# Forecasting and Now-Casting with Disparate Predictors: Dynamic Factor Models and Beyond

FEMES 2006 Meetings
Beijing

James. H. Stock
Harvard University

Joint work with Mark W. Watson, Princeton University

# Introduction

- The history of macroeconomic forecasting has been an uneasy coexistence of "structural" models and "time series" models.
- This talk focuses on a class of models that can incorporate economic theory (as much or as little as desired) into a time series structure – dynamic factor models, using a large number of series.
- The data and economic forecasting environment:
  - there are many predictors ("large *n*")
  - variables are measured with error (possibly large)
  - the available time series might be short, might have different start dates and might have different sampling frequencies (mixed monthly-quarterly)
  - there might be breaks in the individual series, e.g. changes in definitions, collection methods, etc.

In this talk I will:

- Summarize an exciting modeling framework that has received a lot of recent attention: the dynamic factor model (DFM)
  - One main message is that, in the DFM, having many time series is a "blessing" of dimensionality, not a "curse" – having many series can make up for deficiencies in any one series. (This will be made more precise.)
- Discuss main theoretical results for DFMs
- Go through an empirical example for U.S. data with $n = 132$ variables
- Provide a general framework for optimal linear forecasting in a stationary environment and compare the DFM forecasts to the "optimal" (in a specific sense) forecasts – do forecasts based on a small number of factors omit potentially useful information?

Outline

1. Introduction
2. Background – VARs and their limitations
3. Dynamic factor models: some theory, VARs v. DFMs, and a survey of recent theoretical results
4. An empirical DFM – US data, 132 series
5. Econometric theory of forecasting using many predictors
6. Empirical forecast evaluation of DFMs vs. other many-predictor methods – US data

## References

*Stock, J.H. and M.W. Watson (2006), "Forecasting with Many Predictors," *Handbook of Economic Forecasting*, ch. 10

Stock, J.H. and M.W. Watson (2006), "Implications of Dynamic Factor Models for VAR Analysis," manuscript, Harvard University

Stock, J.H. and M.W. Watson (2006), "An Empirical Comparison of Methods for Forecasting with Many Predictors," manuscript, Harvard University

# 2. VARs and their Limitations

Vector Autoregression (VAR) (Sims, 1980):

$$x_{1t} = A_{11}(L)x_{1t-1} + A_{12}(L)x_{2t-1} + u_{1t}$$
$$x_{2t} = A_{21}(L)x_{1t-1} + A_{22}(L)x_{2t-1} + u_{2t}$$

or

$$X_t = A(L)X_{t-1} + u_t$$

In general, $X_t$ is $n \times 1$ and the VAR has $n$-variables with $p$ lags of each variable in each equation

Drawbacks of VARS

- $pn^2$ parameters – so $n$ cannot be large (6, 9,…)
- Can address dimensionality problem using priors, but most priors are ad-hoc (statistical, not economic)
- mediocre forecasting performance: too many parameters, sensitive to mis-specification in one of the equations

# 3. Dynamic Factor Models

Introduced by Geweke (1975, 1977)

***(a) Key DFM ideas:***

- A handful of structural shocks cause the comovements among macro variables at all leads/lags.
- That is, the economy follows a dynamic factor model
- The "handful" of shocks might be as few as 2!

  Sargent and Sims (1977), Sargent (1989), Quah and Sargent (1992), Stock and Watson (1989, 1999, 2002b), Giannone, Reichlin, and Sala (2004),…
- Recent work on DFMs has focused on large *n* is a blessing:

  Stock and Watson (1999, 2002), Ding and Hwang (2001), Forni, Lippi, Hallin, Reichlin (2001), Bai and Ng (2002, 2004, 2006), Bai (2003),…

## *(b) The dynamic factor model*

$$X_{it} = \lambda_i(\mathrm{L})f_t + u_{it},\ i = 1,\ldots,n,$$
$$\Gamma(\mathrm{L})f_t = \eta_t,$$

$X_{it}$ = $t^{\text{th}}$ observation on $i^{\text{th}}$ observable variable

$f_t$ = unobserved factors, $q \times 1$ ($q$ dynamic factors)

$\lambda_i(\mathrm{L})f_t$ = "common component"

$\lambda_i(\mathrm{L})$ = lag polynomial ("dynamic factor loadings")

$u_{it}$ = idiosyncratic disturbance (possibly serially correlated)

$\mathrm{cov}(f_t, u_{is}) = 0$ for all $i$, $s$

***(c) The exact DFM***:

$Eu_{it}u_{jt} = 0$, $i \neq j$ (idiosyncratic disturbances uncorrelated)

***(d) Spectral factorization***:

$$S_{XX}(\omega) = \lambda(e^{i\omega})S_{ff}(\omega)\lambda(e^{-i\omega})' + S_{uu}(\omega),$$

where $S_{uu}(\omega)$ is diagonal under the exact DFM.

***(e) Estimation when n is small***

$$X_{it} = \lambda_i(\mathrm{L})f_t + u_{it},\ i = 1,\ldots,n,$$
$$\Gamma(\mathrm{L})f_t = \eta_t,$$

This is a linear state space model, so it can be estimated in the time domain by Gaussian MLE using the Kalman filter to compute the likelihood (Sargent (1989), Stock and Watson (1989))

## *(f) Forecasting equation for one variable, $y_t$:*

- Denote one of the $X$'s as $y_t$ (a variable of special interest)
- Suppose $u_{yt}$ follows an autoregression; then

$$y_t = \lambda_y(\mathrm{L})f_t + u_{yt},$$

$$u_{yt} = \gamma(\mathrm{L})u_{yt-1} + \varepsilon_t,\ \varepsilon_t \text{ serially uncorrelated}$$

Then

$$E[y_{t+1} | X_t, y_t, f_t, X_{t-1}, y_{t-1}, f_{t-1}, \ldots] = \beta(\mathrm{L})f_t + \gamma(\mathrm{L})y_t$$

so

$$Y_{t+1} = \beta(\mathrm{L})f_t + \gamma(\mathrm{L})Y_t + \varepsilon_{t+1}$$

**No other *X*'s are needed if the *f*'s are known – optimal forecasts can be made using only lagged *f*'s and lagged *Y***

## *(g) The approximate DFM*

- Recall that the exact DFM assumes that all the idiosyncratic disturbances are uncorrelated: $Eu_{it}u_{jt} = 0$, $i \neq j$
- The approximate DFM relaxes this assumption

  Chamberlain-Rothschild (1983), Stock and Watson (1999, 2002a,b), Forni, Hallin, Lippi, Reichlin (2000, 2003a,b, 2004)
- The general idea is to bound the eigenvalues of $S_{uu}(\omega)$ – the correlations among the $u$'s cannot be "too large"
  - e.g. Stock and Watson (2002a):

$$lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} | E(u_{it}u_{jt}) | < \infty.$$

## *(h) Estimation of the factors by principal components*

When $n$ is large, the factors can be estimated by principal components. The starting point is the static form of the DFM.

