Forecasting Germany’s printing and writing paper imports

Lauri Hetemäki, FFRI (METLA)
Jarmo Mikkola, FFRI (METLA)

Contents

I  Background
II  Motivation
III Models an evaluation criterion
IV  Empirical results
V  Conclusions
I Background

• *Case study* - Short term statistical forecasting
• Updated and extended version of the earlier SSFE paper presented in Gilleleje (2002)
• *Finnish Forest Sector Economic Outlook* (Yearly publication, export & production forecasts for six quarters)
• Germany is one of the Finnish paper industry’s biggest export countries

II Motivation

• In *2002 study* the forecasting model was chosen purely on the basis of *in the sample* performance
  → All models failed to handle the structural break, which changed the seasonal dynamics just before the forecast horizon
  → Over fitting?
• *Our focus of interest here;* What kind of econometrical time series model (if any) would be good for forecasting this kind of time series?
II Motivation (Continues)

• Lately it has been emphasized that the actual *purpose of an empirical model* should have a bigger role in the modeling process
  • See for example Granger (1999) *Empirical modeling in Economics*
  • Forecasting model;
    → Model evaluation with the *out of sample* forecasting ability!

III Models and evaluation criterion

• 1 Variables
• 2 Models and model families
• 3 Forecast combinations
• 4 Model evaluation criterion
III.1 Variables

1. Coated printing and writing paper imports to Germany (in tons)
2. Unit price of pulp
3. German gross domestic product
4. German industrial sales volume index
5. German industrial orders volume index
6. Combined activity index \([(3+4+5)/3]\)

III.2 Models and model families

i) Univariate models
   - *Deterministic trend models* (Trend, seasonal dummies etc.)
   - *Smoothing models* (Holt, Brown, exponential etc.)
   - *ARIMA models*
   - *Univariate random coefficient model* (Harvey, Kalman filter)

ii) Single equation multivariate models
   - *Autoregressive dynamic lag models* (ADL)
   - *Multivariate random coefficient models*
III.2 Models and model families

iii) Multivariate system models (VAR models)
   • VAR(\( p \)), DVAR(\( p \))
   • VARX(\( p, s \)), DVARX(\( p, s \))
   • BVAR(\( p, \lambda, \theta \))

iv) Forecast combinations
   • Average forecasts
   • Optimally combined forecasts
     • Model weights optimized to minimize RMSE
     • Weights constant through the forecast horizon (6 quarters)

III.3 Forecast combinations

1 Average combination forecasts (\( k \) models)
   \[ AF_t = \frac{1}{k} \sum_{i=1}^{k} F_{ti}, \quad t = 1, \ldots, 6 \]

2 Optimal combination forecasts
   \[ CF_t = \sum_{i=1}^{k} w_i F_{ti}; \quad \sum_{i=1}^{k} w_i = 1, \quad t = 1, \ldots, 6 \]
   where weights are obtained by minimizing the function
   \[ RMSE = \sqrt{\frac{1}{6} \sum_{i=1}^{6} (CF_t - A_t)^2} \]

\( k \) is number of models, \( F_{ti} \) is model’s \( i \) forecast for time period \( t \) and \( A_t \) is the observed value for period \( t \).
III.4 Model evaluation criterion

General model evaluation criterion was

\[ RMSE = \sqrt{\sum_{t=1}^{6} \frac{1}{6} (F_t - A_t)^2} \]

where \( F_t \) is a forecast and \( A_t \) an observed value for period \( t; t = 1, 2, ..., 6 \)

Note. Diversity of the models limited the general usability of the criterion

- The criterion was used in specification searches, when it was feasible

IV Empirical results

1 Descriptive analysis
2 Estimation results
   2.1 Best models
   2.2 Model family comparisons
   2.3 Combinations
IV.2 Estimation results

1. Best models
2. Model family comparisons
3. Combinations
IV.2.1 Best models

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time trend (II)</td>
<td>0.099</td>
</tr>
<tr>
<td>Brown exponential smooth. (I)</td>
<td>0.094</td>
</tr>
<tr>
<td>ADL (I)</td>
<td>0.085</td>
</tr>
<tr>
<td>Multiv. RC model (II)</td>
<td>0.082</td>
</tr>
<tr>
<td>VARX(3,2) (I)</td>
<td>0.070</td>
</tr>
<tr>
<td>DVAR(1,7) (II)</td>
<td>0.066</td>
</tr>
<tr>
<td>Average combination (I-III)</td>
<td>0.088</td>
</tr>
<tr>
<td>Optimal combination (I-III)</td>
<td>0.062</td>
</tr>
</tbody>
</table>

Note. Sub-sample number in the brackets.
Combination forecasts

Earlier forecasts (Gilleleje 2002)
IV.2.2 Model family comparisons

Forecasts generally improved when models got more complicated!

<table>
<thead>
<tr>
<th>Model family (Best model)</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Univariate models (Brown smoothing)</td>
<td>0.094</td>
</tr>
<tr>
<td>Single eq. multivariate models (MvRC)</td>
<td>0.082</td>
</tr>
<tr>
<td>Multivariate system models (DVARX)</td>
<td>0.066</td>
</tr>
<tr>
<td>Forecast combinations (Opt. combin.)</td>
<td>0.062</td>
</tr>
</tbody>
</table>

IV.2.3 Forecast combinations

- System models dominated combinations; The optimal combination of all the best models included only VARX model (14%) and DVARX model (86%)
- Optimization improved the combination forecast RMSE from 0.088 (0.066) to 0.062
  - Average; 15 models
  - Optimal; 2 models
V Conclusions

1 Out of sample evaluation is recommendable especially in short term forecasting!
   • Earlier study with in the sample evaluation led to bad short term forecasts!
   • Here statistically rather ’non-significant’ system models gave best forecasts!

