the related monthly indicators and fixed capital formation is introduced by requiring the average of the three computed monthly latent variables to equal the development of the standard quarterly capital formation statistic. Defining the quarterly growth rate as a simple average of the monthly growth rates is an approximation, which will result in an error. But it also greatly simplifies the computation, and the error introduced will be acceptable as long as there is not much difference between the months in the quarters. Important is that the monthly state variables are defined as having individual standard deviations and therefore separate innovations.

A slight modification of this set-up produces *variant B*, tested in this study as well.

State Space formulation Variant B

$$\begin{pmatrix} I_t^Q \\ I_t^{Q*} \end{pmatrix} = \begin{pmatrix} 0 \; \tfrac{1}{3} \; \tfrac{1}{3} \; \tfrac{1}{3} \\ 1...0...0...0 \end{pmatrix} * \begin{pmatrix} I_t^Q \\ I_t^{m1} \\ I_t^{m2} \\ I_t^{m3} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} I_t^Q \\ I_t^{m1} \\ I_t^{m2} \\ I_t^{m3} \end{pmatrix} = \begin{pmatrix} 0...0...0...0 \\ 0...0...0...0 \\ 0...0...0...0 \\ 0...0...0...0 \end{pmatrix} * \begin{pmatrix} I_{t-1}^Q \\ I_{t-1}^{m1} \\ I_{t-1}^{m2} \\ I_{t-1}^{m3} \end{pmatrix} + (c) * \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \beta_{IP}..\beta_m..0..0..0..0..0..0 \\ 0..0..\beta_{IP}..\beta_m..0..0..0..0 \\ 0..0..0..0..\beta_{IP}..\beta_m..0..0 \\ 0..0..0..0..0..0..\beta_{IP}..\beta_m \end{pmatrix} * \begin{pmatrix} IP_t^{Q1} \\ M_t^Q \\ IP_t^{m1} \\ M_t^{m1} \\ IP_t^{m2} \\ M_t^{m2} \\ IP_t^{m3} \\ M_t^{m3} \end{pmatrix} + \begin{pmatrix} \eta_t^Q \\ \eta_t^{m1} \\ \eta_t^{m2} \\ \eta_t^{m3} \end{pmatrix}$$

Where: I_t^Q = change in fixed capital formation in quarter t

I_t^{Q*} = auxiliary fixed capital formation variable, necessary to estimate the model

IP^Q_t = industrial production indicator growth rate in quarter t

M^Q_t = imports indicator growth rate in quarter t

C = constant

β_{IP} = coefficient of industrial production indicator

β_M = coefficient of imports indicator

ε_t = disturbance term at t in measurement equation for monthly variables

η_t^Q = disturbance in state equation for quarterly growth rate at t

The modification lies in defining a state vector consisting of the observed quarterly growth rate in fixed capital formation and the three unobserved monthly growth rates. The state equation then contains respectively quarterly growth rates in industrial production and imports and the corresponding monthly growth rates. In this formulation the relationship between target variable and related indicators at a quarterly frequency is used to aid the estimation of the monthly equations.

Other approaches, not used here, will define the y_t as consisting of both the high- and low-frequency target variables, with the high-frequency ones inserted as missing values (e.g. four missing quarterly values and an observed annual one) to be estimated by the model.

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.

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 gebruik	*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 time series 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

The flexibility of the state space framework allows for many different modelling strategies. This section will report on the ones tested in this study and their results. First, some basic options on how to construct the models were tested. The bulk of the results concerns different strategies for using the related indicators. Finally, a few more complex state space modelling strategies were tried out.

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.

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

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.

After defining the basic structure of the state space model, as done in section 3, there are three different approaches for estimating the coefficients of the monthly indicators in the model. Clar et al. (1998) advise using coefficients estimated beforehand in a OLS-model, using lower-frequency data. This reduces the computational burden of the state space model by reducing the number of parameters to be estimated in the state space step. This approach was tested in this study, but yielded relatively large deviations from the reference series realisations, and was thus rejected. Using these estimates to produce starting values for the parameters was useful though. The second approach, using different coefficients for the indicators in each state/transition equation, was problematic as well. In effect, this means estimating three separate equations, one for each month. The resulting increase in the number of degrees of freedom was too great, the models did not converge. Overall, the clearly superior approach was to estimate the coefficients in the state space estimation, but restricting the coefficients of the related indicators to be equal in all state equations, i.e. for all months in the quarter.

The first set of results is for the so-called endogenous models, which only contain import and industrial production indicators. In the next section, other related indicators are added. This is done for state space models of variant A, where the state vector consists only of the three unobserved monthly growth rate of fixed capital formation. After this, the results of the autoregressive variant are reported, and finally the results on variant B. In that variant the quarterly growth rates are added to the state and indicator vectors.