Suppose $\lambda(\mathrm{L})$ has degree $p$ and let $F_t = [f_t' \ \ldots \ f_{t-p+1}']'$:

Dynamic form:

$$X_{it} = \lambda_i(\mathrm{L})f_t + u_{it}$$

$$f_t = \Gamma(\mathrm{L})f_{t-1} + \eta_t$$

Static form:

$$X_{it} = \Lambda_i F_t + u_{it} \qquad (1)$$

$$F_t = \Phi(\mathrm{L})F_{t-1} + G\eta_t \qquad (2)$$

where $G$ is $r \times q$; $r = \dim(F_t)$ = number of static factors.

## *DFM estimation by principal components analysis, ctd.*

Static form: $X_t = \Lambda F_t + u_{it}$ ($X_t$ is $n\times 1$, $\Lambda$ is $n\times r$) (1)

By analogy to regression, estimate $\Lambda$ and $\{F_t\}$ by NLLS,

$$\min_{F_1,\ldots,F_T,\Lambda} T^{-1}\sum_{t=1}^{T}(X_t - \Lambda F_t)'(X_t - \Lambda F_t)$$

subject to $\Lambda'\Lambda = I_r$ (identification). Concentrate out $\{F_t\}$:

$$\min_{\Lambda}\ T^{-1}\sum_{t=1}^{T} X_t'[I - \Lambda(\Lambda'\Lambda)^{-1}\Lambda]X_t$$

$\Leftrightarrow \max_{\Lambda}\text{tr}\{(\Lambda'\Lambda)^{-1/2\prime}\ \Lambda'\hat{\Sigma}_{XX}\Lambda(\Lambda'\Lambda)^{-1/2}$ where $\hat{\Sigma}_{XX} = T^{-1}\sum_{t=1}^{T} X_t X_t'$

$\Leftrightarrow \max_{\Lambda}\Lambda'\hat{\Sigma}_{XX}\Lambda$ s.t. $\Lambda'\Lambda = I_r$,

$\Rightarrow \hat{\Lambda}$ = first $r$ eigenvectors of $\hat{\Sigma}_{XX}$

$\Rightarrow \hat{F}_t = \hat{\Lambda}'X_t$ = first $r$ principal components of $X_t$.

## *Distribution Theory for PCA as factor estimator*

- Connor and Korajczyk (1986) (consistency; exact static FM, $T$ fixed, $n \to \infty$)
- Stock and Watson (2002a) (consistency; approximate DFM, $n$, $T \to \infty$, no $n/T$ rate restrictions)
- Bai (2003) (asymptotic normality of PCA estimator of the common component at rate $\min(n^{1/2}, T^{1/2})$; exact DFM,)
- Bai and Ng (2004) (extend Bai (2003) to approximate DFM)
- Bai and Ng (2006) (confidence intervals when estimated factors in subsequent regressions)

***(i) Extension: weighted principal components***.

Infeasible WLS:

$$\min_{F_1,\ldots,F_T,\Lambda} \sum_{t=1}^{T} (X_t - \Lambda F_t)' \Sigma_{uu}^{-1} (X_t - \Lambda F_t).$$

Solution: $\hat{\Lambda}$ = first $q$ eigenvectors of $\Sigma_{uu}^{-1/2} \hat{\Sigma}_{XX} \Sigma_{uu}^{-1/2}$,

***Feasible weighted PCA***:

(a) Forni et. al. (2004): $\hat{\Sigma}_{uu} = \hat{\Sigma}_{XX} - \hat{\Sigma}_{cc}$,

where $\hat{\Sigma}_{cc}$ is estimate of covariance matrix of the common component in the DFM, estimated by dynamic PCA (Forni et. al. (2003b)

(b) Bovin and Ng (2005): $\hat{\Sigma}_{uu}^{diag} = \text{diag}(\hat{\Sigma}_{uu})$

(this accords with exact DFM restrictions)

## *(k) Estimation of the number of factors*

- Number of static factors ($r$):
  - Bai and Ng (2002) (information criterion applied to eigenvalues of $X'X$, approximate DFM)
  - Onatski (2005) – formal test for number of static factors based on eigenvalues of $X'X$ for number of nonzero eigenvalues (number of principal components to include)
- Number of dynamic factors ($q$):
  - Giannoni, D., L. Reichlin and L. Sala (2004) – heuristic methods based on inspection of eigenvalues of residuals of VAR for $F_t$ (static factors)
  - Amengual and Watson (2006) – extend Bai-Ng to estimate the number of dynamic factors ($q$) by applying information criterion to covariance matrix of residuals from VAR for $F_t$

# 4. An Empirical DFM with U.S. Data

***The data***

- $n$ = 132, postwar monthly US
  - real activity
  - prices
  - interest rates and spreads
  - exchange rates
  - stock returns
  - misc
- All transformed to "stationarity" by first differencing, logs, etc

***Base specification***

VAR(2) for $F_t$, 6 lags for $\delta$(L)

***(a) Estimates of the Number of Factors***

No. of static factors (Bai-Ng $ICP_2$)**:** $r = 9$

No. of dynamic factors (Amengual-Watson $ICP_2$)**:** $q = 7$

Comments

- Sargent-Sims (1977) etc. focused on output and prices only – for which 2 or 3 are plausible – we have a much richer data set and find factors other than output and price factors
- What are the factors?

