# Economic forecasting: Models, indicators and data needs

# INDEX

# Economic forecasting: Models, indicators and data needs

## Peder Andersen and Thorbjørn Horn Rasmussen
## Danish Economic Council

## 1. Introduction

Economic forecasting is a difficult 'art' and a good performance demands a balanced use of different models, ad hoc indicators and a huge amount of good data. The better data available and, in general, the more developed models the less need we have for ad hoc indicators. However, economic theory should always play a significant role. On the other hand, if we rely only on economic theory and theory related empirical models significant forecast mistakes are more likely than if we combine model work and non-economic statistical work containing information based on high frequency macroeconomic indicators. Furthermore, many other sources of information are relevant such as survey data.

In this paper we look at the benefit and the shortcoming of using the traditional large-scale macroeconomic model and how these shortcomings can be reduced if a large-scale macroeconomic model is combined with non-economic statistical work.

## 2. The traditional large-scale macroeconomic model

Economic theory gives a good reference for developing large-scale macroeconomic models. In the Danish Economic Council such a model have been used since 1973. The model SMEC (the **S**imulation **M**odel of the **E**conomic **C**ouncil) has changed features several times as statistical methods, economic theory and data made changes relevant.

The present version of the model from 1999 is fully empirically based and contains a description of the Danish economy disaggregated into 8 sectors. The model contains some 600 equations and 1000 variables. In this model as in most of this type of models the short run production is determined by the demand but in the long run production is mainly supply driven, particularly by the supply of labour.

Key areas in the model is the input-output structure, the wage formation and the determination of the demand of inputs. Also the housing market and the consumption related to the housing market plays a significant role in the Danish model.

Estimations of the various relations are based on the national account data. In the Danish model annual data are used but in other models quarterly data are used. In both cases it is relevant to update key variables so adjustments can be made to take into account the most actual information.

Also several exogenous variables, national as well as international, are important to feed into the model before the forecast procedures can be finished. It is important to have the most updated information available and the relevant question is how we gather or produce the information or how we by using other types of models, surveys or indicators can improve the estimates of key economic variables relevant for policy makers.

### 3. Diffusion indexes

One of our latest progresses in the forecast area, still labelled as (early) work in progress, is setting up a forecast model which is based on linear diffusion indexes. The setup we use for this model follows closely James H. Stock´s and Mark W. Watson´s work on this topic. We want to give them full credit for the framework as we have adopted most of the model setup from them and applied it to Danish data.

When forecasting is based on an annual model, up to date information on the development in several variables on higher frequency is important to use. Especially when forecasting the current year at a point in time where information about part of the year is already known. Knowledge about the development through last year can also have a substantial effect on the forecasted annual growth rate. In other words, generally, an annual model needs to be supplemented by formal or informal higher frequency models when forecasting the current year=s growth rates. When an assessment of the actual state of the economy is ongoing there are some variables that typi-cally/historically are used as leading indicators. These variables are for example car purchases, industrial production, prices, wages unemployment rates etc. There is a very large literature on small-scale macroeconometric models that uses leading indicators in short term (2-1 year ahead) forecasting of specific variables. However, most of these models only include a small subset of variables that might have an influence on the variables to forecast. Therefore in practice, forecasters often find it useful to extract the information from many more series than typically are included in these small scale high frequency models. Because of this procedure the short term forecast easily ends up being based on informal methods to extract the information available and the forecast might reflect the forecaster=s insight more than the models used. Another major drawback in this procedure is that it is nearly impossible to make explicit analyses of what went wrong/right ex post.

The diffusion index makes it possible to extract the information contained in a very big number of high frequency macroeconomic indicators, in both a systematically and consistent, but a non-economic framework. One main advantage of forecast models using diffusions index is their ability to use the information from up to several hundred macroeconomic time series and to do forecasting based on these on central macroeconomic variables. The model setup with diffusion indexes even allows to incorporate explanatory variables with different frequencies and different end periods.

As the model setup is purely statistical it will be of little use in understanding the interdependences between macroeconomic variables and therefore at least the model must have better forecast performance than traditionally indicator models. Evaluation of this model is therefore primarily based on a comparison of forecasts found by using different competing forecast models.

