Economic Forecasting with Mixed Frequency Data

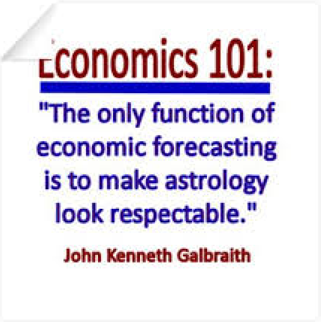
Prediction and Forecasting Conference

Scott A. Brave[1]
Federal Reserve Bank of Chicago

CCASA, Mar 2015

[1] The views herein are my own and do not necessarily represent those of the Federal Reserve Bank of Chicago or the Federal Reserve System.

Economic forecasting has taken a hit in the aftermath of the Crisis...



Source: quotespictures.com

or maybe it never really had that great of a reputation to begin with!

Principal Components Analysis (PCA)

Principal components are consistent estimates of the **common factors** F_t in a large panel of time series, e.g. Stock and Watson (2002).

$$\begin{aligned} x_t &= \Gamma F_t + \upsilon_t \\ \upsilon_t &\sim N(0, \sigma^2 I) \end{aligned}$$

Requires that x_t be stationary (typically also demeaned and standardized)

and a scale normalization ($\Gamma'\Gamma = I$).

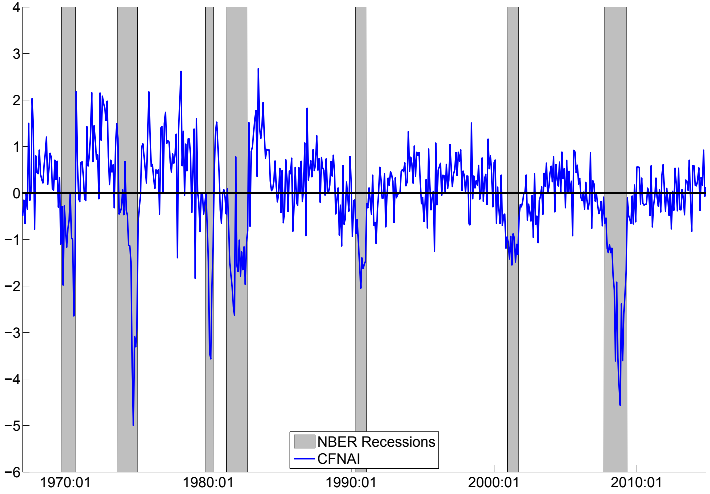
For almost 15 years, the Chicago Fed has produced a monthly index of national economic activity (the **CFNAI**) constructed in this way.

Ex: The Chicago Fed National Activity Index (CFNAI)

The CFNAI is the **first principal component** of 85 monthly indicators of U.S. economic activity, e.g. IP, Payroll Employment, Housing Starts, etc.

Some of its past uses include:

1. Stock and Watson (1999)/Fisher (2000): forecasting inflation
2. Evans, Liu, Pham-Kanter (2001)/Brave (2009,2012): business cycles
3. Brave and Butters (2010,2013,2014): nowcasting GDP growth



Source: chicagofed.org

Dynamic Factor Analysis

1) HAC idiosyncratic errors, and 2) VAR(p) factor dynamics.

$$
\begin{aligned}
x_t &= \Gamma F_t + \upsilon_t \\
F_t &= T F_{t-1} + R v_t \\
\upsilon_{it} &= \rho \upsilon_{it-1} + \epsilon_{it} \\
v_t \sim N(0, \Omega) & \qquad \epsilon_t \sim N(0, \Sigma) \\
\Sigma_{ii} = \sigma^2_{\epsilon_i} & \qquad \Sigma_{ij} = 0 \;\; i \neq j
\end{aligned}
$$

Large, approximate factor structure literature assumes Σ is diagonal.

Requires a scale normalization on the **common shocks** v_t ($\Omega = I$).

Ex: Mariano and Murasawa (2003), Aruoba, Diebold, and Scotti (2009)

Collapsed Dynamic Factor (CDF) Analysis

Began as a way to handle large dimensional problems in a state-space framework, e.g. Jungbacker, Koopman, and van der Wel (2011).

Apply $A = (\Gamma'\Sigma^{-1}\Gamma)^{-1}\Gamma'\Sigma^{-1}$ to $\tilde{x}_t \equiv x_t - \rho x_{t-1}$.

$$\begin{aligned} A\tilde{x}_t \equiv \hat{F}_t &= \tilde{F}_t + \epsilon_t \\ F_t &= TF_{t-1} + Rv_t \\ v_t \sim N(0, \Omega) & \qquad \epsilon_t \sim N(0, \Gamma'\Sigma^{-1}\Gamma) \end{aligned}$$

Essentially Cochrane-Orcutt estimation combined with factor GLS.

