A look into the factor model black box

Publication lags and the role of hard and soft data in forecasting GDP

Marta Bańbura and Gerhard Rünstler

Directorate General Research
European Central Bank

November 2006

PRESENTATION OUTLINE

ECB uses DFM by Doz et al. (2005)

Forecasting from unbalanced monthly data

Integration of interpolation and forecasting

Good forecasting performance

Statistics to assess role of individual series

PRESENTATION OUTLINE

Unbalanced data & dynamic model

- Kalman filter weights
- Uncertainty measures

Application: hard & soft data in forecasting euro area GDP

Publication lags matter a lot

MODEL: DFM

$$
\begin{aligned}
x_t &= \Lambda f_t + \xi_t, & \xi_t &\sim \mathbb{N}(0, \Sigma_\xi), \\
f_t &= \sum_{i=1}^{p} A_i f_{t-i} + \zeta_t, & & \\
\zeta_t &= B\eta_t, & \eta_t &\sim \mathbb{N}(0, I_q).
\end{aligned}
$$

MODEL: interpolation

Forecast for 3-month growth in GDP $\widehat{y}_t^{(3)}$

$$\begin{aligned} y_t^Q &= \frac{1}{3}(y_t^{(3)} + y_{t-1}^{(3)} + y_{t-2}^{(3)}) \\ y_t^{(3)} &= y_t + y_{t-1} + y_{t-2} \end{aligned}$$

Evaluated only in 3^{rd} month of the quarter

Fcst equation for monthly growth rates y_t

$$y_{t+1}^{(3)} = \mu + \lambda' f_{t+1} + \varepsilon_{t+1}^{(3)}, \qquad \varepsilon_{t+1}^{(3)} \sim \mathbb{N}(0, \Sigma_\varepsilon^{(3)})$$

MODEL: State space form

$$\begin{bmatrix} x_t \\ y_t^Q \end{bmatrix} = \begin{bmatrix} \Lambda & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_t \\ y_t^{(3)} \\ Q_t \end{bmatrix} + \begin{bmatrix} \xi_t \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} I_r & 0 & 0 \\ -\lambda' & 1 & 0 \\ 0 & -\frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} f_{t+1} \\ \widehat{y}_{t+1}^{(3)} \\ Q_{t+1} \end{bmatrix} = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Xi_{t+1} \end{bmatrix} \begin{bmatrix} f_t \\ y_t^{(3)} \\ Q_t \end{bmatrix} + \begin{bmatrix} \zeta_{t+1} \\ \varepsilon_t^{(3)} \\ 0 \end{bmatrix}$$

MODEL: interpolation

Forecast for monthly GDP $\widehat{y}_t$

$$
\begin{aligned}
y_t^Q &= \frac{1}{3}(y_t^{(3)} + y_{t-1}^{(3)} + y_{t-2}^{(3)}) \\
y_t^{(3)} &= y_t + y_{t-1} + y_{t-2}
\end{aligned}
$$

Evaluated only in 3^{rd} month of the quarter

Fcst equation for monthly growth rates y_t

$$y_{t+1} = \mu + \lambda' f_{t+1} + \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim \mathbb{N}(0, \Sigma_\varepsilon)$$

MODEL: Kalman filter & smoother

For state space form

$$
\begin{aligned}
z_t &= W\alpha_t + u_t, & u_t &\sim N(0, \Sigma_u) \\
\alpha_{t+1} &= T_t \alpha_t + v_t, & v_t &\sim N(0, \Sigma_v),
\end{aligned}
$$

and any unbalanced data $\mathcal{Z}_t$ the KF provides

$$
\begin{aligned}
a_{t+h|t}^{-j} &= \mathbb{E}\left[\alpha_{t+h} | \mathcal{Z}_t\right] \\
P_{t+h|t}^{-j} &= \mathbb{E}\left[a_{t+h|t} - \alpha_{t+h}\right]\left[a_{t+h|t} - \alpha_{t+h}\right]',
\end{aligned}
$$

Euro area data set

Real activity		32	
	Industrial production		6 weeks
	Retail sales		6 weeks
	Labour market		6-8 weeks
Surveys (EC)		22	0 weeks
	Business		
	Consumer		
	Retail & construction		
Financial data		22	0 weeks
	Exchange & interest rates		
	Stock price indices		
	Other		