**The model**
In the following a short introduction to the model is given. A much more detailed description can be found in Stock and Watson (1998). Our presentation will focus on the factor structure, the forecast model, data and some results.

**Factor structure**

The factor structure captures the co-movement in macroeconomic time series and can be represented in terms of a statistical factor model such as;

$$\textbf{(1)} \qquad\qquad X_t = \Lambda F_t + \boldsymbol{m}_t$$

where $X_t$ is a matric with N time series variables that contains useful information in forecasting the variable of interest. Notify that in general $X_t$ will also contain the variable that is going to be forecasted. $F_t$ are the common factors that correspond to the principal components in $X_t$, and $\Lambda$ is the coefficients (factorloadings). In the present approach $\Lambda$ is assumed to be time invariant, but in a more general setting it can be time dependent. The disturbance $\mu_t$ is generally correlated across time and series. Our data set contains 169 series based on monthly frequency and 90 on quarterly frequency (N=259).

The factor model explains the co-movement in $X_t$ based on a small number of *k* common factors. These types of models are however not new[1] and have been used in traditional indicator models[2]. The EM (Expectation Maximization) algorithm is used to estimate the common factors ($F_t$) and the factor loadings ($\Lambda$). The standard principal component analysis does not apply or is at least infeasible as we are using an unbalanced data panel characterised by missing observations, different frequencies and series that are available over shorter time spans. The last observation date is usually not the same among the time series, due to different publication/collection procedures. In practise this results in a data set where some time series only have observations up to three or four months before the most updated variable. Some variables get published very quickly as for example the interest rate, exchange rates etc. There is a tradeoff in setting the end period, when estimating the common factors. In forecasting there is an obvious desire to use all the newest information - but setting an end period where only a very limited number of variables have observations results in a weaker estimation of the common factors. As a compromise, which deserves further analyse, we require that there are common end period observations for at least 1/2 of the variables.

**Forecast model**

The statistical model is used to predict the growth rate in variables of interest for example in unemployment, GDP or inflation. The variable of interest is denoted $y_t$ and our goal is to estimate the growth rate in $y_t$ - 1, 3, 6 or 12 month ahead given the information contained in the common factors $F_t$ and allowing for autoregresive lags of $y_t$. The forecast model has the following general specification for a forecast horizon on 12 months;

$$\textbf{(2)} \qquad \log\left(\frac{y_{t+12}}{y_t}\right) = \boldsymbol{d}_0 + \sum_k \boldsymbol{j}_k F_{k,t} + \sum_{i=0}^{j} \boldsymbol{b}_i \log\left(\frac{y_{t-i}}{y_{t-1-i}}\right) + \boldsymbol{e}_t$$

---

[1]Sargent and Sims (1977), Geweke (1977), Stock and Watson 1998

[2]Stock and Watson 1989 discuss the index of NBER=s index of coincident of leading indicators using a model much like 1.1.

Where $E(e_{t+1}|\{y_{t-i},X_{t-i},F_{t-i}\}_{i=1}^{-4} = 0$. For specific assumptions used for asymptotic analysis see Stock and Watson (1998). A two-step procedure is used to select the number of common factors and the lags of $y_t$. Information criteria are used in this model selection. Firstly, we find which common factors that should be included in the model. In this step no lags of the endogenous variable are included in the estimation. Secondly, the number of lags are determined. Stock and Watson provides sufficient conditions under which an information criterion such as BIC [3] or AIC[4] consistently estimates the numbers of factors and produces an asymptotically efficient forecast. The exact procedure used by Stock and Watson implies that the model is selected by recursively estimation of the information criteria for the first common factor then for the second and the third common factor and so on. The resulting model then consists of a sequence of common factors starting with the first common factor. In our setup we also use an information criterion in choosing the number of common factors, but we allow the model to consist of common factors that are not necessarily in a following sequence. Our approach is similar to a General to Specific procedure based on an information criterium (AIC). In practice this means hundreds of thousands of regressions for each model selection. In the present model we constrain our self to testing only the first twenty common factors.