Brauning and Koopman (2014) instead let $\epsilon_t \sim N(0, V)$ and exclude one element of x_t, a *"target variable"*, from the collapse.

CDF Analysis with a Target Variable

Perform a trend-cycle decomposition of a lower frequency **target** (y_t).

$$y_t = \underbrace{\mu_t}_{trend} + \underbrace{\alpha F_t}_{cycle} + \epsilon_t$$

F_t are principal components/factors ($\hat{F}_t$) minus *measurement error* (u_t)

$$\begin{aligned} \hat{F}_t &= F_t + u_t \\ F_t &= TF_{t-1} + Rv_t \end{aligned}$$

and follow a dynamic process where ϵ_t, u_t, v_t are i.i.d. normally distributed.

Similar to a rotation of $\hat{F}_t$ in the space of the target variable, or instrumenting for an *"errors-in-variables"* problem.

Handling Mixed Frequency Data in a State-Space System

Requires casting the data in a state-space system that can accommodate:

- Missing observations; Durbin and Koopman (2001)
- Temporal aggregation; Harvey (1989)
- Time-varying parameters; Brave, Butters, and Justiniano (2015)

Augment the state-space with **"Harvey" accumulators**, ζ_t, which act as **constraints** on the high frequency latent states to preserve temporal aggregation within a lower frequency period of observation.

Offers an alternative mixed frequency approximation to various forms of interpolation (ADL) or "bridge" equations (MIDAS).

Simple Sum Accumulator:

Typically associated with flows where the value reported, y_t, is cumulative over the base frequency being modeled, e.g. quarterly growth rates in logs.

$$\psi_t = \begin{cases} 0 & \text{first period within aggregation period} \\ 1 & \text{second period within aggregation period} \\ \vdots & \vdots \\ 1 & \text{last period within aggregation period} \end{cases}$$

$$\zeta_t = y_t + \psi_t \zeta_{t-1}$$

Simple Average

Generalized Kalman Filter/Smoother

Estimation is made computationally efficient by fast Kalman filter/smoother algorithms with time-varying parameters.

Univariate version of the Kalman filter (Durbin and Koopman, 2012) transforms a multivariate observation equation into N univariate systems.

1. Eliminates the need to calculate the inverse of a $N \times N$ matrix
2. Lowers computational burden when N is small relative to size of state

See Brave, Butters, and Justiniano (2015) for a technical note summarizing the Generalized Kalman filter/smoother including the accumulators.

Maximum Likelihood Estimation

State-space for panel of time series, X_{it}, and latent factors, F_t, is given by

$$\begin{aligned} x_t &= Z\alpha_t + \epsilon_t \\ \alpha_t &= T\alpha_{t-1} + R\eta_t \end{aligned}$$

where $\alpha_{t-i} = [F_{t-i}, \zeta_{t-i}]$ for $i = 0, 1, \ldots, p$, and $x_t = [X_{it}]$ for $i = 1, \ldots, N$ and $\epsilon_t \sim N(0, H)$ and $\eta_t \sim N(0, I)$.

Lower frequency X_{it} load onto the appropriate Harvey accumulator ζ_t via Z. Deterministic dynamics of the accumulators are embedded in T and R.

An EM algorithm estimator is described in Banburra and Modugno (2010). A full-ML estimator is described in Mariano and Murasawa (2010).

What is Nowcasting?
Current quarter (or release) GDP forecasting.

Why should I care?
GDP comes out with a lag and is twice revised (3-month lag). There exists more frequently available data informative for GDP and its revisions.

How do I do it?
Principal component or factor analytic methods have been commonly used to summarize the available data for this purpose.

- Stock and Watson (2002) SW
- Giannone, Reichlin, and Small (2008) GRS
- Banbura and Rustler (2011) BR

Here, I use the Brauning and Koopman (2014) CDF method of nowcasting.

Updating Stock and Watson's (1998) Trend-Cycle Decomposition

An unobserved components model for monthly GDP growth, GDP_t.

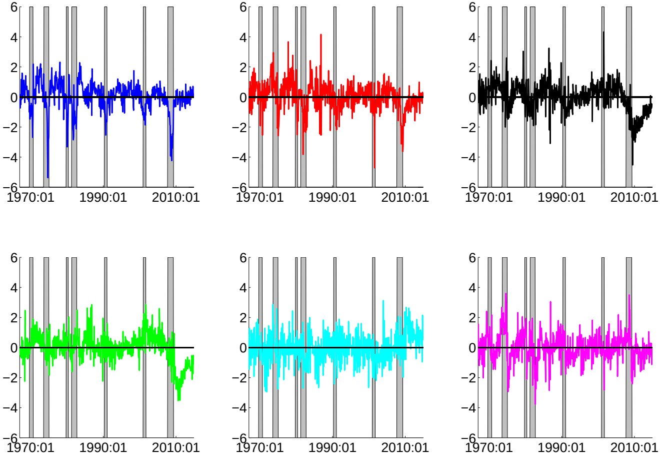
Measurement Equations:

$$\begin{aligned} \hat{F}_{t+1} &= F_{t+1} + u_{t+1} && (CDF) \\ GDP_{3t} &= \delta_t(L) GDP_t && \end{aligned}$$