2 It seems worthwhile to combine forecasts, if the structure of the process or system isn’t clear.
   • Intuitively it might be reasonable to use diverse models when combining forecasts
   • Without more precise knowledge about the ’optimal’ weighting structure, simple averages is one solution

V Conclusions (Continues)

3 Modeling differences might be one way to proceed when forecasting time series, that seem to have structural breaks

4 System modeling (VAR modeling) can provide good forecasts even, when descriptive analysis doesn’t suggest it’s use or when the models don’t look statistically that significant

5 Bayesian extension doesn’t necessarily improve the VAR forecasts
V Conclusions (Continues)

6 Starting from a scratch, this kind of statistical forecasting exercise is much too laborious to be useful in practise!
   • In order to save time the process for this kind of forecasting approach should be as automatic as possible

The end

(Below extra material!)
III Practical matters

- 1 General features of estimation
- 2 Variables
  - 2.1 Variable selection
  - 2.2 Final model variables
- 3 Models and model families
- 4 Model evaluation and specification
- 5 Technical tools

III.1 General features of estimation

- Case study
  - Single series
- Small sample
  - Observations in the whole sample 42 + 6
  - Observations in sub-samples: II; 34 + 6 and III; 16 + 6 obs.
- Pragmatic approach;
  - Theoretical considerations in the background
  - Parameter estimate significance tested only, when it was necessary for the model specification
  - Model diagnostics in a minor role!
III.2.1 Variables selection process

i) Potential explanatory variables (Chosen using economics theory and expert knowledge)
   - Unit price (Pulp price)
   - Activity variables (GDP, sales, orders)
   - Product demand variables

ii) Variables in final models (Gave best forecasts)
   - Imports
   - Activity variables (GDP, Sales, Orders and a combination of these)

III.4 Model evaluation and specification

2) Specification within the model families
i) Univariate models
   - In the sample estimations with roughly 30 different models per series
   - Model comparisons with RMSE

ii) Single equation multivariate models
   - ’General to specific’
   - Model comparisons and variable selection with RMSE, when it was possible
III.4 Model selection and specification

iii) VAR models
   • Lag length selections with RMSE (Macro)
   • Choice of the exogenous variables with RMSE

iv) Forecast combinations
   • RMSE as quality measure of combinations
   • RMSE also the optimization criterion in the selection of the optimal model combination weights (Macro)

III.5 Technical tools

i) Softwares
   • SAS/ETS 8.2
   • Eviews 4.1
   • PcGive 9.10

ii) Macro for VAR lag length selection
   • SAS macro that estimated models and printed the forecast RMSE (for given parameter intervals)
     • $p$ in VAR($p$) model
     • $p$ and $s$ in VARX($p,s$) model
     • $\theta$ and $\lambda$ for BVAR($p, \theta, \lambda$) model
IV.1 Descriptive results

- i) Trend structures
- ii) Seasonal fluctuations
- iii) Stationarity
- iv) Structural breaks
- v) Unit roots
- vi) Variable relationships

The first difference of the imports series
Desc. characteristics - Imports

- Local time-trend like behaviour
- The series seems to have several breaking points, both in trend and seasonal dimension (1Q1993, 3Q1997 and 2Q1999). These were also confirmed by the Chow test
- Possible seasonal behaviour seems very irregular and couldn’t be modeled by stable seasonal dummy model for the whole sample
- Small significant autocorrelation up to the seasonal lag four (in levels)
Desc. characteristics – Imports (Cont.)

- The general irregularity of the series makes the unit root testing difficult
- Testing difference stationarity against the trend stationarity, the usual unit root tests (PP, ADF) indicate trend stationarity

Pulp price
First difference – pulp price

Descriptives (Pulp price)

- No seasonal behaviour
- No trend-like structures
- No distinct seasonal variation
Descriptives (GDP)

- The most deterministic of the model series;
  - Clear time trend
  - The seasonal variation seems pretty stable
  - The $R^2$ series with time trend and seasonal dummies is 95.6%
Industry sales volume index

Industrial new orders index
Combined activity index

- Sales, orders, combined ind.
  - Quite clear shift downwards in the third quarter of 1993
  - Upward-trend like behaviour after the shift
  - No clear seasonal fluctuations
Variable relations

- Correlations
  - Some significant correlations between the imports and activity variable but not between the imports and price variable

- Integration and co-integration
  - Mixed results
  - System testing (Johansen) indicated to the possible co-integration between the imports and indices
  - Both differences and levels included in the modeling

- No pairwise relationship seemed beforehand clearly useful in the forecasting of the imports!

IV.2.2 Model family comparisons

- Introduction of Bayesian extension in the VAR model or random coefficient models didn’t improve the out of the sample forecasts

- The irregularity of the imports series is confirmed by the fact that smoothing models and plain time trend gave best forecasts among the univariate models
  → No deterministic structures were found significant!

- Using differences improved forecasts with VAR and VARX models
IV.2.2 Model family comparisons

- Extensive model search was impossible with the multivariate single equation models due to the huge number of model combinations
- Combination optimization led to the inclusion of only a couple of best models (VARX, DVARX)

V Conclusions (Continues)

- Starting from a scratch, this kind of statistical forecasting exercise is much too laborious to be useful in practise!
  - In order to save time the procedure for this kind of forecasting approach should be as automatic as possible
    - One should have particular groups of models ready for the series with different characteristics!
    - Out of sample evaluation should be made with standard approach
      - Pseudo-horizon evaluation + whole sample estimation?
    - There should be standard approach to combining forecasts
      - Averages etc.