5.1 Endogenous models, variant A

In this step of model construction there are several possibilities for entering the information available from the related monthly indicators. 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. On the other hand, this can result in overidentified or inefficient models. As mentioned before, the composition of the national accounts indicators on private fixed capital formation was used to make a first selection of relevant indicators. Further information on the most relevant indicators was gained by constructing OLS models for capital formation at the quarterly level.

Table 5.1; Composition of state space models Variant A, industrial production and import indicators at different levels of aggregation.

State Space Model	*Industrial production indicators*	*Level of aggregation*	*Imports indicators*	*Level of aggregation*
Model Low A	Construction, Electronics, Machines, transport equipment	2-digit	M71, M72, M73, M74, M75, M76, M77, M78, M79, M87	2-digit
Model Low B	Construction, Electronics, Machines, transport equipment	2-digit	M71, M75, M76, M77, M87	2-digit
Model Low C	Construction	2-digit	M75	2-digit
Model Middle D	Construction, electronics	2-digit	M7,M8	1-digit
Model High E	IP total	1-digit	M7	1-digit
Model Low F	Construction, Electronics, Machines, transport equipment	2-digit	-	-

State space models at different levels of aggregation were constructed and tested. The best performing ones are described in the table 5.1. The preliminary simple regressions indicated that the low-aggregation level indicators contain the most information on capital formation. This is logical as not all components of imports and industrial production are equally relevant for fixed capital formation. Using high aggregation indicators does enable the construction of parsimonious models. As the results in table 5.2 show, this advantage does not outweigh the superior accuracy of the low aggregation indicators. On the one hand, these results indicate that it is possible to measure monthly capital formation via the latent variable approach. The in-sample (filtered) errors at 1.83%-point show that the outcomes are close enough to the original series to be credible.

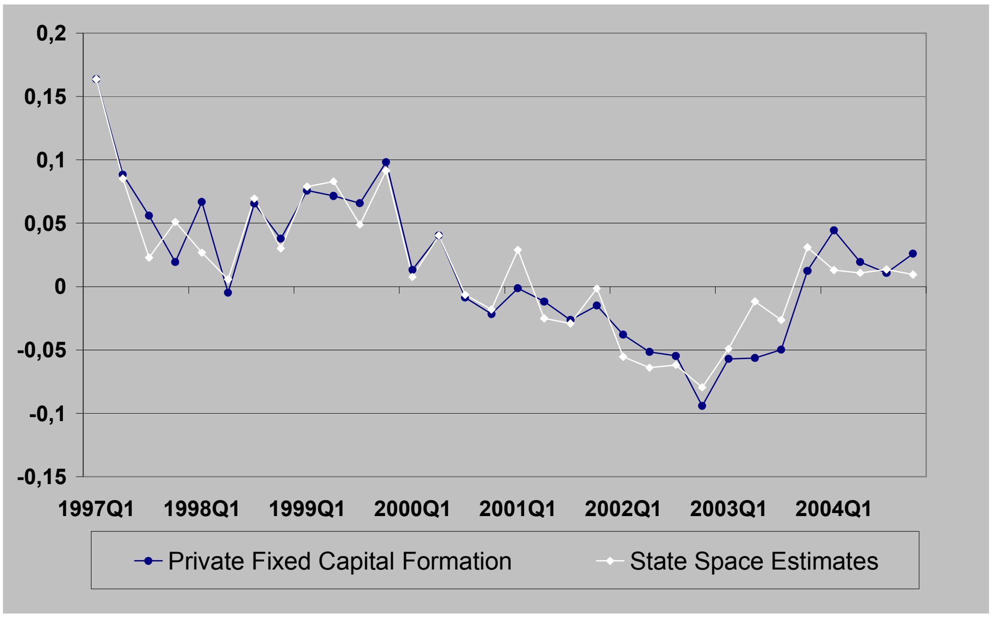
Table 5.2; Estimation results for models (variant A) consisting of industrial production and import indicators at different levels of aggregation. Reference is quarterly realisations of growth rate of private fixed capital formation.

Model	*RMSE filtered (%-points)*	*RMSE real-time (%-points)*	*Log Likeli-hood*	*AIC*	*Q-prob 4 lags*	*Jaque-Berra probability*
Model Low A	**1.83%**	**4.16%**	80.0	-3.81	0.570	0.97
Model Low B	**3.13%**	**6.09%**	63.4	-3.09	0.154	0.65
Model Low C	**3.73%**	**4.16%**	50.2	-2.7	0.005	0.73
Model Middle D	**4.17%**	**7.46%**	46.7	-2.36	0.057	0.97
Model High E	**4.11%**	**6.17%**	44.2	-2.33	0.014	0.97
Model Low F	**5.05%**	**7.76%**	47.1	-2.12	0.000	0.71

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. However, the out-of-sample real-time simulations show that these formulations will perform quite badly when used to produce actual statistics. This cast doubts on the approach tested here. In the following sections, it is tested whether adding other indicators and modifying the state-space set-up can improve results.

