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

mbanbura@ulb.ac.be
Université Libre de Bruxelles

gerhard.ruenstler@ecb.int
Directorate General Research
European Central Bank

November 2006

Introduction

Forecasting GDP from monthly data: large unbalanced data sets

Dynamic factor models (DFMs) have become a common tool

One drawback is the black box nature of DFMs

- Which series drive the forecast?
- How to handle unbalancedness?
- What are the sources of forecast revisions?

This paper

We use a DFM by Doz et al. (2005)

Statistics to assess role of individual series

- Kalman filter weights
- Uncertainty measures

Hard & soft data in forecasting euro area GDP

Important role for surveys once unbalancedness is accounted for

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

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

  In the 3^{rd} month of each quarter, evaluate the fcst for quarterly GDP growth, $\widehat{y_t^Q}$

$$
\begin{aligned}
\widehat{y}_t &= \beta' f_t \\
\widehat{y_t^Q} &= \frac{1}{3}(\widehat{y}_t + \widehat{y}_{t-1} + \widehat{y}_{t-2})
\end{aligned}
$$

$$\begin{bmatrix} x_t \\ y_t^Q \end{bmatrix} = \begin{bmatrix} \Lambda & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_t \\ \widehat{y}_t \\ \widehat{y}_t^Q \end{bmatrix} + \begin{bmatrix} \xi_t \\ \varepsilon_t^Q \end{bmatrix}$$

$$\begin{bmatrix} I_r & 0 & 0 \\ -\beta' & 1 & 0 \\ 0 & -\frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} f_{t+1} \\ \widehat{y}_{t+1} \\ \widehat{y}_{t+1}^Q \end{bmatrix} = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Xi_{t+1} \end{bmatrix} \begin{bmatrix} f_t \\ \widehat{y}_t \\ \widehat{y}_t^Q \end{bmatrix} + \begin{bmatrix} \zeta_{t+1} \\ 0 \\ 0 \end{bmatrix}$$

Kalman filter & smoother

For state space form

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

and any data $\mathcal{Z}_t^{-j}$ the KF provides MMSE estimates of α_{t+h},

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

Flexible handling of missing observations

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		

Recursive pseudo real-time forecasts

(End-of-month)

	Example Q2	Real data	Surveys	Financial
Q(-2) M1	(Oct)			
Q(-2) M2	(Nov)			
Q(-2) M3	(Dec)			
Q(-1) M1	Jan			
Q(-1) M2	Feb			
Q(-1) M3	Mar			
Q(0) M1	Apr			
Q(0) M2	May			
Q(0) M3	Jun			
Q(+1) M1	**Jul**			
Q(+2) M2	Aug			

Recursive pseudo real-time forecasts

(End-of-month)

	Example Q2	Real data	Surveys	Financial
Q(-2) M1				
Q(-2) M2				
Q(-2) M3				
Q(-1) M1	Jan			
Q(-1) M2	Feb			
Q(-1) M3	Mar			
Q(0) M1	Apr			
Q(0) M2	May			
Q(0) M3	**Jun**			
Q(+1) M1	Jul			
Q(+2) M2	Aug			

Recursive pseudo real-time forecasts
(End-of-month)

	Example Q2	Real data	Surveys	Financial
Q(-2) M1				
Q(-2) M2				
Q(-2) M3				
Q(-1) M1	Jan			
Q(-1) M2	Feb			
Q(-1) M3	Mar			
Q(0) M1	**Apr**			
Q(0) M2	May			
Q(0) M3	Jun			
Q(+1) M1	Jul			
Q(+2) M2	Aug			

Recursive pseudo real-time forecasts

(End-of-month)

	Example Q2	Real data	Surveys	Financial
Q(-2) M1				
Q(-2) M2				
Q(-2) M3				
Q(-1) M1	**Jan**			
Q(-1) M2	Feb			
Q(-1) M3	Mar			
Q(0) M1	Apr			
Q(0) M2	May			
Q(0) M3	Jun			
Q(+1) M1	Jul			
Q(+2) M2	Aug			

Forecast weights

Express forecasts$\widehat{y}^{Q}_{t+h|t}$ as the weighted sum of available observations in $\mathcal{Z}_t$