**Data**
A large number of Danish data series have been applied to the model. So far the model setup has been theoretically not economic. However, in the selection of data categories economic theory is the guideline. The applied categories follow Stock and Watson=s setup. Some categories are however added to reflect that this model is based on data from a small open economy. The data set is grouped in nine categories, which are

Output and Income
Employment and hours
Consumption
Investments
Real inventories and inventory-sales ratios
Prices and wages
Foreign trade
Interest rates
Exchange rates and stock prices

Each group contains approximately 25 time series. All series are transformed to achieve stationary time series.

An analysis of the $k$ common factors reveals that groups of variables such as real output, employment, prices and investments are more or less represented by specific common factors. As the common factors only are identified up to a $k$x$k$ matrix, it is not warranted to push this analysis too far. However, to get an idea about the character of the common factors we have shown the development in two common factors[5] and the development in unemployment and net consumer prices (net of taxes and duties CPI). A more systematically analysis of the interpretation of the common factors are not yet done.

---

[3]Sawa=s Baysian information criterium

[4] Aike=s information criterion

[5] The presented common factors are linear transformed in this presentation.
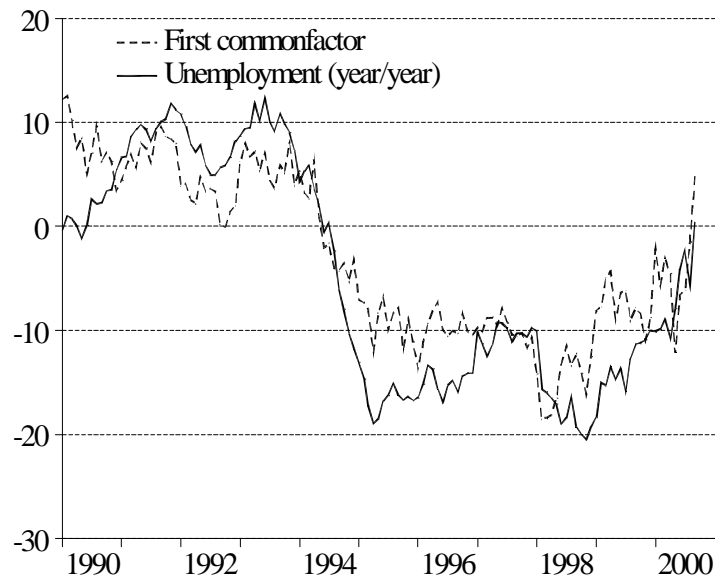
*Figure 1 First common factor*



*Figure 2 Fourth common factor*
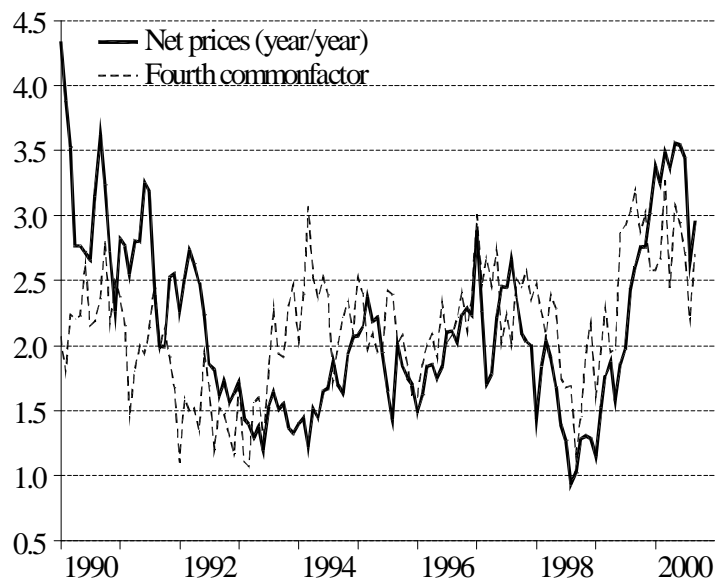


**Results and further work**

Our works are still concentrated on making the model applicable and several issues have to be dealt with before a fully integrated version will be incorporated in our standard forecast procedure. So far we have only investigated the forecast performance based on - in sample - forecasts. It is our intention to compare forecast performance of the factor model based on real time out of sample forecasts.

The forecast performance is measured here by the mean square error (MSE) relatively to a univariate autoregresive process. The univariate autoregressive forecast model has the following specification, based on a 12-month horizon.