Here, using all relevant monthly indicators of production and imports (model A low) gives the best results. The in-sample filtered error is acceptable at 1.83 %-points, though far from negligible. This model also has the highest accuracy in the real-time estimate, but this is still an unacceptable 4.16%-points. The residual statistics for the two best performing indicator sets are good, indicating no misspecification. The calculated quarterly averages of the constructed monthly indicators from model low A 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 for model A low 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 reasonably well in-sample, except for the occasional larger deviation. The state space models based solely on the production and imports indicators can be compared to the first step of measuring fixed capital formation in the Quarterly National Accounts system. This is before additional information on capital formation and the state of the economy as a whole is added. Anchoring the state space model on the standard quarterly growth rates introduces part of these corrections, but not sufficiently so.

5.2 Exogenous models, variant A

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. Other Statistics Netherlands research 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. Introducing these related indicators in the state space models is bound to improve accuracy. Several of the most successful set-ups of the previous step were enhanced with the additional indicators, as shown in table 5.3.

Table 5.3; Composition of state space models Variant A, additional indicators next to industrial production and import indicators at different levels of aggregation.

State Space Model	*Industrial production indicators*	*Imports indicators*	*Exogenous indicators*	*Level of aggregation*
Model exo G	Construction, Electronics, Machines, transport equipment	M71, M72, M73, M74, M75, M76, M77, M78, M79, M87	Capacity Utilization (change, lag 1 quarter), 10-year bond yield(change, lag 2 quarters)	low
Model exo H	Construction, Machines, transport equipment	M71, M75, M76, M77, M87	Capacity Utilization, 10-year bond yield	low
Model exo I	Construction, Machines, transport equipment	M71, M75, M76, M77, M87	Capacity Utilization, 10-year bond yield, VAT IT-services, BS employment IT-services	low
Model exo J	Construction, Machines, transport equipment	M71, M75, M76, M77, M87	Capacity Utilization, 10-year bond yield, BS employment IT-services	low
Model exo K	Construction, Machines, transport equipment	M71, M75, M76, M77, M87	Capacity Utilization, 0-year bond yield, VAT IT-services	low
Model exo L	Construction, Machines, transport equipment	M71, M75, M76, M77, M87	VAT IT-services	low

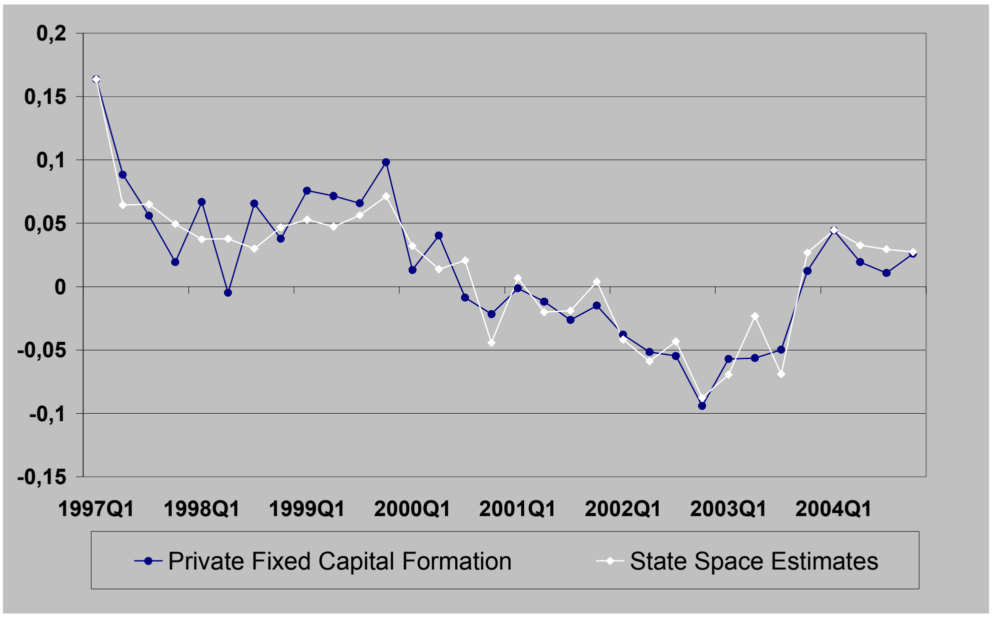
The results in the table below show that introducing additional information on economic conditions via these related indicators improves results in all cases. Again, the best results are achieved by including many or all relevant indicators, and at a low level of aggregation**.** As expected, the state space estimates now are better at tracing the actual realisations, but only for the in-sample results.