Triangle Average

State Equations:

$$\begin{aligned} GDP_t &= \mu_t + \alpha F_t + \beta F_{t-1} + \epsilon_t && \\ \mu_{t+1} &= \mu_t + \eta_{t+1} && (Trend) \\ F_{t+1} &= TF_t + v_{t+1} && (Cycle) \\ \rho(L)\epsilon_{t+1} &= \varepsilon_{t+1} && (Irregular) \end{aligned}$$



Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | Factor 6

Predicting Business Cycles

Consider the following problem faced in medical statistics:

Given a known incidence of a disease in a population, how likely is it that a positive test result is reflective of a true occurrence in sample?

Consider the similar problem with respect to business cycle detection:

Given known incidences of recession in U.S. history, how likely is it that a reading of the factors truly reflects a contraction in economic activity?

We can use the past as a guide to judge our common factors as indicators of the business cycle, i.e. Berge and Jorda (2011).

Receiver Operating Characteristics (ROC) Analysis

We compare the accuracy of the factors in capturing NBER business cycles by calculating the area under their ROC curves.

The ROC curve is defined over the range of possible threshold rules c:

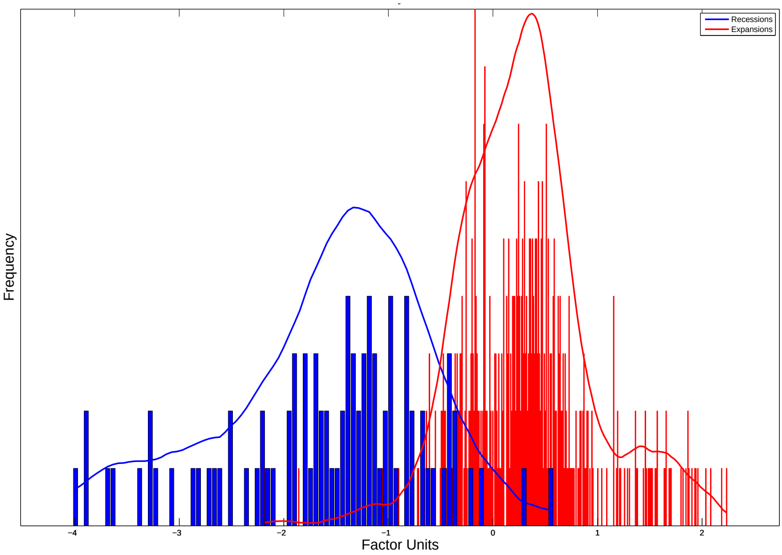
$$\begin{aligned} TP(c) &= P[F_t \geq c | S_t = 1] \\ FP(c) &= P[F_t \geq c | S_t = 0] \end{aligned}$$

where S_t is a binary variable with 1 representing an NBER recession. The Cartesian convention is given as:

$$\{ROC(r), r\}_{r=0}^{1}$$

where $ROC(r) = TP(c)$ and $r = FP(c)$.

We can non-parametrically fit this curve and find its area, **AUROC**.


Recessions
Expansions
Frequency
Factor Units
−4
−3
−2
−1
0
1
2

Area Under the ROC curve (AUROC) at Monthly Lags and Leads

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
-6m	**0.74**	**0.77**	**0.62**	**0.61**	**0.59**	**0.37**
-3m	**0.87**	**0.74**	0.55	**0.61**	**0.59**	0.52
0m	**0.94**	**0.62**	0.52	0.54	0.53	**0.67**
+3m	**0.86**	0.53	0.55	0.51	0.50	**0.76**
+6m	**0.75**	0.49	0.54	0.49	0.49	**0.75**

Bold denotes statistical significance with 95% confidence relative to 0.50

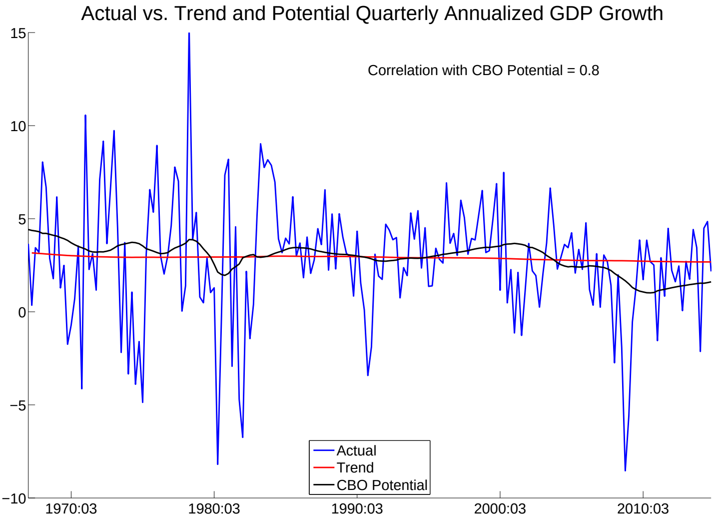
1. Factor 1 (CFNAI) is best *coincident* indicator of the business cycle
2. Factors 2, 3, 4, and 5 *lag* the business cycle
3. Factor 6 *leads* the business cycle