Algorithm by Harvey and Koopman (2003) to calculate weights

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

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

Inspect

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

Uncertainty measures

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

x_t^{-j} : all observations of x_t^j eliminated

$\mathcal{Z}_t^{-j} = \left\{ x_s^{-j} \right\}_{s=1}^{t}$

Marginal contribution to forecast precision of x_t^{-j} :

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

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

Recursive pseudo real-time forecasts

(End-of-month)

	Example Q2	Real data	Surveys	Financial
Q(-2) M1				
Q(-2) M2				
Q(-2) M3				
Q(-1) M1	Jan			
Q(-1) M2	Feb			
Q(-1) M3	Mar			
Q(0) M1	Apr			
Q(0) M2	**May**			
Q(0) M3	Jun			
Q(+1) M1	Jul			
Q(+2) M2	Aug			

Filter uncertainty (Giannone et al., 2005)

Decompose the variance of the forecast error for y^Q_{t+h}

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

- σ^2_ε is residual uncertainty
- $\pi^{-j}_{t+h|t}$is the uncertainty from $f^{-j}_{t+h|t}$.

  Our SSF: $\pi^{-j}_{t+h|t}$ is obtained from $P^{-j}_{t+h|t}$.

Inspect increase in $\pi^{-j}_{t+h|t}$ against $\pi_{t+h|t}$.

RMSE

Filter uncertainty measure ignores parameter uncertainty

Consider out-of-sample forecasts based on recursive parameter estimates

But: same parameters for all data $\mathcal{Z}_t^{-j}$.

Pseudo realtime forecasts

Data downloaded on 30, June 2006

We apply publication lags from this day (but final data!)

