Improved Forecasting with Leading Indicators: The Principal Covariate Index

Christiaan Heij
Dick van Dijk
Patrick Groenen

Econometric Institute
Erasmus University Rotterdam
The Netherlands

1-st MFC, Rome
March 27, 2009

Outline

- Forecasting with Index Models
- Methodology: PCR and PCOVR
- Data and Forecast Evaluation
- Forecast Comparison
- Recession Probabilities
- Conclusion

Notation:

- y : variable of interest to be predicted (CCI of the Conference Board)

- $(x_1, \ldots, x_N)$: set of N predictor variables ($N = 10$ leading indicators of CLI)

- f : index (latent factor, to be constructed) ($f = \sum_{i=1}^{N} \gamma_i x_i$, weighted average)

Data compression and (static) forecasting:

$$(y, x_1, \ldots, x_N) \;\rightarrow\; f = \sum_{i=1}^{N} \gamma_i x_i \;\rightarrow\; \widehat{y} = \alpha + \beta f$$

(f can be seen as 'state of the economy')

Forecasting with Index Models - 4/22

Two popular index methods to forecast the Composite Coincident Index (CCI) of the Conference Board:

- **CLI: Composite Leading Index**

  weighted average of 10 leading indicators, with highly specific procedure to compute the weights

- **PCR: Principal Component Regression**

  weights chosen so as to approximate the N predictors by a single factor in an optimal way; this method is followed by, e.g., CFNAI of the Chicago Fed, by Stock & Watson, by Bernanke & Boivin

Forecasting with Index Models - 5/22

We propose an alternative method, **Principal Covariate Regression** (PCOVR).

In contrast with PCR, PCOVR takes also the variable to be forecasted into account in constructing the index.

For **illustrative purposes**, we restrict the database to $y =$ CCI and the 10 leading indicators x_i of the CLI and compare three methods:

- CLI (Conference Board)
- PCR (Chicago Fed)
- PCOVR (our new method)

Methodology - PCR - 6/22

Data compression and forecasting:

$$(y, x_1, \ldots, x_N) \;\rightarrow\; f = \sum_{i=1}^{N} \gamma_i x_i \;\rightarrow\; \widehat{y} = \alpha + \beta f$$

PCR constructs f to approximate $(x_1, \ldots, x_N)$, NOT y.

$$\min_{\gamma_i} \left[\min_{\delta_i} \quad S(x) = \sum_{i=1}^{N} ||x_i - \delta_i f||^2 \right]$$

The solution is the *first principal component* of the N predictor variables (that should be normalized, e.g., to z-scores).
Next, $\widehat{y} = \alpha + \beta f$ is obtained by regression:

$$\min_{\alpha,\beta} \quad S(y) = ||y - \alpha - \beta f||^2$$

Methodology - PCOVR - 7/22

$$(y, x_1, \ldots, x_N) \;\rightarrow\; f = \sum_{i=1}^{N} \gamma_i x_i \;\rightarrow\; \widehat{y} = \alpha + \beta f$$

Basic idea: Construct the index **not only** from predictors x_i, **but also** from predicted y, by *weighted average* of $S(x)$ and $S(y)$ of PCR:

$$\min_{\gamma_i, \delta_i, \alpha, \beta} \left[\; w_1 S(x) + w_2 S(y) \; \right]$$

- $0 < w < 1$, $w_1 = \frac{w}{||y||^2}$, $w_2 = \frac{1-w}{\sum_{i=1}^{N} ||x_i||^2}$

  $w = 0.5$: equal weights on normalized var.

- $w = 0 \rightarrow w_1 = 0 \rightarrow$ OLS

  $w = 1 \rightarrow w_2 = 0 \rightarrow$ PCR

- Choose grid on [0,1] and select w by cross validation (min MSE_y on hold-out sample).