Fraction of Data Variance Explained

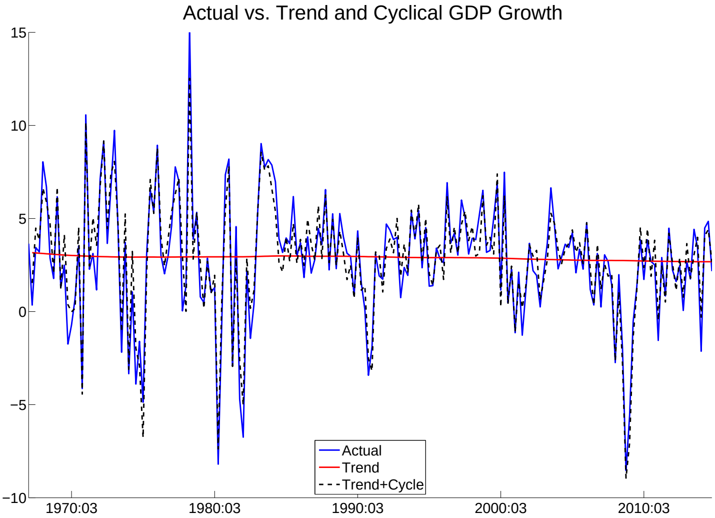
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Total	0.25	0.10	0.07	0.05	0.04	0.01
P&I	0.38	0.07	0.20	0.06	0.37	0.33
EU&H	0.37	0.16	0.10	0.12	0.09	0.31
C&H	0.07	0.24	0.53	0.40	0.05	0.06
SO&I	0.18	0.53	0.17	0.42	0.49	0.30

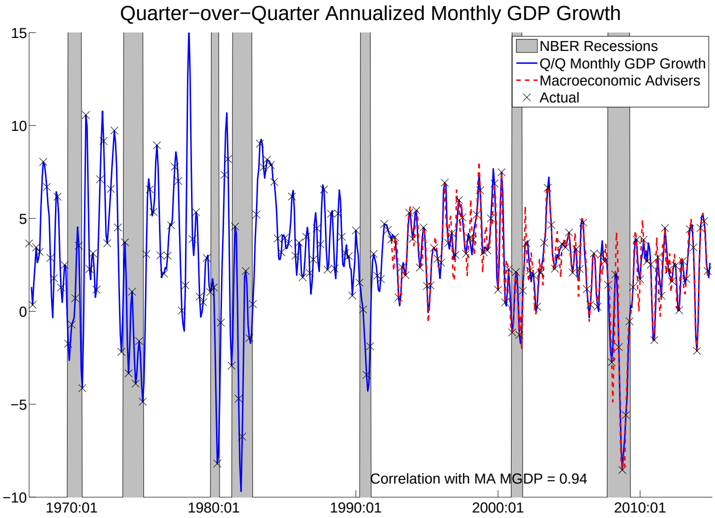
Red denotes share explained $\geq$ 20%

- Production-related indicators explain 1st, 3rd, 5th, and 6th factors
- Employment-related indicators explain 1st and 6th factors
- Consumption and housing indicators explain 2nd, 3rd, and 4th factors
- Sales, orders, and inventories explain 2nd, 4th, 5th, and 6th factors



Source: CBO Real Potential GDP growth via Haver Analytics


Actual vs. Trend and Cyclical GDP Growth
15
10
5
0
−5
−10
1970:03
1980:03
1990:03
2000:03
2010:03
Actual
Trend
Trend+Cycle



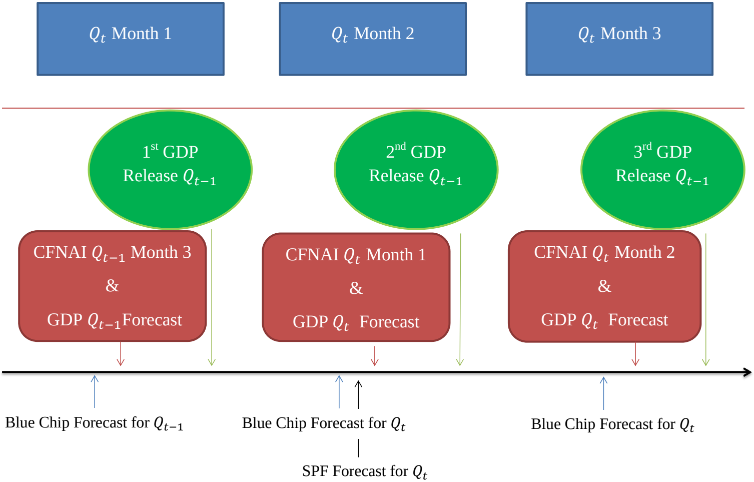
Source: Macroeconomic Advisers Monthly GDP via Haver Analytics