Seven forecasts from Q(-1) M1 to Q(+1) 1

2 data sets

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

Chart 1: Cumulative forecast weights across data sets

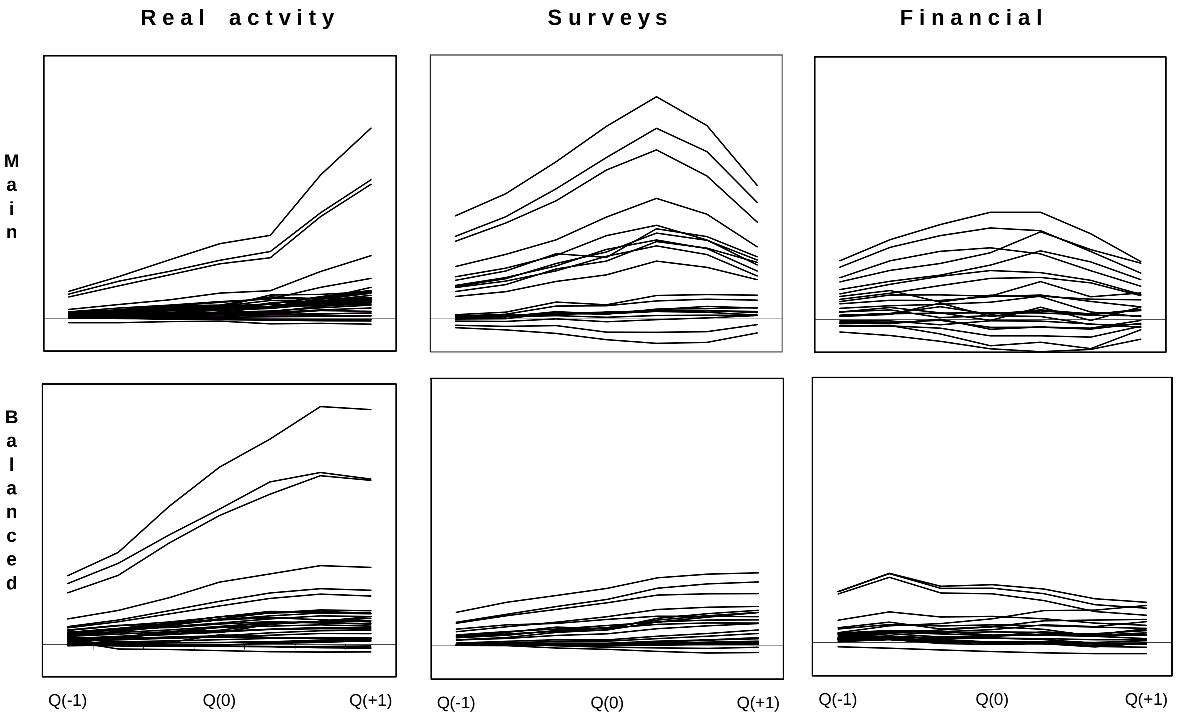


Table 1: Mean absolute contributions (MAC) of data groups

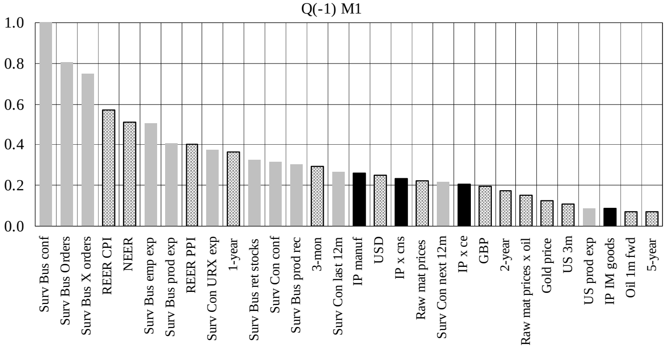
(1998 Q1 - 2005Q4)

Data	Main				Balanced			
	Fcst	Contributions (%)			Fcst	Contributions (%)		
	$\mathcal{Z}$	$\mathcal{S}$	$\mathcal{F}$	$\mathcal{R}$	$\mathcal{Z}$	$\mathcal{S}$	$\mathcal{F}$	$\mathcal{R}$
Q(-1) M1	0.158	60 %	57 %	14 %	0.135	34 %	46 %	46 %
Q(-1) M2	0.183	58 %	57 %	15 %	0.163	32 %	46 %	44 %
Q(-1) M3	0.196	61 %	56 %	16 %	0.192	28 %	42 %	49 %
Q(0) M1	0.227	62 %	50 %	16 %	0.188	34 %	40 %	48 %
Q(0) M2	0.245	63 %	42 %	17 %	0.199	35 %	35 %	47 %
Q(0) M3	0.230	61 %	37 %	25 %	0.206	35 %	29 %	52 %
Q(+1) M1	0.210	53 %	35 %	32 %	0.200	37 %	29 %	50 %

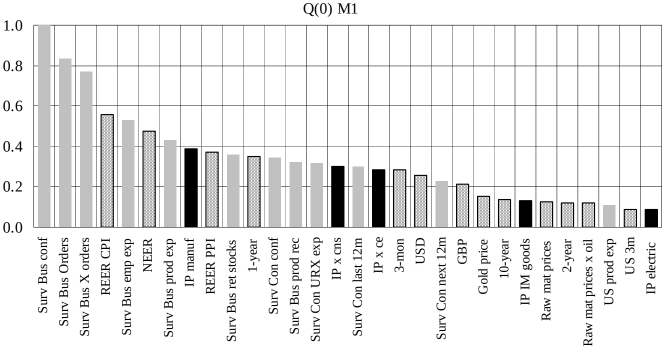
Chart A.1: Absolute cumulative weights

Main data

Q(-1) M1



Q(0) M1



Q(+1) M1

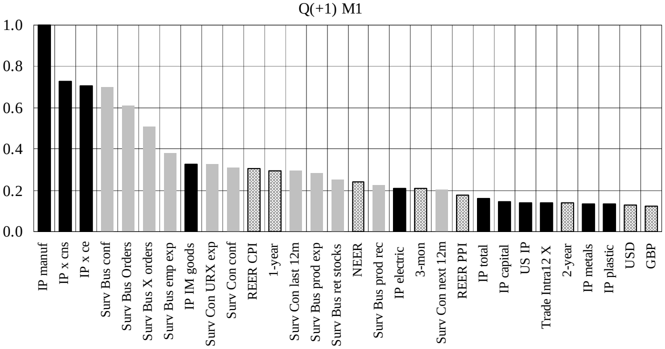
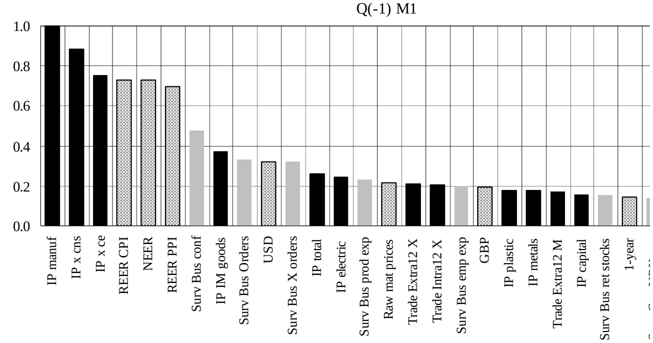