Data and Forecast Evaluation - 8/22

Variable to be predicted: growth rate of CCI

Composite Coincident Index (Conference Board), constructed from 4 key macroeconomic series: employment, income, production, and sales

Predictor variables : 10 leading indices

Ten components of CLI (Composite Leading Index): 5 production related variables, 3 financial variables, employment, and consumer expectations

- all variables transformed to stationarity (as in Stock & Watson)
- monthly data 1959.01 - 2003.12
- CCI growth rate for different horizons: 3, 6, 12, and 24 months

Data and Forecast Evaluation - 9/22

- *Recursive forecasts* (simulated out-of-sample)

  at time t, use data up to t to construct factor and to estimate model, and forecast h-month growth over $[t+1, t+h]$

- *Moving window* of 120 months
  use data on $[t-119, t]$

- Forecast error: actual h-month growth minus predicted growth over $[t+1, t+h]$

- MSE: mean squared forecast error, shifting forecast time t from 1969.12 to 2003.12-h

Forecast Comparison - 10/22

6-month CCI growth rate and 3 forecasts:

- CLI (top)
- PCR (middle)
- PCOVR (bottom)

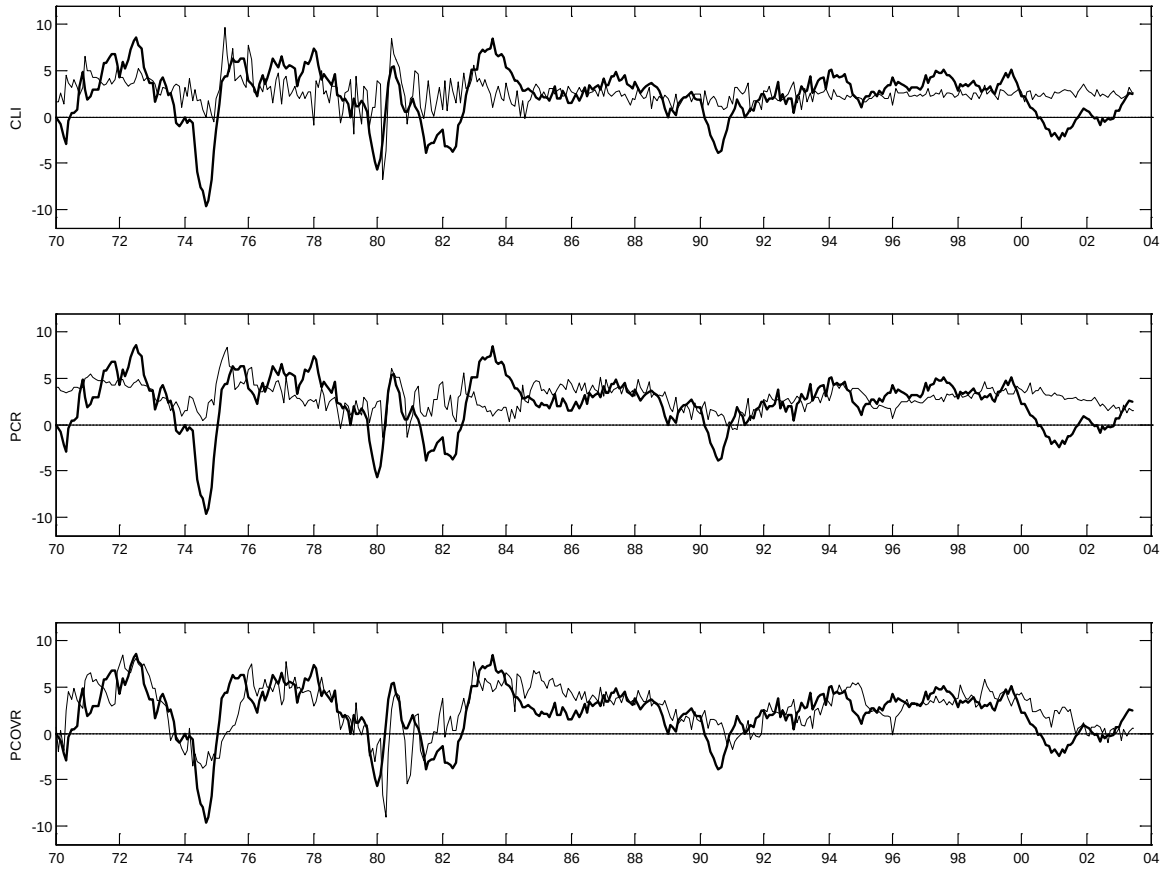

Table 1: Out-of-sample correlations between forecasts and actual CCI

Period	Index	Forecast Horizon			
		3	6	12	24
1970-2003	CLI	0.32	0.32	0.33	0.27
	PCR	0.31	0.36	0.44	0.43
	PCOVR	**0.62**	**0.66**	**0.68**	**0.54**
1970-1983	CLI	0.40	0.42	0.42	0.37
	PCR	0.34	0.45	0.60	0.61
	PCOVR	**0.65**	**0.71**	**0.80**	**0.70**
1984-1993	CLI	0.14	0.18	0.26	0.40
	PCR	0.45	0.43	**0.51**	0.44
	PCOVR	**0.52**	**0.51**	**0.51**	**0.64**
1994-2003	CLI	-0.21	-0.11	-0.17	-0.22
	PCR	0.22	0.14	-0.13	-0.64
	PCOVR	**0.53**	**0.54**	**0.41**	**-0.06**