FORECAST EVALUATION

	Example Q2	Real activity	Surveys	Financial
	(Oct)			
	(Nov)			
	(Dec)			
1	Jan			
2	Feb			
3	Mar			
4	Apr			
5	May			
6	Jun			
7	**Jul**			

FORECAST EVALUATION

	Example Q2	Real data	Surveys	Financial
	(Oct)			
	(Nov)			
	(Dec)			
1	Jan			
2	Feb			
3	Mar			
4	Apr			
5	May			
6	**Jun**			
7	Jul			

FORECAST EVALUATION

	Example Q2	Real data	Surveys	Financial
	(Oct)			
	(Nov)			
	(Dec)			
1	Jan			
2	Feb			
3	Mar			
4	Apr			
5	**May**			
6	Jun			
7	Jul			

FORECAST EVALUATION

	Example Q2	Real data	Surveys	Financial
	(Oct)			
	(Nov)			
	(Dec)			
1	**Jan**			
2	Feb			
3	Mar			
4	Apr			
5	May			
6	Jun			
7	Jul			

FORECAST EVALUATION

Forecast performance (2000 Q1 - 2006 Q2)

Fcst Nr	**Example Q2**	AR(1)	QVAR	Bridge eqs	DFM
1	Jan	.82	.82	.84	.70
2	Feb	.82	.82	.85	.72
3	Mar	.82	.82	.88	.74
4	Apr	.98	.98	.87	.73
5	Jun	.98	.98	.89	.73
6	Jul	.98	.98	.93	.80
7	Aug	1.03	1.05	.95	.81

INDIVIDUAL SERIES: Forecast weights

Express $\widehat{y}^{Q}_{t+h|t}$ as the weighted sum of observations in $\mathcal{Z}_t$ (Harvey and Koopman, 2003)

$$\widehat{y}^{Q}_{t+h|t} = \sum_{k=0}^{t-1} \omega_{k,t}(h) z_{t-k} \ ,$$

Weights are time-invariant for our definition of $\mathcal{Z}_t$.

- Cumulative forecast weights $\sum_{k=0}^{t-1} \omega_{k,i}(h)$ for series i,
- Historical contributions of series i to the forecast

INDIVIDUAL SERIES: Uncertainty measures

Define *subsets of indicators* $x_t = (x_t^{1\prime}, x_t^{2\prime}, x_t^{3\prime})'$

Form data $\mathcal{Z}_t^{-j}$: all observations of x_t^j eliminated

Consider difference in precision from data $\mathcal{Z}_t$ and $\mathcal{Z}_t^{-j}$

Advantage: no re-estimation (maintain original factor loadings)!

INDIVIDUAL SERIES: Uncertainty measures

	Example Q2	Real data	Surveys	Financial
	(Oct)			
	(Nov)			
	(Dec)			
1	Jan			
2	Feb			
3	Mar			
4	Apr			
5	**May**			
6	Jun			
7	Jul			

INDIVIDUAL SERIES: Uncertainty measures

a) Filter uncertainty (Giannone et al., 2005)

$$\mathrm{var}(\widehat{y}^{Q,-j}_{t+h|t} - y^{Q}_{t+h}) = \pi^{-j}_{t+h|t} + \sigma^2_{\varepsilon} \ ,$$

σ^2_{ε} is residual uncertainty

$\pi^{-j}_{t+h|t}$ is uncertainty from $f^{-j}_{t+h|t}$.(from $P^{-j}_{t+h|t}$)