**(3)**
$$\log\left(\frac{y_{t+12}}{y_t}\right) = d_0 + \sum_{i=0}^{j} b_i \log\left(\frac{y_{t-i}}{y_{t-1-i}}\right) + e_t$$

This model is used as a first approach, to compare the forecast results from the diffusionsindexs-model. It should be notified that other, more intelligent, models should be compared to the diffusionsindex model before any superiority can be claimed. Some results from our model applied on Danish data are shown below. These results are based on forecast models that all have a horizon of 12 months. The diffusions indexes model reduces the mean square error, in these models, with approximately 20-45 pct. compared to a univariat autoregressiv forecast model. In other words there seems to be valuable information in the common factors when forecasting. Comparisons of forecast results from other models such as leading indicators, structural VAR=s etc. are however more appropriate in determining the gain of using leading indicators. This work remains to be done with our data set. Stock and Watson show some really impressing results where the linear diffusions index model outperforms several competing models.

*Table 1 Forecast performance*

| Model | Diffusions index | | | Autoregressive[1] | Relative MSE[4] |
|---|---|---|---|---|---|
| | # factors[2] | # lags[3] | Adj. $R^2$ | Adj. $R^2$ | |
| Unemployment total | 8 | 0 | 0,69 | 0,44 | 0,54 |
| Unemployment construction | 4 | 0 | 0,45 | 0,14 | 0,63 |
| Unemployment building sector | 8 | 0 | 0,39 | 0 | 0,58 |
| Netprice indeks | 10 | 0 | 0,65 | 0,37 | 0,62 |
| Engros prices | 4 | 0 | 0,21 | 0 | 0,8 |
| Merchandise turnover | 9 | 0 | 0,35 | 0 | 0,7 |
| Industri turnover | 5 | 1 | 0,21 | 0,02 | 0,78 |

1) Autoregressive univariate model, with five lags of the dependent variable.
2) Number of common factors (F).
3) Number of lagged dependent variables.
4) Means Square Error (MSE) of the diffusion index model divided by the MSE of the autoregressive model.

So far the presented results from the diffusions index model are based on - in sample- estimation of the growth rate in different variables of interest. As an example we will now show the procedure used in forecasting -out of sample- and some results. The date of forecast is November 13, 2000. Our objective is to estimate the growth rate in net consumer prices up to 12 months ahead. The data set on a quarterly frequency has at this time observations up to second quarter of 2000. The data on a monthly frequency has a few observations for October, but most has September as the last observation date. September is chosen as last observation date for estimation of the common factors. The applied EM- algorithm is used to fill out any missing observations. This involves that time series with end periods before September 2000 are assigned estimated values by the EM algorithm. In the next step the common factors are estimated. To show the development through the next 12 months, it is necessary to set up 12 forecast models that have horizons corresponding to 1 - 12 months. For each of the horizons, the previously described selection method, chooses the Abest@ model - that is the relevant common factors and the number of lags. The forecast model with a horizon on 12 months is shown in equation (2). This equation estimates the yearly growth rate 12 months ahead. In table 2 is shown the 12 models (for the respective 12 horizons) and the entering explanatory variables in these models. The models only slightly differ in what explanatory variables that are included in these models. As an illustration of the models for a forecasting horizon of 1, 2, 11 and 12 months can be seen in the appendix. The forecast from the 12 forecast models, and the corresponding forecasts 1 - 12 months ahead, are linked together and presented in figure 3. The diffusions indexes model predicts a lower inflation, measured by the net consumer priceindex in the coming months. Notify that before September 2000, the estimated -in sample - values is derived from the model with a horizon of 12 months.
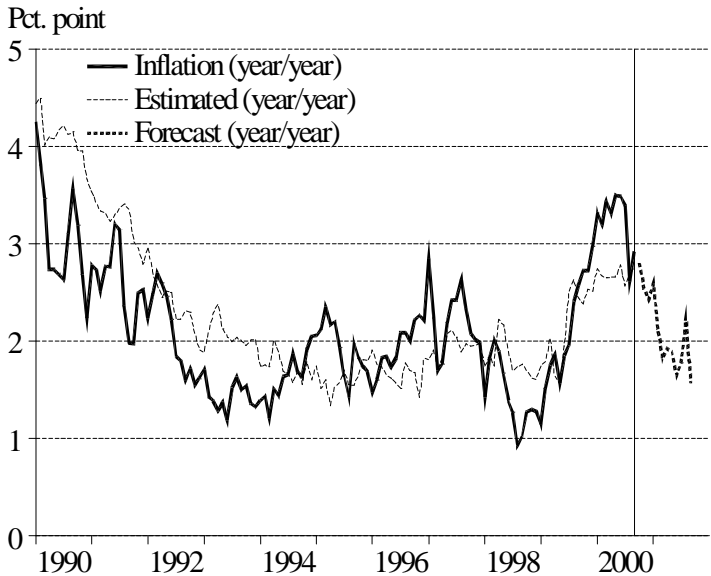
*Figure 3 Net consumer prices*