  #1: real variables (93% of IP)

  #2 and #3: price inflation (66% of CPI inflation)

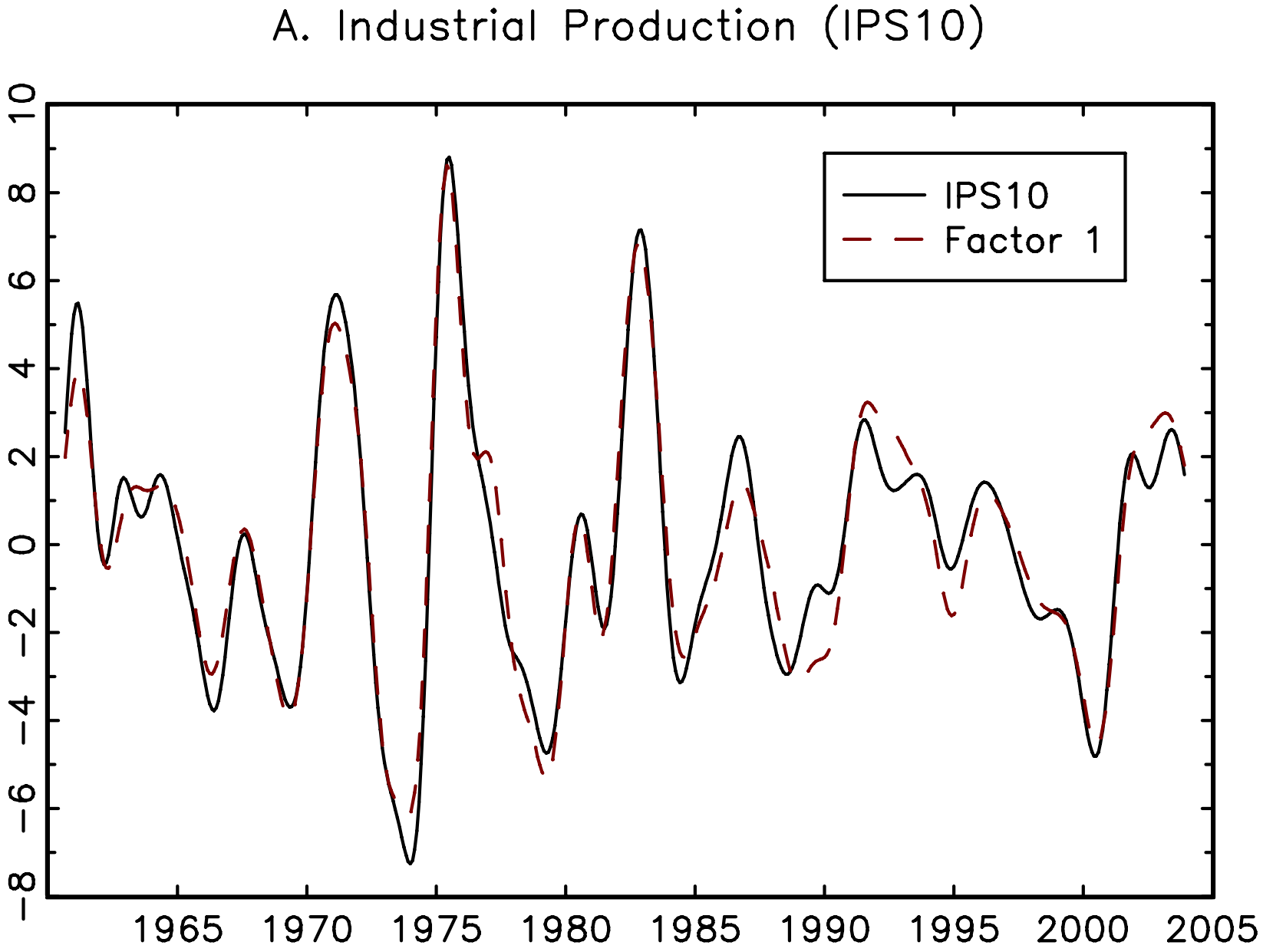
  #4: long-term interest rates (31% of 10-yr T-bond)

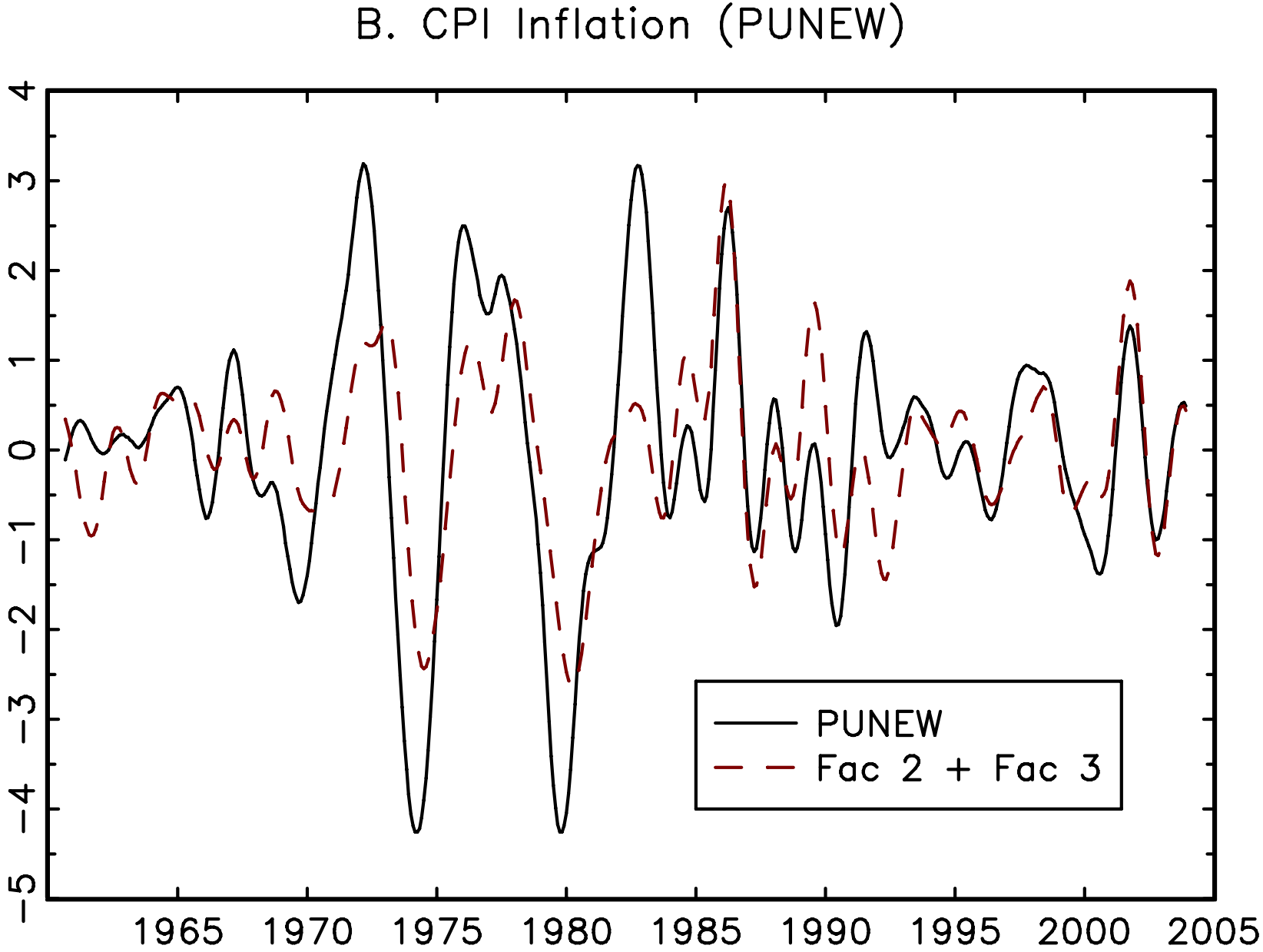
  #5: long-term unemployment (31% of mean duration)