Nowcasting GDP Growth in Real-time

We design a real-time out-of-sample nowcasting exercise to back test our model akin to Banbura, Giannone, Modugno, and Reichlin (2013).

- Sample: 2004:Q3 - 2014:Q3
- Use real-time monthly indicators to construct the factors
- Use real-time GDP data to both forecast and evaluate models
- Evaluate against third real-time GDP release

To benchmark model performance, we make MSFE comparisons against the *Blue Chip Consensus* and *Survey of Professional Forecasters*.


Q_t Month 1
Q_t Month 2
Q_t Month 3
1st GDP Release Q_{t-1}
2nd GDP Release Q_{t-1}
3rd GDP Release Q_{t-1}
CFNAI Q_{t-1} Month 3 & GDP Q_{t-1} Forecast
CFNAI Q_t Month 1 & GDP Q_t Forecast
CFNAI Q_t Month 2 & GDP Q_t Forecast
Blue Chip Forecast for Q_{t-1}
Blue Chip Forecast for Q_t
Blue Chip Forecast for Q_t
SPF Forecast for Q_t

We use the CFNAI data archives, ALFRED, the Real-time Dataset for Macroeconomists (RDM), and Haver Analytics to obtain vintage data.

1. 85 monthly indicators in the CFNAI (CFNAI)
2. Real imports and exports of goods (ALFRED and Haver)
3. Real federal outlays (Haver)
4. Real retail and food service sales (ALFRED)
5. Real private residential construction spending (CFNAI and Haver)
6. Total business inventories-to-sales ratio (ALFRED and Haver)
7. Aggregate private weekly hours worked (ALFRED)
8. Real GDP growth (RDM)
9. Blue Chip Consensus nowcast (Haver)
10. Survey of Professional Forecasters nowcast (Haver)

Diebold-Mariano Test of Equal MSFE

To compare our model against the survey forecasts we use a modified version of the Diebold and Mariano (1995) test of equal MSFE.

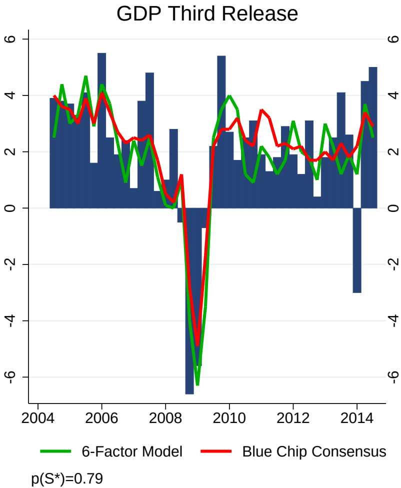
- Harvey, Leybourne, and Newbold (1997) sample size correction κ
- Clark and McCracken (2012) HAC error variances $V^{\frac{1}{2}}$

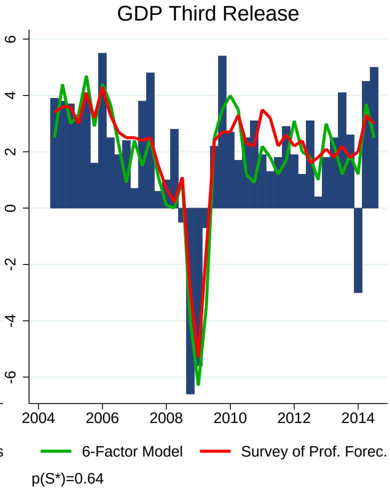
The test statistics are compared against standard normal critical values.