Table 2: Mean squared prediction errors of CCI

Period	h	var(y)	Const	CLI	PCR	PCOVR
1970-2003	3	10.66	1.08	0.93	0.95	**0.68**
	6	8.17	0.86	0.72	0.72	**0.50**
	12	5.93	0.65	0.53	0.50	**0.34**
	24	3.67	0.39	0.34	0.31	**0.27**
1970-1983	3	19.32	1.08	0.88	0.92	**0.60**
	6	14.34	0.83	0.66	0.64	**0.40**
	12	9.79	0.60	0.47	0.39	**0.19**
	24	5.81	0.36	0.31	0.24	**0.21**
1984-1993	3	4.36	1.03	1.04	**0.98**	**0.98**
	6	3.03	**0.72**	0.73	**0.72**	0.88
	12	2.10	0.48	0.49	**0.46**	0.88
	24	1.56	0.36	**0.32**	0.33	0.35
1994-2003	3	4.62	1.15	1.20	1.09	**0.93**
	6	4.00	1.03	1.03	1.03	**0.89**
	12	3.48	0.91	0.92	1.02	**0.80**
	24	2.85	0.74	0.74	0.89	**0.69**

'var(y)' denotes the variance of the predicted variable

'Great Moderation' in the early eighties, var(y) $\downarrow\downarrow$

other columns show MSE of each method relative to this variance

'Const' forecasts by the sample average over the last 10 years

Forecast Comparison - 13/22

Main findings:

- PCOVR correlates best with future CCI
- applies for all horizons and all periods
- PCOVR smallest MSE over full data period
- compared to CLI and PCR, MSE reduction $\approx 30\%$ for $h = 3, 6, 12$, $\approx 10\%$ for $h = 24$
- PCOVR best in 1970-1983 and 1994-2003
- CLI and PCR better in period 1984-1993 PCOVR suffers from 'Great Moderation', with break in macroeconomic volatility
- all methods problem for 1994-2003, $h = 24$

Recession Forecasts - 14/22

Definition of Recessions and Expansions:

- Q_1 and Q_2: CCI growth rate over **future** months 1,2,3 and 4,5,6
- Recession ($R = 1$): $Q_1 < 0$ and $Q_2 < 0$
- Expansion ($R = 0$): $Q_1 > 0$ and $Q_2 > 0$
- Mixed ($R = 0.5$): otherwise

Recession probabilities are estimated by assuming a joint normal distribution for (Q_1, Q_2)

Recession signal ($\widehat{R} = 1$): current recession probability larger than average of last 10 years

Quadratic Probability Score (QPS):

$\frac{1}{T}\sum_{t=1}^{T}(R_t - \widehat{R}_t)^2$ (excluding mixed months)

Recession Forecasts - 15/22

Future 6-months recessions and estimated probabilities:

- empirical recessions & NBER (top)
- emp. recessions & CLI forecasts (2-nd)
- idem for PCR forecasts (3-rd)
- idem for PCOVR forecasts (bottom)