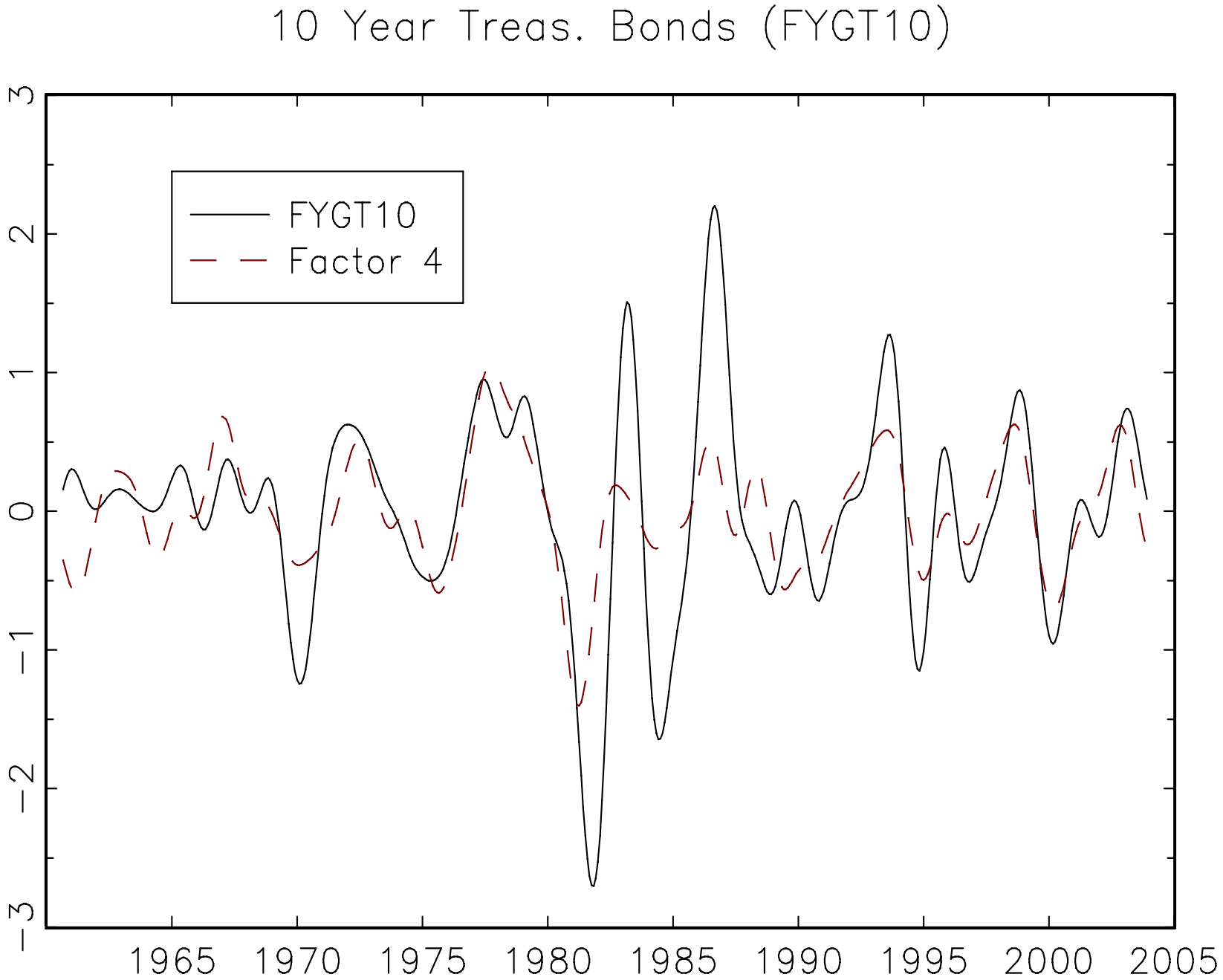
  #6: stock returns, exchange rates (12% of S&P 500)