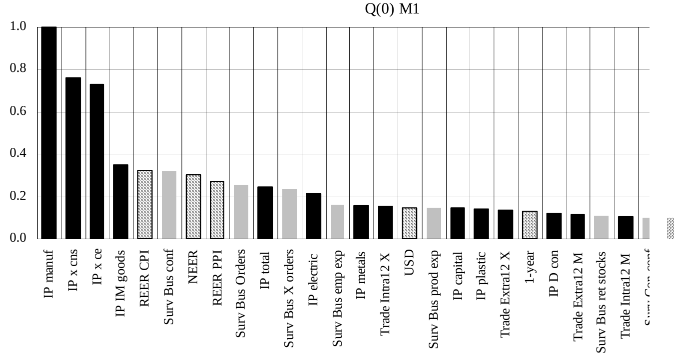
Chart A.3: Absolute cumulative weights

Balanced data

Q(-1) M1



Q(0) M1



Q(+1) M1

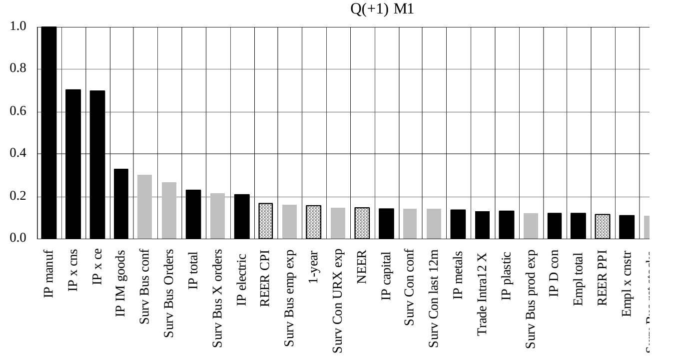


Table 2: Filter uncertainty

(Full-sample parameter estimates)

	Main				Balanced			
	$\mathcal{Z}$	$\mathcal{RS}$	$\mathcal{RF}$	$\mathcal{SF}$	$\mathcal{Z}$	$\mathcal{RS}$	$\mathcal{RF}$	$\mathcal{SF}$
Q(-1) M1	.178	.185	.182	178	.176	.180	.176	178
Q(-1) M2	.160	.172	.166	.160	.158	.165	.158	.160
Q(-1) M3	.137	.152	.145	.138	.135	.143	.135	.138
Q(0) M1	.100	.112	.119	.100	.091	.097	.091	.100
Q(0) M2	.070	.078	.100	.071	.056	.060	.056	.071
Q(0) M3	.037	.042	.068	043	.021	.023	.023	.043
Q(+1) M1	.029	.033	.043	042	.020	.023	.022	.042

Table 3: RMSE from recursive forecasts

(1998 Q1 - 2005Q4)

	AR	Main				Balanced			
		$\mathcal{Z}$	$\mathcal{RS}$	$\mathcal{RF}$	$\mathcal{SF}$	$\mathcal{Z}$	$\mathcal{RS}$	$\mathcal{RF}$	$\mathcal{SF}$
Q(-1) M1	.38	.33	37	.32	.33	.33	.35	.32	.33
Q(-1) M2	.35	.32	.36	.31	.32	.31	.33	.31	.32
Q(-1) M3	.35	.28	.33	.29	.28	.28	.30	.29	.28
Q(0) M1	.35	.28	.30	.30	.28	.26	.27	.26	.28
Q(0) M2	.31	.28	31	.29	.28	.25	.26	.24	.28
Q(0) M3	.31	.25	.28	.27	.27	.24	.25	.24	.27
Q(+1) M1	.31	.24	.25	.24	.27	.23	.24	.23	.27