Table 5.4; Estimation results for models (variant A) consisting of additional indicators next to industrial production and import indicators at different levels of aggregation. 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*
Model exo G	**1.47%**	**4.88%**	86.4	-4.09	0.002	0.56
Model exo H	**2.03%**	**3.81%**	76.6	-3.85	0.017	0.48
Model exo I	**0.41%**	**na**	57.5 (different sample)	-4.77	0.303	0.868
Model exo J	**0.76%**	**na**	47.6 (different sample)	-3.72	0.778	0.109
Model exo K	**0.59%**	**na**	51.8 (different sample)	-4.21	0.283	0.096
Model exo L	**0.75%**	**na**	47.8 (different sample)	-3.98	0.821	0.193

The in-sample deviations from the reference quarterly growth rates are strongly reduced, by half a percentage point to an acceptable 1.47%-point for model exo G and to an impressive 0.41%-point for model exo I. The improved fit is also visible in the smaller AIC values, though there is evidence of serial correlation in the residuals, indicating some form of misspecification. The improvement is however not visible in the real-time simulation results. These have hardly improved, the best one being an unacceptable 3.81%-points for model exo H. It is also unwise to focus here on the impressive performance of the models containing the VAT-data, as the monthly availability of these data is not certain. Another disadvantage of using the IT business survey- and VAT-indicators is the limited amount of data available, which makes a real-time simulation pointless here. Given the results of the other models, it is likely that the in-sample results of the VAT-containing models cannot be replicated in real-time conditions. Therefore, I will first further analyze the best performing set-up without the VAT-data, model exo H. It has the lowest real-time errors, and is therefore to be preferred to the model exo G. This last one performs better in-sample, but is probably overfitted.

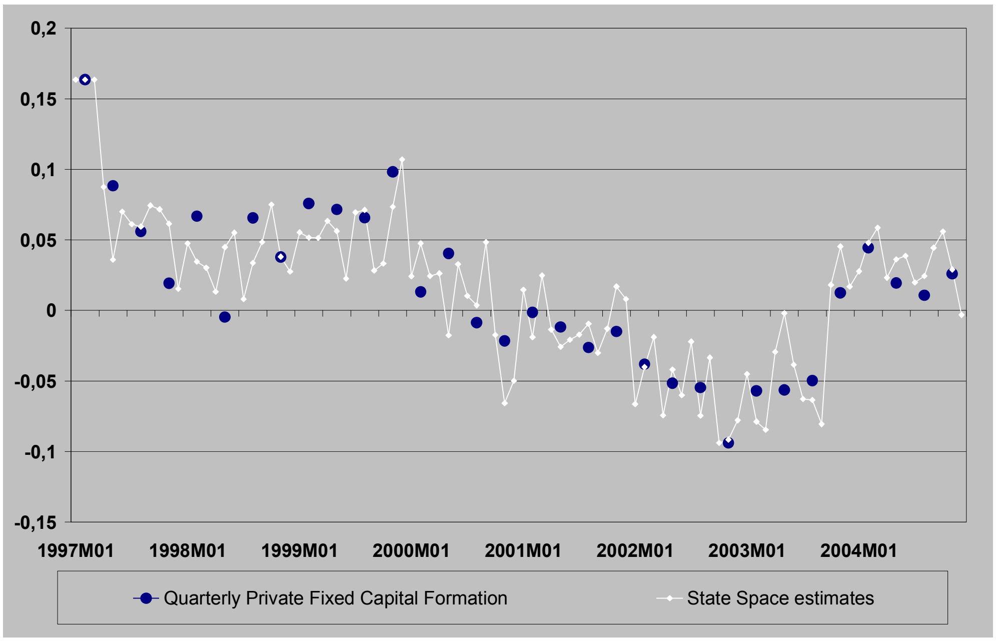
Graph 5.2; Synthetic quarterly year-on-year growth rates of model exo H compared to actual realisations of the change in private fixed capital formation from the National Accounts.



As can be seen in graph 5.2, the in-sample synthetic growth rates from model exo H track the quarterly realisations reasonably well, especially in the second half of the sample. This set-up contains selected import and industrial production indicators and the indicators of capacity utilization and long-term interest rated. Residual correlation statistics are somewhat problematic here, indicating that this specification is not perfect. Ignoring for a moment all obvious problems, the conclusions are that the additional indicators contain valuable information and that this methodology is able to derive a credible indicator of fixed capital formation.