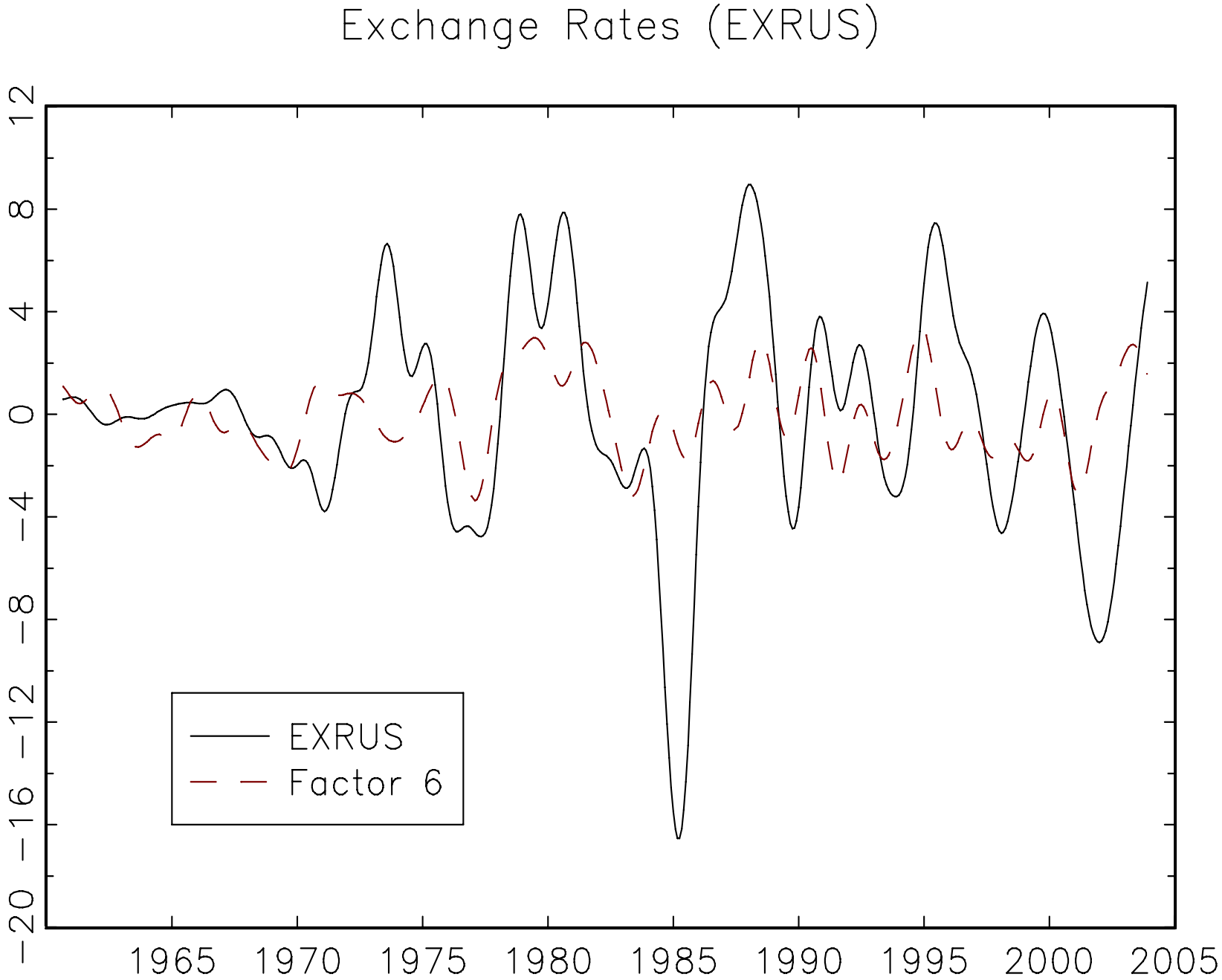
  #7: exchange rates, little else (28% of trade-weighted)

***Some examples….business cycle components:***


A. Industrial Production (IPS10)
IPS10
Factor 1
10
8
6
4
2
0
−2
−4
−6
−8
1965
1970
1975
1980
1985
1990
1995
2000
2005


B. CPI Inflation (PUNEW)
4
3
2
1
0
−1
−2
−3
−4
−5
1965
1970
1975
1980
1985
1990
1995
2000
2005
PUNEW
Fac 2 + Fac 3


10 Year Treas. Bonds (FYGT10)
FYGT10
Factor 4
3
2
1
0
−1
−2
−3
1965
1970
1975
1980
1985
1990
1995
2000
2005


Exchange Rates (EXRUS)
12
8
4
0
−4
−8
−12
−16
−20
1965
1970
1975
1980
1985
1990
1995
2000
2005
EXRUS
Factor 6

## *(b) Test of exact DFM restriction: $X_j$ does not predict $X_i$ given $F_{t-1}$*

This can be tested in a VAR framework:

$$X_{it} = \Lambda_i \Phi(\mathrm{L}) F_{t-1} + \delta_i(\mathrm{L}) X_{it-1} + \delta_{ij}(\mathrm{L}) X_{jt-1} + \varepsilon_t \qquad (*)$$

$$H_0: \delta_{ij}(z) = 0, j = 1,\ldots, 132, j \neq i$$

- 6 restrictions (6 lags) for each $j$
- Total # restrictions = $6 \times (132^2 - 132) = 103{,}752$ (!)

Results:

- There are more rejections at the 5% level than one would expect by random sampling under the null
- However, these rejections are (almost) entirely associated with small marginal $R^2$'s – not economically large.
- In general, the predictive content of $X$'s is greatly reduced (or eliminated) by including $F_{t-1}$ in the forecasting equation (*)

# 5. Theory of Forecasting with Many Predictors

***(a) Optimal forecasting in the i.i.d. Gaussian/strictly exogenous model:***

$$Y_{t+1} = \delta' P_t + \varepsilon_{t+1}, \; t = 1, \ldots, T$$

$Y_{t+1}$ = scalar

$P_t$ = $n$ orthonormal predictors (principal components) so $P'P/T = I_n$

***Suppose (for now) that $P_t$ is strictly exogenous, $\varepsilon_{t+1}$ i.i.d. $N(0, \sigma^2)$***

Well-known results from classical statistics:

- If $n \geq 3$, OLS is inadmissible
- OLS is dominated by shrinkage estimators (James-Stein)
- What is the best shrinkage estimator to use?
- Bayes estimators are obvious candidates

## *The i.i.d. Gaussian/strictly exogenous model, ctd.*

$$Y_{t+1} = \delta' P_t + \varepsilon_{t+1}, t = 1, \ldots, T$$

- $P_t$ strictly exogenous, $\varepsilon_{t+1}$ i.i.d. $N(0,\sigma^2)$ – ***deal with this later***
- Asymptotics: let $n/T \to 0 < c < 1$, $\delta_i = d_i/T^{1/2}$
- Squared error forecast loss $\Rightarrow L(\tilde{\delta}) = \text{tr}[(\tilde{\delta} - \delta)(\tilde{\delta} - \delta)']$
  - Motivation: consider forecast risk,

$$E(Y_{T+1} - \tilde{Y}_{T+1|t})^2 = E[(\tilde{\delta} - \delta)P_T + \varepsilon_{T+1}]^2$$
$$= E\text{tr}[(\tilde{\delta} - \delta)(\tilde{\delta} - \delta)' P_T P_T'] + \sigma^2$$
$$\approx E\text{tr}[(\tilde{\delta} - \delta)(\tilde{\delta} - \delta)'] + \sigma^2$$

- The only part we can work on is $E\text{tr}[(\tilde{\delta} - \delta)(\tilde{\delta} - \delta)']$
- Consider equivariant estimators under permutations of $P_t$