Conclusions

Statistics to investigate the role of individual series in forecasts from a DFM

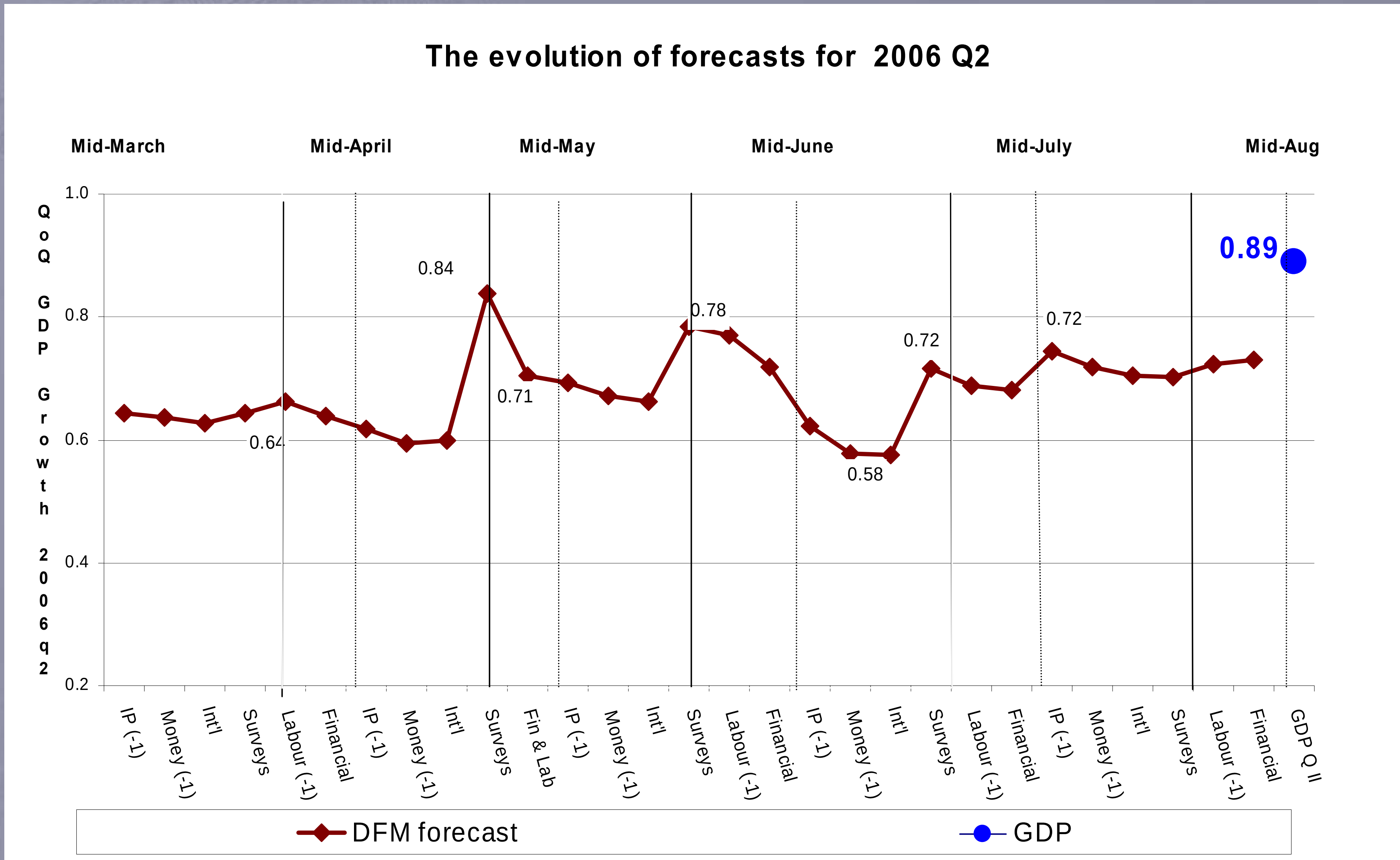
- Based on Kalman filter
- Deals with unbalanced data sets
- No need to re-estimate parameters