7

*Table 2 Forcastmodels for net consumer prices with 12 different horizons*

| Explanatory variables | Model with a horizon on # months: | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| F1 | x | x | x | x | x | x | x | x | x | x | x | x |
| F2 | x |   | x | x | x | x | x | x | x | x | x | x |
| F3 |   | x |   | x | x | x | x | x | x | x | x | x |
| F4 | x | x | x | x | x | x | x | x | x | x | x | x |
| F5 |   |   |   |   |   |   |   |   |   |   |   |   |
| F6 |   |   |   | x | x | x | x | x | x | x | x | x |
| F7 |   |   |   |   |   |   |   |   |   |   |   |   |
| F8 |   |   | x | x | x | x | x | x | x | x | x | x |
| F9 | x | x | x | x | x | x | x | x | x | x | x | x |
| F10 |   |   |   |   |   |   |   |   |   |   |   |   |
| F11 |   |   | x | x |   |   |   |   |   |   |   |   |
| F12 |   |   |   |   |   |   |   |   |   |   |   |   |
| F13 |   |   |   |   |   |   |   |   |   |   |   |   |
| F14 |   | x | x | x | x | x | x | x | x | x | x | x |
| F15 |   |   | x |   |   |   |   |   |   |   |   |   |
| F16 | x |   | x | x |   |   |   |   |   |   |   |   |
| F17 | x | x | x | x | x | x | x | x | x | x | x | x |
| F18 |   |   |   |   |   |   |   |   |   |   |   |   |
| F19 |   | x |   | x | x | x | x | x | x | x | x | x |
| F20 |   |   |   | x | x | x | x | x | x |   |   |   |
| #Lag(y) | 1 |   |   | 1 |   |   |   |   |   |   |   |   |

F<i> is common factor number <i>

## 4. Concluding remarks

The diffusion index models used here are still being worked on. There are still many open questions about how to handle specific practical procedure in making forecast within this framework. The results -so far- however indicates that there is valuable information in using diffusion indexes in forecasting macroeconomics variables.

There will newer be one single or simple way to proceed when economic forecasting is performed. A combination of using large-scale models, small-scale models and various ways of using short

term indicators, including surveys, will, in our opinion, always be around and to a certain degree compete but also supplement each other.

The demand for data will increase and at the same time an increasing focus on quality, comparability and accessibility. However, if the use of diffusion indexes becomes common these demands will be even more prevailing.

## Literature

Det Økonomiske Råds Sekretariat (The Secretariat of the Danish Economic Council). 1999. SMEC (Simulation Model of the Danish Economic Council). *Arbejdspapir 1999:7 (Working Paper 1999:7)*

Geweke, J. 1977. The Dynamic Factor Analysis of Economic Time Series. In D. J. Aigner and A. S. Goldberger (eds.). *Latent Variables in Socio-Economic Models.* North-Holland, Amsterdam

Sargent, T. J. and C. A. Sims. 1977. Business Cycle Modeling without Pretending to have too much a-priori Economi Theory. In C. Sims et al. (eds.). *New Methods in Business Cycle Research.* Federal Reserve Bank of Minneapolis.