## *The i.i.d. Gaussian/strictly exogenous model, ctd.*

Frequentist risk for permutation equivariant estimators:

$$R(\tilde{\delta},\delta) = \sum_{i=1}^{n} E(\tilde{\delta}_i - \delta_i)^2 \qquad \text{(trace MSE loss)}$$

$$= \left(\frac{n}{T}\right) n^{-1} \sum_{i=1}^{n} E(\tilde{d}_i - d_i)^2 \qquad \text{(local to zero)}$$

$$= c \int E(\tilde{d} - d)^2 \, dG_n(d) \qquad \text{(permutation equivariance,}$$

$G_n$ = empirical cdf of $d_i$'s)

$$= R_{G_n}(\tilde{d}) \qquad \text{(Bayes risk of estimator } \tilde{d}$$

w.r.t. $G_n$)

The frequentist risk for permutation equivariant estimators is the Bayes risk wrt the empirical cdf of the $d$'s, $G_n$

***(b) Empirical Bayes heuristics***

Frequentist problem:

$$\min_{\tilde{\delta}} R(\tilde{\delta}, G_n) = c\int E(\tilde{d} - d)^2 \, dG_n(d) \qquad \text{cdf of } d_i$$

Bayes problem:

$$\min_{\tilde{\delta}} R(\tilde{\delta}, G) = c\int E(\tilde{d} - d)^2 \, dG(d) \qquad \text{subjective prior}$$

Empirical Bayes problem:

$$\min_{\tilde{\delta}} R(\tilde{\delta}, \hat{G}) = c\int E(\tilde{d} - d)^2 \, d\hat{G}(d) \qquad \text{estimated prior}$$

Empirical Bayes: under technical conditions,

- asymptotically admissible, asy. optimal (Robbins (1964))
- has certain minimax properties (Zhang, AS (2005))
- $\hat{G}$ can be nonparametric or parametric (e.g. BMA)
- asymptotically, EB is minimum risk equivariant (Edelman (1988), Knox, Stock, Watson (2001))

## *(c) Relax the i.i.d. Gaussian and strict exogeneity assumptions*

- How to extend this to $P_t$ predetermined, not strictly exogenous?
- How to extend to multistep forecasting?
- The Bayes derivation breaks down under these conditions
- Empirical question: Is there any gain to using the remaining 127 factors?

<u>Proposed approach</u>

1. Provides a common shrinkage representation for Bayes, EB, and some other methods – *for predetermined data and multistep forecasts*
2. Empirical comparison of several methods for forecasting 9 monthly U.S. macro time series using 131 predictors

***(d) Shrinkage representations for optimal linear forecasts***

**Bayes, EB estimators (also, BMA, bagging, pretest) have a "shrinkage representation:"** Suppose $P'P/T = I_n$. Then

$$\hat{Y}_{t+1} = \sum_{i=1}^{n} \psi(t_i)\hat{\delta}_i P_{it} + o_p(1)$$

where $\hat{\delta}$ = OLS estimator of $\delta$, $s_e^2 = \sum_{t=1}^{T} (Y_{t+1} - \hat{\delta}' P_t)^2 / (T-n)$, and $t_i = \sqrt{T}\,\hat{\delta}_i / s_e$.

- $0 \leq \psi(x) \leq 1$ – hence "shrinkage" terminology
- The $\psi$ function is a property of the estimation algorithm
- The representation holds under general conditions on true DGP
- Think of this exercise in the same way as "pseudo-ML" – here, we are doing "pseudo-BMA"
- This representation allows us to study the performance of the procedure when the modeling assumptions are false

## Example 1: Bayes and Empirical Bayes

Modeling assumptions:

- $P_t$ strictly exogenous; $\varepsilon_t$ i.i.d. Normal
- $\delta_i$, $i = 1,\ldots, n$ are iid with prior distribution $G$

Posterior mean can be written in "simple" Bayes form as

$$\hat{\delta}_i^B \mid \sigma^2 = \frac{\int x\phi_{\sigma/\sqrt{T}}(\hat{\delta}_i - x)\mathrm{d}G(x)}{\int \phi_{\sigma/\sqrt{T}}(\hat{\delta}_i - x)\mathrm{d}G(x)} = \hat{\delta}_i + \frac{\sigma^2}{T}\ell(\hat{\delta}_i)$$

$\ell(x) = \mathrm{d}\ln(m(x))/\mathrm{d}x$, where $m(x) = \int \phi_{\sigma/\sqrt{T}}(x-\delta)dG(\delta)$ is the marginal distribution of an element of $\hat{\delta}$.

Using the "simple Bayes" formula, given $\sigma^2$

$$\hat{\delta}_i^B|\sigma^2 = \hat{\delta}_i + \frac{\sigma^2}{T}\ell(\hat{\delta}_i)$$

$$= (1 + \frac{\sigma^2}{T}\frac{\ell(\hat{\delta}_i)}{\hat{\delta}_i})\hat{\delta}_i$$

$$= \psi^B(\hat{\tau}_i)\,\hat{\delta}_i,$$

where, by change of variables with $\hat{\tau}_i = \sqrt{T}\hat{\delta}_i/\sigma$,

$$\psi^B(z) = 1 + \tilde{\ell}(z)/z$$

$$\tilde{\ell}(z) = \frac{\mathrm{dln}\tilde{m}(z)}{\mathrm{dz}},\ \tilde{m}(z) = \int \phi(z-\tau)dG_\tau(\tau)$$

$G_\tau$ is a prior defined over $\tau = \sqrt{T}\,\delta/\sigma$.

- Note that this representation is a consequence of the modeling assumptions ($G_\tau$) – *not the true dgp*

*Normal Bayes – integration over posterior*

Next, integrate over the posterior. Then,

$$\hat{\delta}_i^B = E_\sigma[(1 + \frac{\sigma^2}{T}\frac{\ell(\hat{\delta}_i)}{\hat{\delta}_i})|\hat{\delta},\hat{\sigma}^2]\hat{\delta}_i$$

***Empirical Bayes Strategy***

- Use $\{\hat{\delta}_i\}$ to estimate $\tilde{\ell}(z)$ using a parametric or nonparametric estimator – call this $\hat{\tilde{\ell}}(z)$
- Substitute this into the formula for $\psi^B$:

$$\psi^B(z) = 1 + \hat{\tilde{\ell}}(z)/z$$

- Then $\hat{\delta}_i^B = \psi^B(\hat{\tau}_i)\hat{\delta}_i$

## Example 2: Pretest Methods

Pretest estimators include a variable (with the OLS coefficient) if the OLS coefficient exceeds a constant – another term for this is "hard thresholding."