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

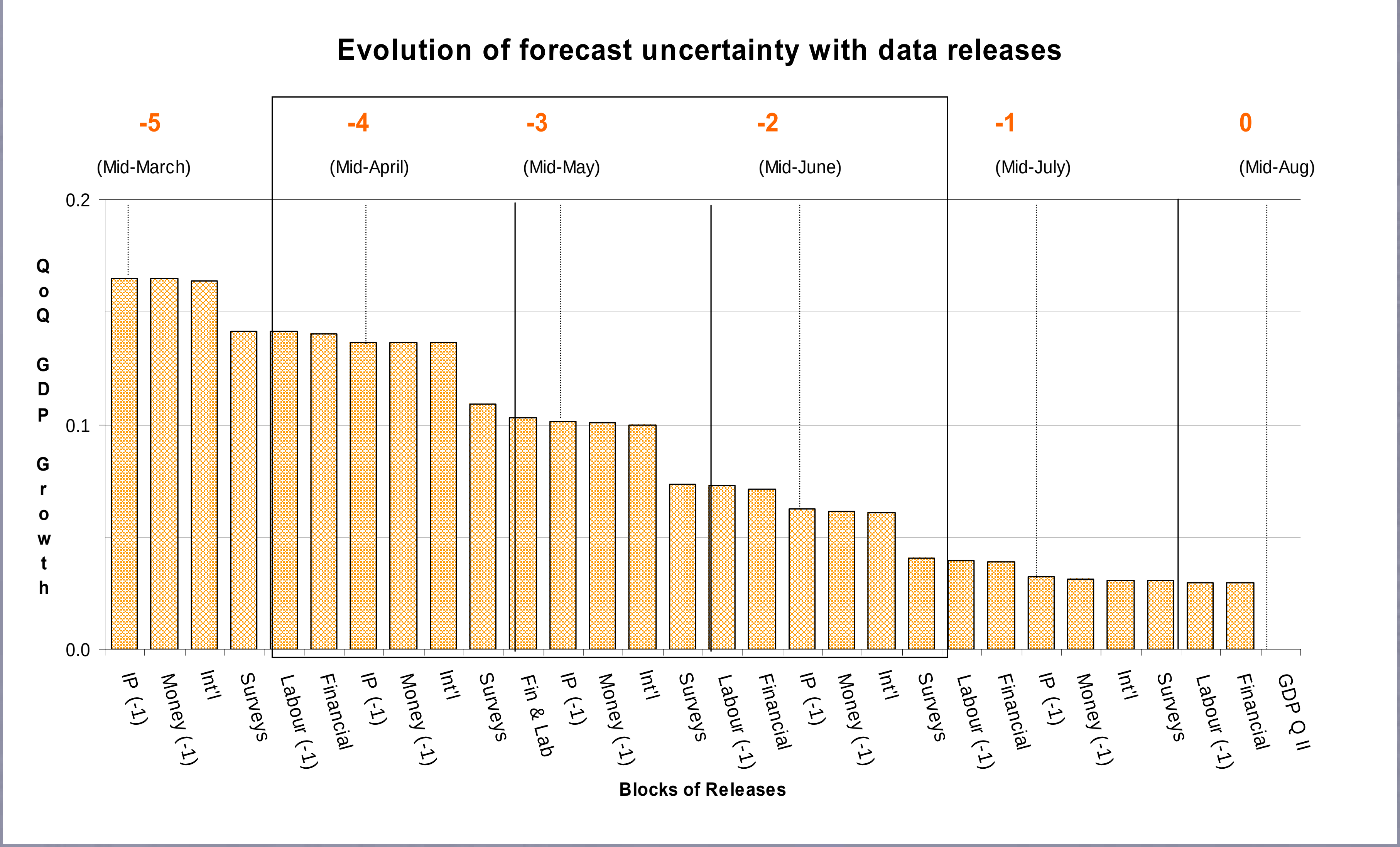
- Surveys are close substitutes to real data - less precise but published earlier
- Financial data provide complementary information

Balanced data give a wrong picture, but many studies have ignored unbalancedness

5 – Reading the News: 2006 Q2



5 – Reading the news: decline in uncertainty



5 – Reading the News: contributions in 2006 Q2