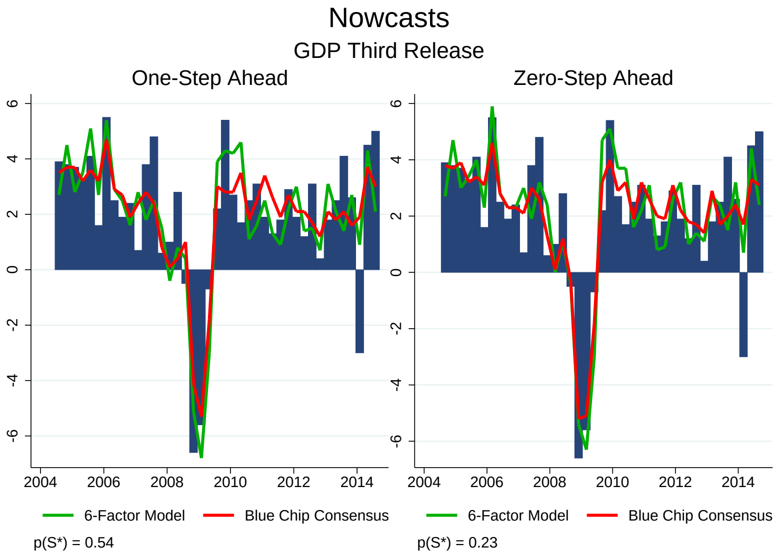
$$
\begin{aligned}
H0 &: MSFE_{survey} - MSFE_{model} = 0 \\
S^* &= (\kappa * (MSFE_{survey} - MSFE_{model}))/V^{\frac{1}{2}}
\end{aligned}
$$

Nowcasts

Two-Step Ahead






Nowcasts
GDP Third Release
One-Step Ahead
Zero-Step Ahead
6
4
2
0
-2
-4
-6
2004
2006
2008
2010
2012
2014
6-Factor Model
Blue Chip Consensus
p(S*) = 0.54
p(S*) = 0.23

Directions for Further Research

There is still considerable room for improvement in estimation and data.

- Gibbs Sampling w/ Bayesian shrinkage and stochastic volatility
- Further expand the scope of the dataset

Alternatively, in Brave, Butters, and Justiniano (2015), we show that a mixed frequency BVAR with a subset of these data in levels

1. Performs similarly in nowcasting, but
2. Produces superior forecasts up to 4 quarters ahead

in comparison to the same surveys of professional forecasters.

Appendix

Simple Average Accumulator:

Typically associated with flows/rates where the value reported is an average for the period, e.g. quarterly PCE.

$$\psi_t = \begin{cases} 1 & \text{first period within aggregation period} \\ 2 & \text{second period within aggregation period} \\ \vdots & \vdots \\ S & \text{last period within aggregation period} \end{cases}$$

$$\zeta_t = \frac{1}{\psi_t} y_t + \frac{(\psi_t - 1)}{\psi_t} \zeta_{t-1}$$

back

Triangle Average Accumulator:

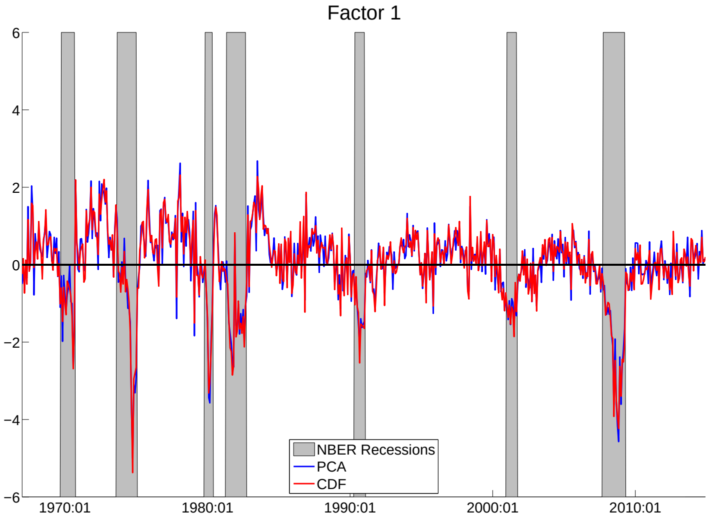
Typically associated with growth rates in values that have been averaged, e.g. quarter-over-quarter growth in monthly GDP.

$$\psi_t = \begin{cases} 1 & \text{first period within aggregation period} \\ 2 & \text{second period within aggregation period} \\ \vdots & \vdots \\ S & \text{last period within aggregation period} \end{cases}$$

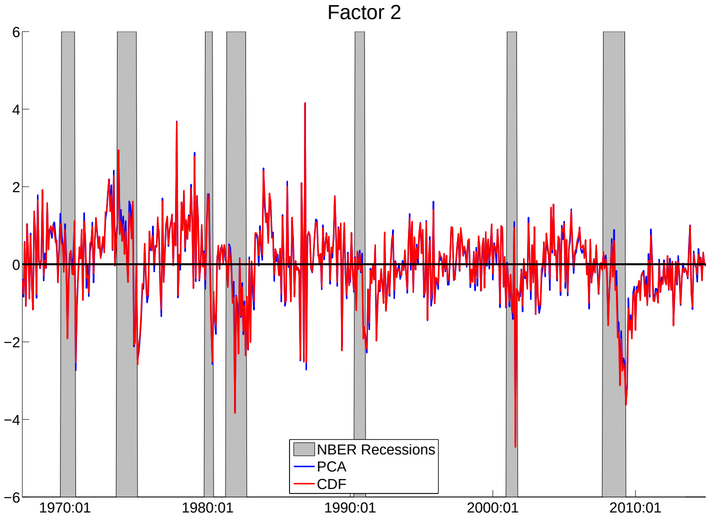
$$\zeta_t = \frac{1}{\psi_t}\left(y_t + y_{t-1} + \cdots + y_{t-H+1}\right) + \frac{(\psi_t - 1)}{\psi_t}\zeta_{t-1}$$