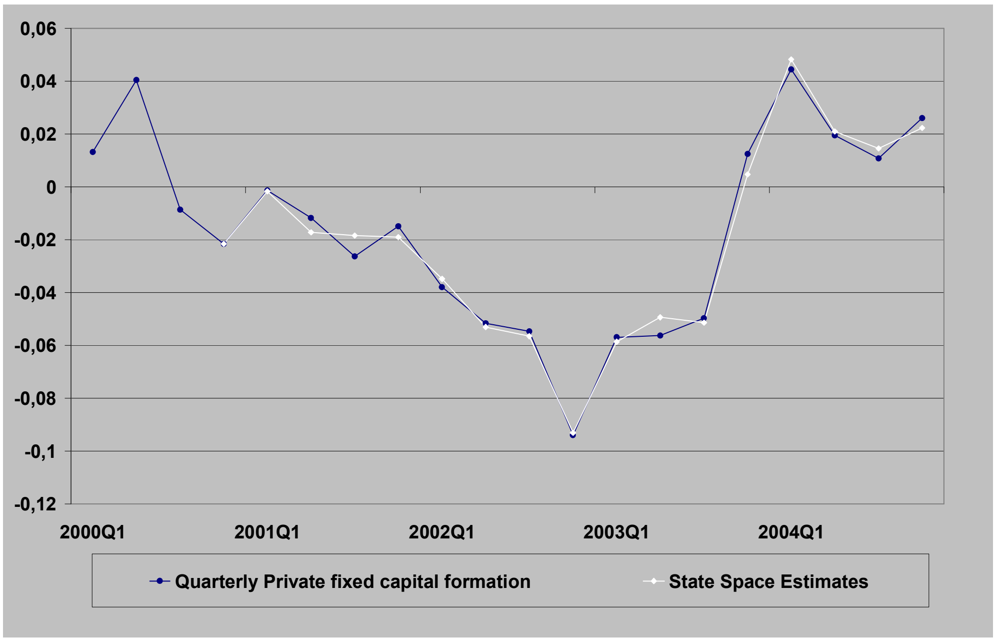
In graph 5.3 the corresponding monthly estimates are shown, next to the quarterly realisations. This shows what the monthly capital formation statistic could look like. The monthly statistic is somewhat more volatile than the quarterly National Accounts measure. But higher frequency statistics are usually more volatile than lower frequency quarterly or yearly statistics. Also, the increased volatility originates in the source statistics, which are more volatile than the quarterly capital formation growth rates. In appendix II, 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. The higher volatility of the monthly source indicators can be seen to be translated into the state space estimates.

Graph 5.3; Synthetic monthly year-on-year growth rates of model exo G compared to quarterly realisations of the change in private fixed capital formation from the National Accounts.



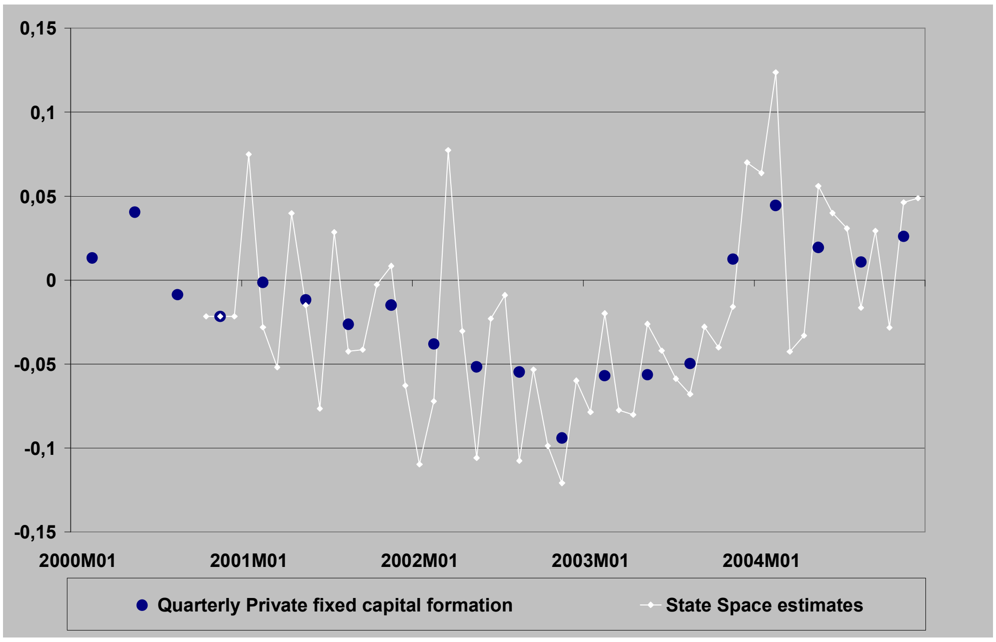
Another interesting result from this section is how well the set-ups containing the IT-sector business survey and VAT-data perform, though evaluating these remains difficult. The best performing one is model exo I, containing all additional indicators. The resulting synthetic quarterly growth rate estimates track the actual quarterly realisations very well, see graph 5.4.

Graph 5.4; Synthetic quarterly year-on-year growth rates of model exo I compared to actual realisations of the change in private fixed capital formation from the National Accounts.