- Because the regressors are orthogonal, using "hard threshold" on the *t*-statistic for model selection is equivalent to including those regressors that have *t*-statistics exceeding a certain threshold.
- If $n$ is fixed and coefficients are local to zero, AIC is asymptotically equivalent to hard threshold *t*-statistic pretest
- For these methods, the $\psi$ function is,

$$\psi^{JC}(\tau) = \mathbf{1}(|\tau| \geq c) \quad (AIC\text{:}\ \ c = \sqrt{2})$$

## Example 3: Bagging

Breiman (1996); Inoue and Kilian (2004), Lee and Yang (2004)

- Start with hard threshold:

$$\hat{\delta}_i^{PT} = \psi^{PT}(t_i)\hat{\delta}_i, \text{ with } \psi^{IC}(\tau)=\mathbf{1}(\tau \geq c)$$

- Bagging: "soften" the threshold by averaging over bootstrap replications of hard threshold estimator.
- Asymptotic form of resulting estimator (Bühlmann and Yu (2002)):

$$\hat{\delta}_i^{Bagging} \approx \mathrm{E}(x \mid x^2 > \sigma^2 c^2), \text{ where } x \sim \mathrm{N}(\hat{\delta}_i, \sigma^2/T)$$

which implies

$$\hat{\delta}_i^{Bagging} \approx \psi^{Bagging}(t_i)\,\hat{\delta}_i$$

where $\psi^{Bagging}(\tau) = \dfrac{\phi(c-\tau)-\phi(-c-\tau)}{\tau} + 1 - \Phi(c-\tau) + \Phi(-c-\tau)$

***(e) Formal results (validity of the shrinkage representation)***

(1) Normal Bayes: If $|p_{iT}| \le p_{max}$ and

(*i*) posterior for $\sigma$ concentrates around $\hat{\sigma}^2$

(*ii*) score function is sufficiently smooth

(*iii*) moments of *t*-statistic and $\hat{\sigma}_Y^2$ exist; then

$$E\left[\hat{Y}_{T+1|T}^{NB} - \sum_{i=1}^{n} \psi^{NB}(\kappa t_i)\hat{\delta}_i p_{iT}\right]^2 \to 0$$

where $\kappa = (1 - n/T)^{-1/2}$, $\psi^{NB}(t) = 1 + \dfrac{\ell(t)}{t}$, $\ell(t) = \dfrac{\mathrm{d}\ln m(t)}{\mathrm{d}t}$,

$m(t) = \int \phi(t-\tau)dG_\tau(\tau)$, and $K_1$, $K_2$, $K_3$, $M$ depend on the prior, and

$0 \le \psi^{NB}(t) \le 1$, and if $g$ is symmetric, $\psi^{NB}(t) = \psi^{NB}(|t|)$.

(2) Bagging. For all $T$, $n$ s.t. for $r = T - n > 8$,

$$E\left[\hat{Y}^{BG}_{T+1|T} - \sum_{i=1}^{n} \psi^{BG}(t_i)\hat{\delta}_i p_{iT}\right]^2 \rightarrow 0,$$

where

$$\psi^{BG}(t) = 1 - \Phi(t + c) + \Phi(t - c) + t^{-1}[\phi(t - c) - \phi(t + c)],$$

***These results make no assumption about the DGP – in particular it does not require strict exogeneity or Gaussian errors – that is, the original model (whereby the estimator is derived) can be mis-specified.***

## *(f) Leading example: Bayesian model averaging with orthogonal regressors*

Clyde, Desimone, and Parmigiani (1996), Clyde(1999a,b), Koop and Potter (2003), Wright (2004a,b)

BMA modeling assumption:

$$\delta_i|\sigma \sim \begin{cases} N(0, \sigma^2/g) & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

then

$$\psi^{BMA}(t_i) = \frac{pb(g)\phi(b(g)t_i)}{(1+g)[pb(g)\phi(b(g)t_i) + (1-p)\phi(t_i)]}$$

where $b(g) = \sqrt{g/(1+g)}$, $\omega^2 = \sigma^2/gT$, and $\phi$ is normal pdf.

***(g) Comments on shrinkage representations:***

1. These representations are consequences of the algorithm + weak assumptions (moments) on the true DGP
2. They tell us what the algorithm does mechanically when the strong assumptions of the derivation fail
   - "*pseudo BMA*" – analogous to pseudo-ML
   - weak exogeneity
   - serially correlated errors – direct multistep forecasting
3. For strictly exogenous $X$, these results extend Bühlmann and Yu (2002) from fixed $n$ to $n/T \to c > 0$
4. We can ask whether non-Bayes methods (e.g. bagging) are admissible in the exogenous $X$/Gaussian model.

***Comments, ctd.***

5. Shrinkage representation provides a justification for direct estimation of flexible forms of $\psi$ (rather than indirect via EB estimation of prior $G$). In preliminary work we use the logistic function, $\psi(t) = [1 + \exp(-\beta_0 + \beta_1|t|)]^{-1}$
6. Estimation of $\psi$ function parameters
   - NLLS? No: this leads to OLS including all regressors if the $\psi$ function nests $\psi(t) = 1$. E.g. for logistic, $\beta_1 = 0$, $\beta_0 \rightarrow -\infty$; for BMA, $p = 1$ with no shrinkage.
   - In the empirical work we estimate $\psi$ parameters by predictive least squares (PLS)

# 6. Empirical Forecast Evaluation of DFMs vs. Other Many-Predictor Methods with US Data

*Data:*

- 131 monthly U.S. macro time series from 1959:1 – 2003:12.
- 9 of these variables are forecasted.