where H denotes horizon of the **difference** not necessarily the **frequency**

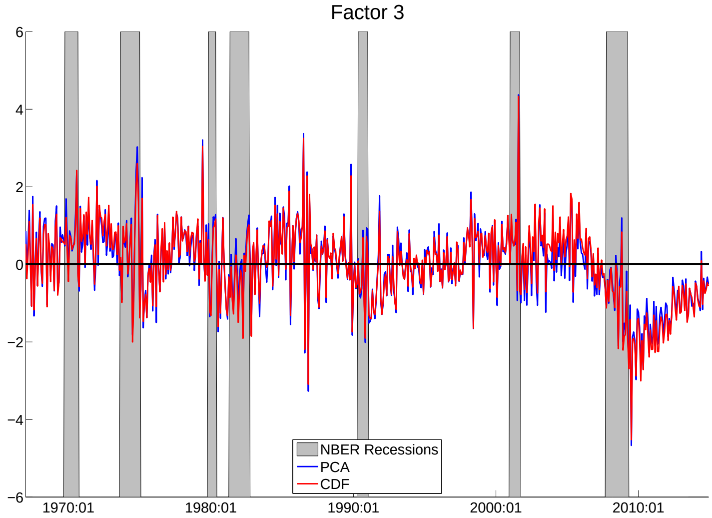
back


Factor 1
6
4
2
0
−2
−4
−6
1970:01
1980:01
1990:01
2000:01
2010:01
NBER Recessions
PCA
CDF

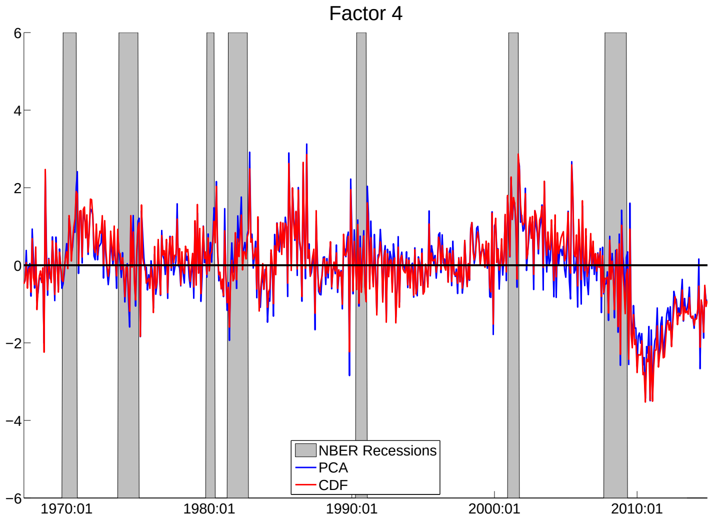
back


Factor 2
6
4
2
0
−2
−4
−6
1970:01
1980:01
1990:01
2000:01
2010:01
NBER Recessions
PCA
CDF

back

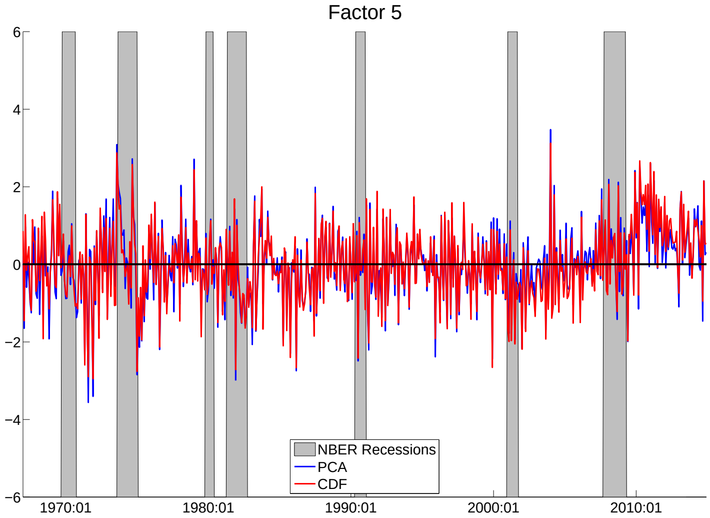

Factor 3
6
4
2
0
−2
−4
−6
1970:01
1980:01
1990:01
2000:01
2010:01
NBER Recessions
PCA
CDF