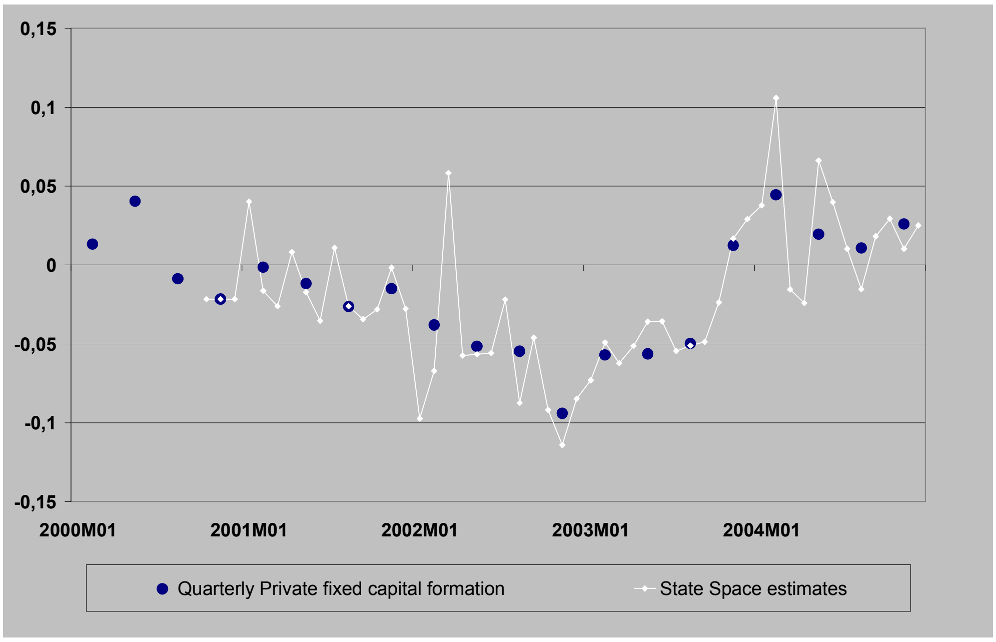
Deviations are small and the patterns match very well. The corresponding monthly estimates are slightly volatile, see graph 5.5 (note the reduced period under consideration).

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



On this front, set-up exo K which lacks the IT-sector business survey data but does contain the VAT-data, performs better. Its fit is somewhat inferior to the previously considered set-up containing all additional indicators, but its monthly evolution, depicted in graph 5.6, is more smooth and therefore credible.

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



This set-up contains a selection of import- and industrial production indicators, a bond yield indicator, an indicator of capacity utilization and finally the VAT-data of the IT-sector. It possesses a very good in-sample fit and a very plausible monthly evolution. The analysis of the residuals also indicates no problems. A disadvantage is that it requires VAT-data, whose availability is uncertain. If this should become a problem, set-up exo J is available, which only adds the IT-sector business survey data. It also performs very well, with low errors and good statistical properties.

5.3 Autoregressive specification

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. It was tested what the effects were of introducing in the state space models a simple autoregressive component next to the monthly indicators. The autoregressive component links to the value of each state space vector to it realisation in the previous period, i.e. the previous quarter (see section 3):

$$sv_t^i = c(1) + c(2) * sv_{t-1}^i + ...$$

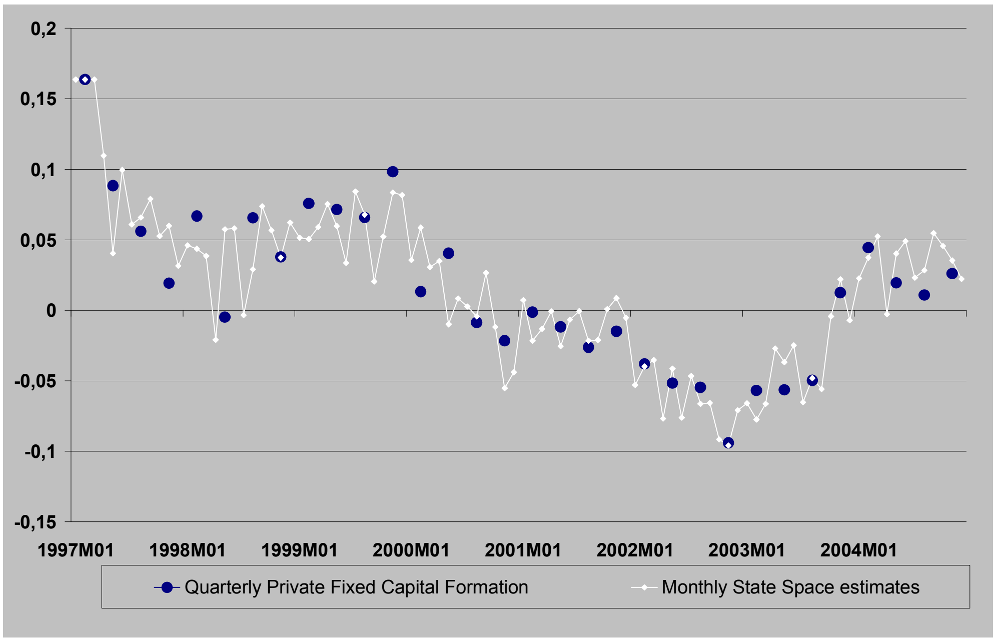
The resulting models seem to perform quite well, at least when considering the filtered estimates see table 5.5. Also, defining a separate autoregressive process for each monthly state vector proved to be superior to using the previous' quarters realisation in the autoregressive specification.