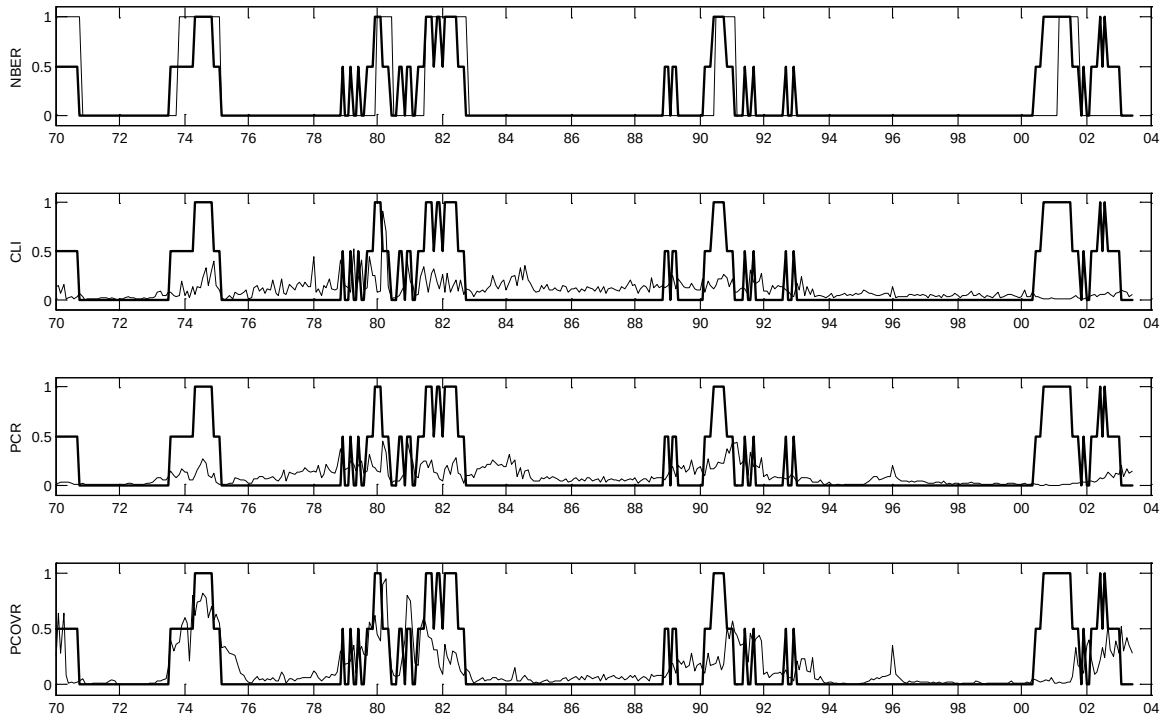

Table 3: Recessions: Actual and Forecasts

	CLI			PCR			PCOVR		
Actual	Yes	No	Total	Yes	No	Total	Yes	No	Total
Forecast Yes	19	78	97	25	96	121	24	46	70
Forecast No	19	214	233	13	196	209	14	246	260
Actual Total	38	292	330	38	292	330	38	292	330
% Correct Forecasts	50	73	71	**66**	67	67	63	**84**	**82**

Yes: Recession in coming six months (38)

No: Expansion in coming six months (292)

Months with 'Mixed regime' are excluded (72)

Forecast period: 1970.01 to 2003.06

Table 4: Quadratic Probability Scores

Period (Observations)	CLI	PCR	PCOVR
Expansions (292)	0.014	**0.012**	0.017
Recessions (38)	0.788	0.778	**0.560**
Both (330)	0.103	0.100	**0.080**

Future six-month recessions and expansions

Months with 'Mixed regime' are excluded (72)

Forecast period: 1970.01 to 2003.06

Recession Forecasts - 18/22

Main findings:

- PCOVR provides overall best signals
- PCOVR has 82% correct forecasts

  PCR 67% and CLI 71%
- PCOVR has lowest QPS at recessions

  and also for recessions and expansions jointly

Conclusion - 19/22

Main conclusions:

- Index models help to forecast macroeconomic variables
- PCOVR often performs better than conventional PCR
- PCOVR combines the objectives of forecasting and predictor compression

Conclusion - 20/22

Extensions in paper:

- Use of dynamic forecast models

  (including lags of y and f)

- Use of richer predictor sets

  (128 Stock-Watson variables)

- Use of multiple factors

Conclusion - 21/22

Some points for further research:

- Forecasting with more recent data (up to 2009)
- Forecasting with real-time data (vintage data)
- Targeted predictor selection (compare previous talk)
- Non-linear factor extraction (breaks, kernel methods, ...)

References 22/22

- Heij, Van Dijk and Groenen (2007), Improved forecasting with leading indicators: the Principal Covariate Index, EI 2007-23.

- Heij, Van Dijk and Groenen (2006), Time series forecasting by principal covariate regression, EI 2006-37.

- Marcellino (2006), Leading indicators, in Handbook Economic Forecasting, Vol 1.

- Stock and Watson (2002), Forecasting using principal components from a large number of predictors, JASA.

- De Jong and Kiers (1992), Principal covariate regression, Chemometrics and Intelligent Laboratory Systems.