back

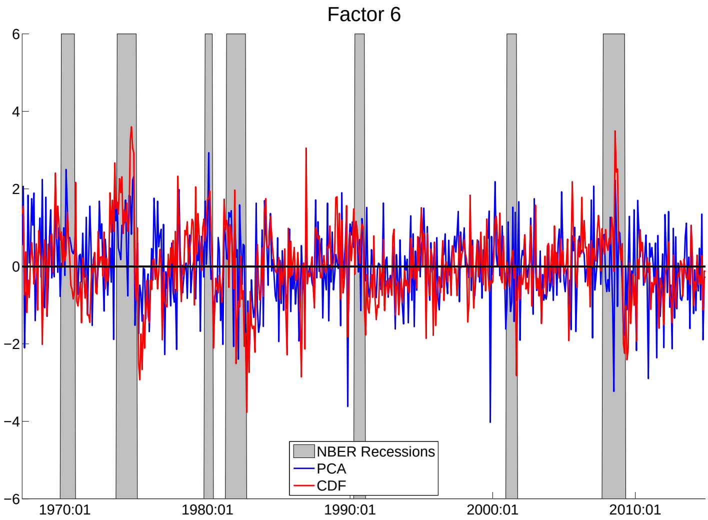


Factor 4
6
4
2
0
−2
−4
−6
1970:01
1980:01
1990:01
2000:01
2010:01
NBER Recessions
PCA
CDF

back


Factor 5
6
4
2
0
−2
−4
−6
1970:01
1980:01
1990:01
2000:01
2010:01
NBER Recessions
PCA
CDF

back



Factor 6
6
4
2
0
−2
−4
−6
1970:01
1980:01
1990:01
2000:01
2010:01
NBER Recessions
PCA
CDF

back

Ex: Stock and Watson (2002)

$$\begin{bmatrix} X_t \\ GDP_{3t} \end{bmatrix} = \begin{bmatrix} \Gamma & \cdots & 0 \\ \beta(0) & \cdots & \beta(L) \end{bmatrix} \begin{bmatrix} F_t \\ \vdots \\ F_{t-L} \end{bmatrix} + \begin{bmatrix} 0 \\ d \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}$$

$$\begin{bmatrix} F_t \\ \vdots \\ F_{t-L} \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 \\ I & 0 & \cdots \\ 0 & I & 0 \end{bmatrix} \begin{bmatrix} F_{t-1} \\ \vdots \\ F_{t-L-1} \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} + R\xi_t$$

$$\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2 I), \;\; \eta_t \sim \mathcal{N}(0, \sigma_\eta^2), \;\; \xi_t \equiv 0$$

back

Ex: Giannone, Reichlin, and Small (2008)

$$\begin{bmatrix} X_t \\ GDP_{3t} \end{bmatrix} = \begin{bmatrix} \Gamma & \cdots & 0 \\ \beta(0) & \cdots & \beta(L) \end{bmatrix} \begin{bmatrix} F_t \\ \vdots \\ F_{t-L} \end{bmatrix} + \begin{bmatrix} 0 \\ d \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}$$

$$\begin{bmatrix} F_t \\ \vdots \\ F_{t-L} \end{bmatrix} = \begin{bmatrix} \alpha(1) & \cdots & \alpha(L) \\ I & 0 & 0 \\ 0 & I & 0 \end{bmatrix} \begin{bmatrix} F_{t-1} \\ \vdots \\ F_{t-L-1} \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} + R\xi_t$$

$$\varepsilon_t \sim \mathcal{N}(0, diag(\sigma^2_{\varepsilon_i})), \ \eta_t \sim \mathcal{N}(0, \sigma^2_\eta), \ \xi_t \sim \mathcal{N}(0, I)$$

back

Ex: Banbura and Rustler (2011)

$$\begin{bmatrix} X_t \\ GDP_{3t} \end{bmatrix} = \begin{bmatrix} \Gamma & \cdots & 0 & 0 \\ 0 & \cdots & 0 & \delta_t(L) \end{bmatrix} \begin{bmatrix} F_t \\ \vdots \\ F_{t-L} \\ GDP_t \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} F_t \\ \vdots \\ F_{t-L} \\ GDP_t \end{bmatrix} = \begin{bmatrix} \alpha(1) & \cdots & \alpha(L) & 0 \\ I & 0 & \cdots & 0 \\ 0 & I & 0 & 0 \\ \beta(0) & \cdots & \beta(L) & 0 \end{bmatrix} \begin{bmatrix} F_{t-1} \\ \vdots \\ F_{t-L-1} \\ GDP_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ c \end{bmatrix} + R \begin{bmatrix} \xi_t \\ \eta_t \end{bmatrix}$$

$$\varepsilon_t \sim \mathcal{N}(0, diag(\sigma^2_{\varepsilon_i})), \; \eta_t \sim \mathcal{N}(0, \sigma^2_\eta), \; \xi_t \sim \mathcal{N}(0, I)$$

back