Table 5.5; Estimation results for models (variant A) enhanced with an autoregressive component. 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*
Model M Laexoallesar1	**0%**	**na**	79.8	-3.61	0.002	0.79
Model N Laendoallesar1	**0%**	**4.02%**	78.1	-3.63	0.021	0.33
Model O Laexo2ar1	**1.8%**	**3.14%**	72.6	-3.61	0.004	0.59

For models M and N, the filtered, in-sample estimates give a perfect fit with the quarterly realisations. However, the real-time errors indicate that this performance will probably not be attainable in practice. The good results are probably due to the good in-sample qualities of the methods used. Also, the filtered monthly estimates of these two autoregressive models are more volatile than the monthly estimates of the previous set-ups, see appendix II. However, set-up model O has the overall best results in the real-time simulation, in spite of its relatively poor in-sample performance. Its volatility also matches that of the quarterly fixed capital formation growth rates. As is visible in graph 5.7, this means that this monthly indicator tracks the quarterly realisations relatively smoothly, resulting in a more credible monthly indicator. These results show that including a autoregressive component significantly improves the estimates. Further study into the possibilities of introducing some variant of the local linear trend model is advisable.

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



5.4 Endogenous and exogenous models, variant B

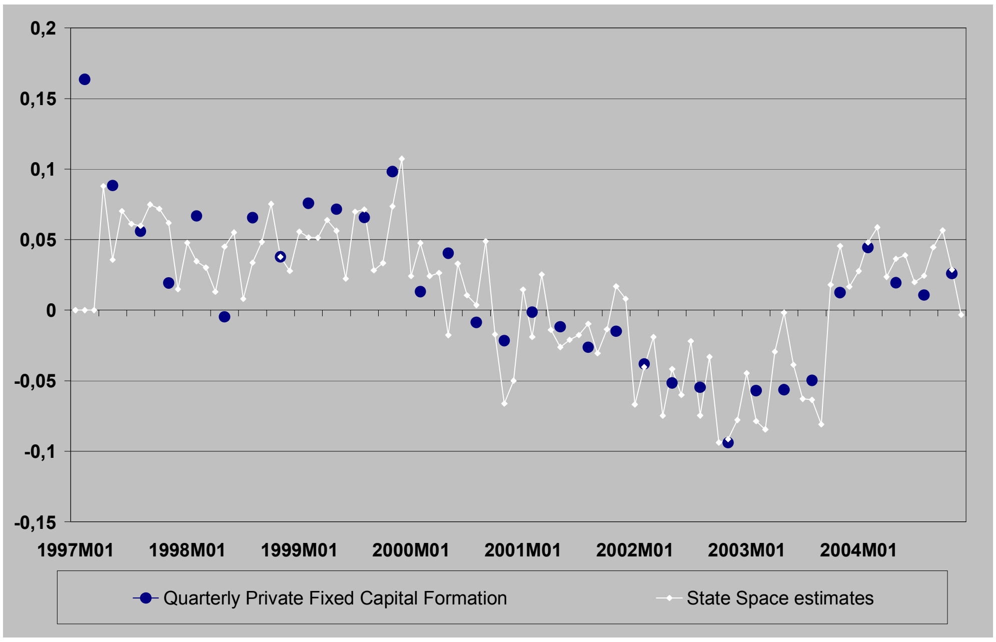
Another possibility was to construct a state space set-up where the quarterly realisations of the changes in fixed capital formation are included in the state space vector (Variant B, see section 3). In effect, this puts the monthly and quarterly observations on an equal footing, allowing the relationship between the observed quarterly realisations and the quarterly values of the related indictors to be used in estimating the monthly equations. Table 5.6 shows that this results in a perfect in-sample fit, suggesting that this is the approach to use. However, the results from the real-time simulations are disappointing here as well. An acceptable level of accuracy is not reached. This suggests that this approach is better suited for backwards extrapolation of high-frequency data than for actual monthly production of new statistics. Note that due to the different model set-up, log likelihood and AIC values cannot be compared to those of variant A. It seems probable that variant B suffers somewhat from over-fitting, which means that variant A will probably be more robust in the daily practice.

Table 5.6; Estimation results for models of variant B, both with and without additional indicators. No autoregressive components. 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*
Model exo variant B	**G**	**0%**	**na**	156.9	-8.43	0.456	0.000
Model exo variant B	**H**	**0%**	**3.94%**	137.9	-7.62	0.718	0.000
Model Low variant B	**A**	**0%**	**4.30%**	143.3	-7.07	0.603	0.000

A small experiment supports this analysis. It is possible to use the values of the parameters resulting from variant B as starting values for the estimation of the corresponding model of variant A. This results in fast convergence, but to slightly different outcomes than in the normal estimation. Now, the in-sample errors for variant A are zero as well, indicating a perfect fit. But prediction errors for variant A are now higher and log-likelihood values lower than after the normal estimation procedure, indicating an inferior fit.

Graph 5.8; Synthetic monthly year-on-year growth rates of model exo G variant B compared to quarterly realisations of the change in private fixed capital formation from the National Accounts.