b) RMSE from recursive forecasts

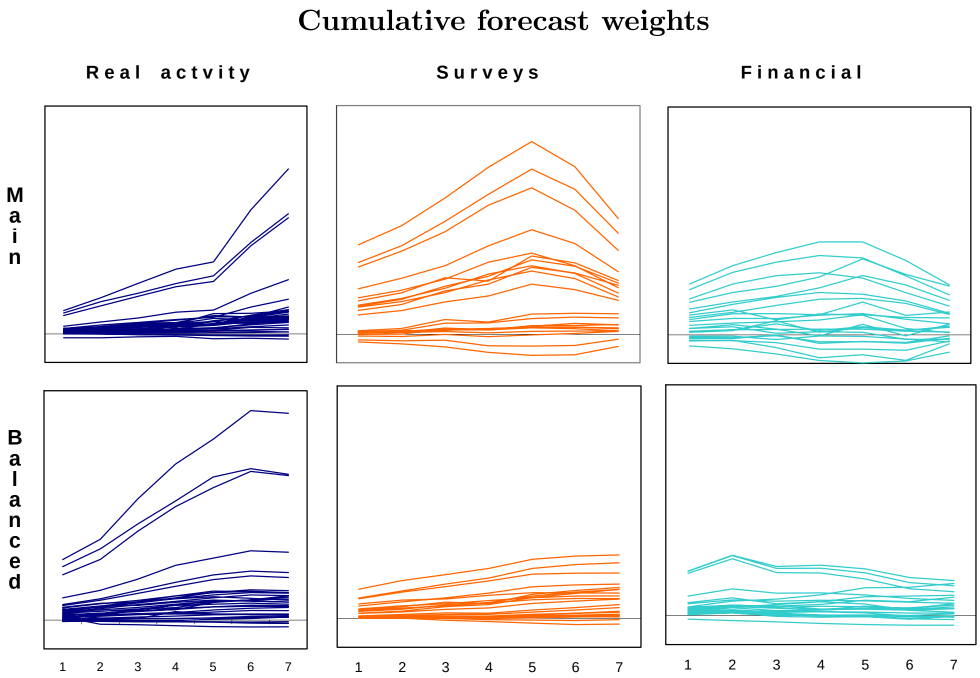
APPLICATION

Data downloaded on 30, June 2006

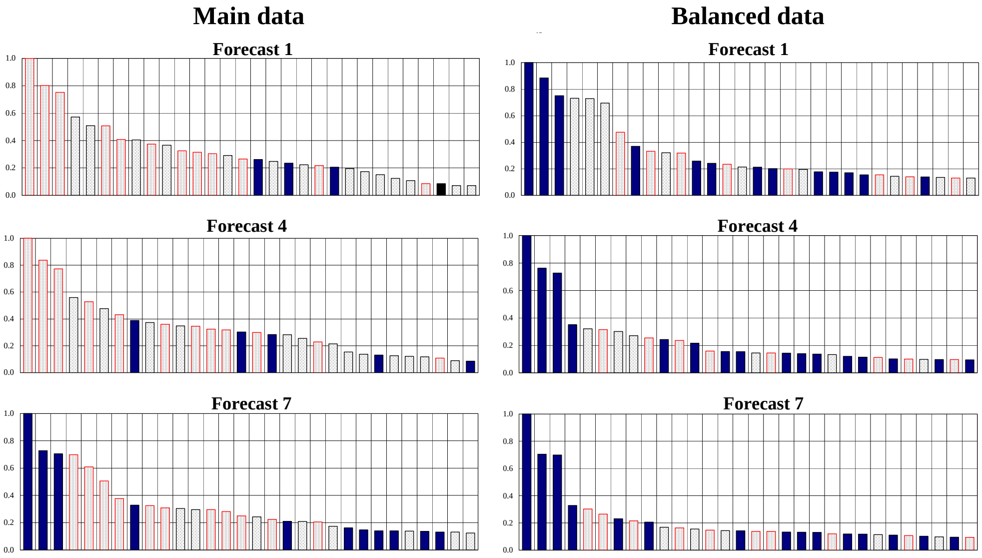
Pseudo real-time design

2 data sets

- Main (original publication lags)
- Balanced (w/o publication lags in real data)


Cumulative forecast weights
Real actvity
Surveys
Financial
Main
Balanced
1
2
3
4
5
6
7

Cumulative forecast weights

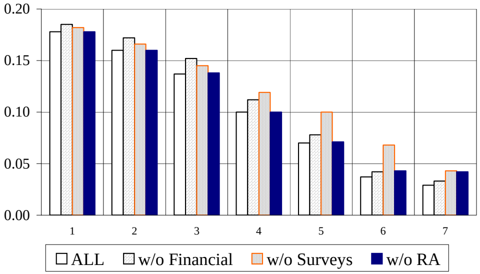


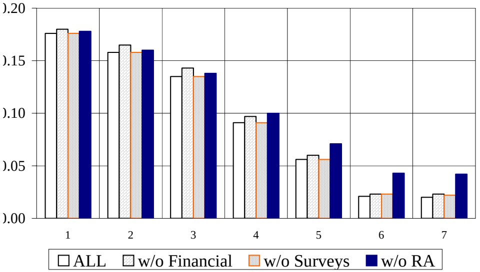
Mean absolute contributions of data groups