Stock, J. H. and M. W. Watson. 1998. Diffusion Indexes. *NBER Working Paper 6702.* www.nber.org/paper/w6702

Stock, J. H. and M. W. Watson. 1998. Diffusion Indexes, Revised June 1999. www.wws.princeton.edu/~mwatson/publi.html

Watson, M. W. 2000. Macroeconomic Forecasting Using Many Predictors. *Working Paper, July 2000 (Revised August 2000).* www.wws.princeton.edu/~mwatson/html.wp

Lahiri, K. and Moore, G. H.1991. Leading economic indicators. Cambridge University Press.

# Appendix
**Estimated forecastmodels for net consumer prices with 4 different horizons**.

$y_t$ are the net consumer priceindex

f<i> are the commonfactor number <i>

## Model with a 12 months horizon
```
 Ordinary Least Squares
 MONTHLY data for  153 periods from  JAN 1988 to  SEP 2000
```

$\log(y_{t+12}/y_t)$
```
    =  - 0.00455 * f1 + 0.00291 * f2 + 0.00196 * f3
         (8.08153)        (2.91884)        (2.58279)

       - 0.00612 * f4 - 0.00303 * f6 + 0.00298 * f8
         (8.53154)        (3.89994)        (3.54816)

       + 0.00415 * f9 + 0.00185 * f17 - 0.00136 * f19
         (5.53227)        (3.03992)        (2.34600)

       + 0.00197 * f14 + 0.02630
         (2.73200)        (38.1762)

 Sum Sq      0.0073   Std Err     0.0072   LHS Mean   0.0257
 R Sq        0.6761   R Bar Sq    0.6533   F 10,142   29.6423
 D.W.( 1)    0.6272   D.W.(12)    1.7247
```

## Model with a 11 months horizon
```
Ordinary Least Squares
 MONTHLY data for  154 periods from  DEC 1987 to  SEP 2000
```

$\log(y_{t+11}/y_t)$
```
   =  - 0.00446 * f1 + 0.00337 * f2 + 0.00143 * f3
         (8.63210)        (3.70531)        (2.06621)

      - 0.00558 * f4 - 0.00254 * f6 + 0.00223 * f8
        (8.48555)        (3.55737)        (2.89523)

      + 0.00380 * f9 + 0.00157 * f17 - 0.00124 * f19
        (5.52660)        (2.80378)        (2.34330)

      + 0.00198 * f14 + 0.02440
        (3.00919)        (38.7682)

 Sum Sq      0.0062   Std Err     0.0066   LHS Mean   0.0237
 R Sq        0.6788   R Bar Sq    0.6563   F 10,143   30.2148
 D.W.( 1)    0.5919   D.W.(12)    1.7124
```

## Model with a 2 months horizon
Ordinary Least Squares
 MONTHLY data for  163 periods from  MAR 1987 to  SEP 2000

$\log(y_{t+2}/y_t)$
   =  - 0.00093 * f1 + 0.00063 * f3 - 0.00123 * f4
      (4.90774)         (2.77840)          (5.13800)

      + 0.00116 * f9 - 0.00054 * f19 + 0.00067 * f17
      (4.67040)         (2.67625)          (3.22974)

      + 0.00050 * f14 + 0.00423
      (2.01378)          (20.3980)

```
 Sum Sq      0.0010    Std Err    0.0026    LHS Mean    0.0044
 R Sq        0.3926    R Bar Sq   0.3652    F  7,155  14.3115
 D.W.( 1)    1.1928    D.W.(12)   2.1266
```

## Model with a 1 month horizon
 Ordinary Least Squares
 MONTHLY data for  162 periods from  APR 1987 to  SEP 2000

$\log(y_{t+1}/y_t)$
   =  - 0.00067 * f1 + 0.00071 * f2 - 0.00076 * f4
      (5.11757)         (3.55288)          (4.78190)

      + 0.00045 * f9 - 0.00019 * f16 + 0.00049 * f17
      (2.57961)         (1.30883)          (3.66148)

      - 0.14018 * $\log (y_t/y_{t-1})$ + 0.00259
      (1.96272)              (11.6288)

```
 Sum Sq      0.0004    Std Err    0.0017    LHS Mean    0.0022
 R Sq        0.3349    R Bar Sq   0.3046    F  7,154  11.0763
 D.W.( 1)    2.0789    D.W.(12)   2.2162
```