An added disadvantage is the high volatility of the synthetic monthly indicators resulting from variant B, see graph 5.8 and appendix II. Volatility is higher than that of the original quarterly series and also than the volatilities of previous state space set-ups. It is of course possible to reach a correct average value in a quarter with three very large estimated monthly growth rates. All this casts doubts on the usefulness of this particular approach.

6. Conclusions

The main question which this study set out to answer is whether it is possible to use state space techniques to construct high-frequency statistics from scratch. This means that econometric techniques are used to derive the desired information from related indicators instead of constructing a formal statistical framework for the measurement. Specifically, the aim was to construct a monthly indicator for private fixed capital formation. For each quarter, three unobserved, latent, monthly capital formation indicators were defined. In the state space framework, these were anchored on the quarterly growth rates from the standard quarterly National Accounts. Information from the National Accounts was also used to select the monthly related indicators from which the monthly fixed capital formation growth rates were derived. This approach proved to be somewhat successful.

In all formulations tested, latent variables could be extracted and an at least credible monthly indicator of fixed capital formation development was produced. Accuracy varied among the specifications, but the basic latent variable approach is viable. However, only the in-sample results possessed acceptable to very good accuracies. When performing real-time simulations via a rolling regression-type procedure, deviations from the quarterly realisations increased dramatically. Even the best set-up possessed an average real-time error of over three percentage-points. This is clearly unacceptable, even for a monthly statistic which is meant to give a first impression of developments.

The elementary form of the models tested here contained just import and industrial production indicators, at the two-digit level. Low-level aggregates were better sources of information than high-level ones. These real indicators were selected based on information from the supply and use tables from the National Accounts. The estimated monthly indicators of capital formation were credible, but not very accurate. 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 dramatically improved results. Business survey data and bond yields, which reflect general economic conditions, were shown to contain important information. VAT-data on the IT-sector seemed to improve accuracy as well, though the time series available was too short for a thorough analysis.

A state space formulation containing an autoregressive component performed the best. This models also contained selected production and imports indicators and data on developments in capacity utilization and long-term interest rates The real-time deviations of this model were lowest, and the corresponding monthly indicator tracked the quarterly realisations relatively smoothly. This is important, as some formulations yielded wildly varying monthly growth rates. Monthly indicators usually are more volatile than quarterly ones, and the volatility was shown to originate in the source statistics. But a monthly indicator which resembles the evolution of the quarterly statistic more closely is to be preferred. By comparing the volatilities of the monthly indicators to that of the reference quarterly statistic, an additional quality measure for the model outcomes was created. The same goes for adding a simple autoregressive component. Possibly using a more sophisticated version of these approaches could yield better results.

Methodologically speaking, the state space models performed satisfactorily. Using the same coefficients in each monthly state equation was necessary for successful estimation, though these could be estimated in the state space model itself. A slightly more complex formulation adding variables at the quarterly frequency to the state vector was less successful. Convergence tended to be fast and as noted to credible outcomes, whilst residual properties were generally good as well. This is a little surprising, as the most successful formulations here contained quite a few indicators. Usually, state space models are constructed with as few parameters as possible to keep the estimation process manageable. This might be an explanation for the good in-sample, but poor real-time properties of the models. Another one is that the basic state space formulation tested here is just too unstable for real-time purposes.

The disappointing real-time results mean that this approach is unsuitable for use in a statistical production process. But this study has confirmed the principle of using the state space approach to construct new statistics without the need of a formal statistical framework. These econometric techniques can be used to extract new information from existing, traditional and alternative, data sources and combining this into credible statistics. If a reference statistic is present at a lower frequency, a high frequency statistic can be constructed by using related high-frequency indicators. However, an approach with better real-time properties is needed. The relatively good results of the autoregressive specification and the literature on high-frequency interpolation of statistics indicate that there may be possibilities in extending that approach. This is done in the companion study, Van Ruth (2006), and there the results are better indeed.

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), "Constructing a monthly indicator of fixed capital formation via high frequency interpolation. The state space approach" 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]:

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

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:

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

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:

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

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)***
Private 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 Low A	0.015	0.065
Model Low B	0.016	0.052
Model Low C	0.021	0.039
Model Middle D	0.027	0.045
Model High E	0.026	0.040
Model Low F	0.020	0.050
Model exo G	0.015	0.060
Model exo H	0.015	0.054
Model M	0.015	0.064
Model N	0.015	0.068
Model N	0.015	0.055
Model exo G variant	0.015	0.066
Model exo H variant B	0.015	0.063
Model Low A variant B	0.015	0.075

Indicator	*Mean of relative growth rate* ***(2000-2004)***	*Standard deviation of relative growth rate* ***(2000-2004)***
Fixed capital formation	-0.021	0.037
Model Low A	-0.018	0.054
Model exo G	-0.021	0.047
Model exo H	-0.018	0.043
Model exo I	-0.021	0.055
Model exo J	-0.021	0.046
Model exo Kc	-0.021	0.045
Model exo L	-0.021	0.057

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