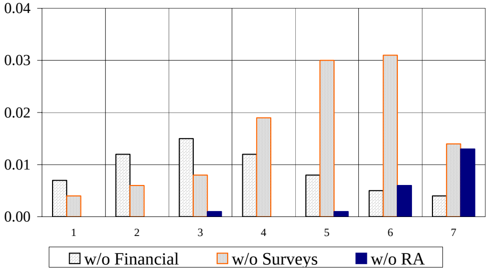
(1998 Q1 - 2005Q4)

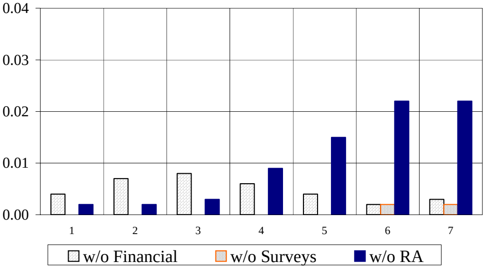
Data	Main				Balanced			
	Fcst	Contributions (%)			Fcst	Contributions (%)		
	$\mathcal{Z}$	$\mathcal{S}$	$\mathcal{F}$	$\mathcal{R}$	$\mathcal{Z}$	$\mathcal{S}$	$\mathcal{F}$	$\mathcal{R}$
1	0.158	60 %	57 %	14 %	0.135	34 %	46 %	46 %
2	0.183	58 %	57 %	15 %	0.163	32 %	46 %	44 %
3	0.196	61 %	56 %	16 %	0.192	28 %	42 %	49 %
4	0.227	62 %	50 %	16 %	0.188	34 %	40 %	48 %
5	0.245	63 %	42 %	17 %	0.199	35 %	35 %	47 %
6	0.230	61 %	37 %	25 %	0.206	35 %	29 %	52 %
7	0.210	53 %	35 %	32 %	0.200	37 %	29 %	50 %

Filter uncertainty









Filter uncertainty

(Full-sample parameter estimates)

	Main				Balanced			
	$\mathcal{Z}$	$\mathcal{RS}$	$\mathcal{RF}$	$\mathcal{SF}$	$\mathcal{Z}$	$\mathcal{RS}$	$\mathcal{RF}$	$\mathcal{SF}$
1	.178	.185	.182	178	.176	.180	.176	178
2	.160	.172	.166	.160	.158	.165	.158	.160
3	.137	.152	.145	.138	.135	.143	.135	.138
4	.100	.112	.119	.100	.091	.097	.091	.100
5	.070	.078	.100	.071	.056	.060	.056	.071
6	.037	.042	.068	043	.021	.023	.023	.043
7	.029	.033	.043	042	.020	.023	.022	.042

Table 3: RMSE from recursive forecasts

(1998 Q1 - 2005Q4)

	AR	Main				Balanced			
		$\mathcal{Z}$	$\mathcal{RS}$	$\mathcal{RF}$	$\mathcal{SF}$	$\mathcal{Z}$	$\mathcal{RS}$	$\mathcal{RF}$	$\mathcal{SF}$
1	.38	.33	37	.32	.33	.33	.35	.32	.33
2	.35	.32	.36	.31	.32	.31	.33	.31	.32
3	.35	.28	.33	.29	.28	.28	.30	.29	.28
4	.35	.28	.30	.30	.28	.26	.27	.26	.28
5	.31	.28	31	.29	.28	.25	.26	.24	.28
6	.31	.25	.28	.27	.27	.24	.25	.24	.27
7	.31	.24	.25	.24	.27	.23	.24	.23	.27

CONCLUSIONS

Investigating the role of individual series in a DFM

- Based on Kalman filter
- Deal with unbalanced data sets
- Avoid re-estimation of parameters
- Usage for selection of series

CONCLUSIONS

Hard & soft data: differences in publication lags matter a lot!

- Surveys are close substitutes to real data
- Financial data provide complementary information

Balanced data give a grossly wrong picture