*General Form of forecasting model and data transformations:*

- "Direct" forecasting
- General form of model used for forecasting at horizon $h$ :

$$Y_{t+h}^{h} = \alpha + \sum_{i=1}^{n} \beta_i X_{t,i} + \sum_{i=1}^{p} \phi_i Y_{t+1-i} + u_{t+h}$$

  - $Y_{t+h}^{h}$ = transformed value of the variable being forecast
  - $Y_{t-i}$ = autoregressive lags
  - $X_{t,i}$ denotes the $i^{\text{th}}$ predictor variable ***or*** $P_{t,i}$
- Transformations: logarithms and differencing, as appropriate

# Series being forecasted and their transformations

| Series | | $Y_{t+h}^{h}$ | $Y_t$ |
|---|---|---|---|
| Personal Income | PI | $(1200/h)\ln(Z_{t+h}/Z_t)$ | $\Delta\ln(Z_t)$ |
| Ind. Production | IP | $(1200/h)\ln(Z_{t+h}/Z_t)$ | $\Delta\ln(Z_t)$ |
| Unemployment | UR | $(Z_{t+h} - Z_t)$ | $\Delta Z_t$ |
| Employment | EMP | $(1200/h)\ln(Z_{t+h}/Z_t)$ | $\Delta\ln(Z_t)$ |
| 3-Mth Tbill Rate | TBILL | $(Z_{t+h} - Z_t)$ | $\Delta Z_t$ |
| 10-Yr TBond Rate | TBOND | $(Z_{t+h} - Z_t)$ | $\Delta Z_t$ |
| Prod. Price Index | PPI | $1200[(1/h)\ln(Z_{t+h}/Z_t) - \Delta\ln(Z_t)]$ | $\Delta^2\ln(Z_t)$ |
| Cons. Price Index | CPI | $1200[(1/h)\ln(Z_{t+h}/Z_t) - \Delta\ln(Z_t)]$ | $\Delta^2\ln(Z_t)$ |
| PCE Deflator | PCED | $1200[(1/h)\ln(Z_{t+h}/Z_t) - \Delta\ln(Z_t)]$ | $\Delta^2\ln(Z_t)$ |

## *Pseudo-out-of-sample forecasts – some details*

- First Estimation Period 1960:1
- Each in-sample estimated regression is restricted to contain a minimum of 120 observations
- For regressions involving all regressors the minimum number of in-sample regression observations is 130/.75 = 174
- Forecast period is 1974:7. 2003:12 – $h$.

# Summary of Forecasting Methods for Empirical Comparison

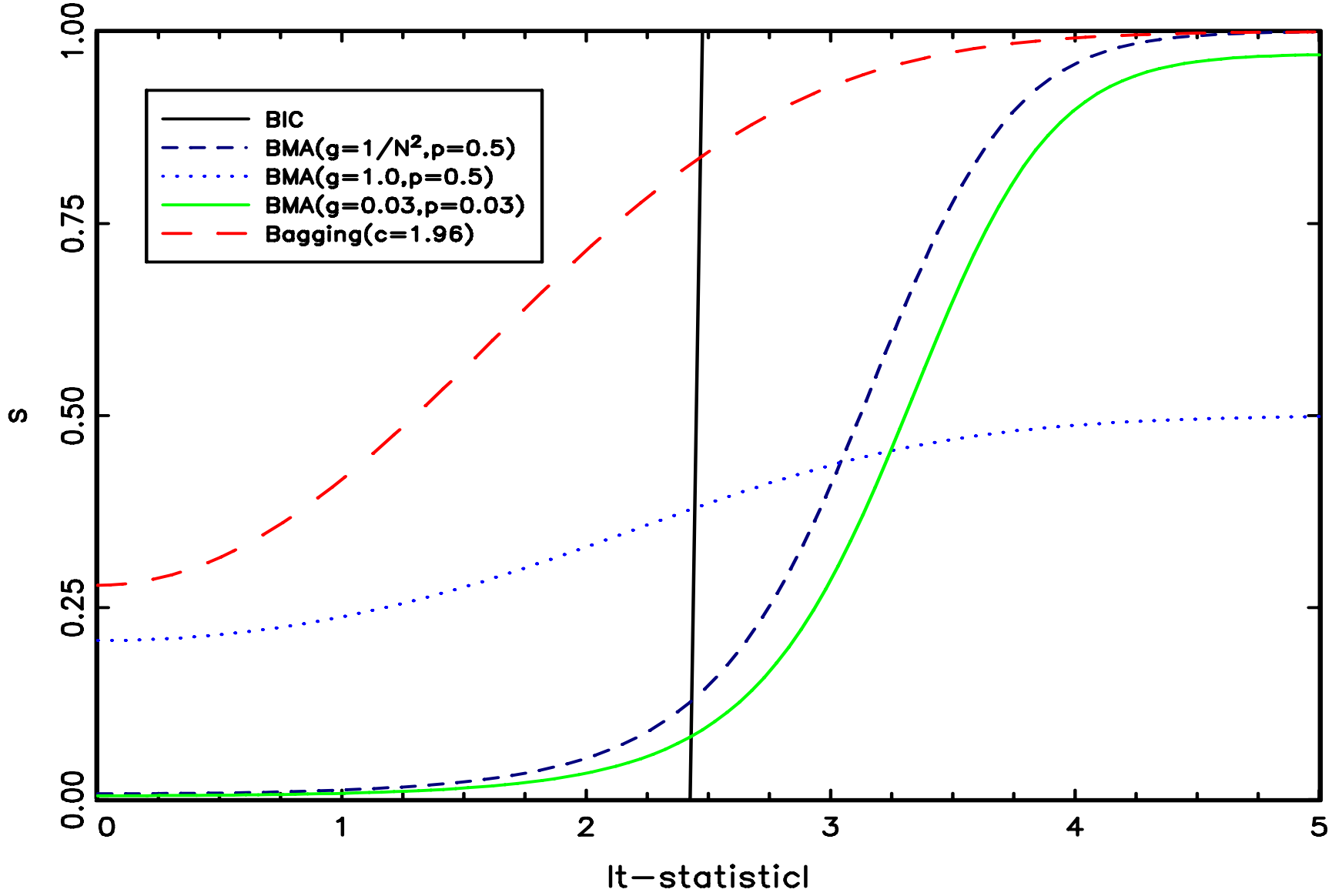
| Method | Description |
|---|---|
| Combined-Mean | Combined ADL Models, AIC Lag Selection, sample mean . |
| AR | AR Model, AIC Lag Selection |
| OLS | All $X$ Variables, $p_Y = 4$, all coefficients estimated by OLS |
| Combined-SSR | Combined ADL Models, $p_Y = 4$, $\alpha$ chosen by PLS |
| FAAR-OLS | Factor Augmented AR model, OLS estimation of Factors (PC), AIC selection of factors and AR lags |
| FAAR-GLS | Factor Augmented AR model, GLS estimation of Factors (PC), AIC selection of factors and AR lags |
| FAAR-WLS | Factor Augmented AR model, WLS estimation of Factors (PC), AIC selection of factors and AR lags |

# Forecasting methods, ctd.

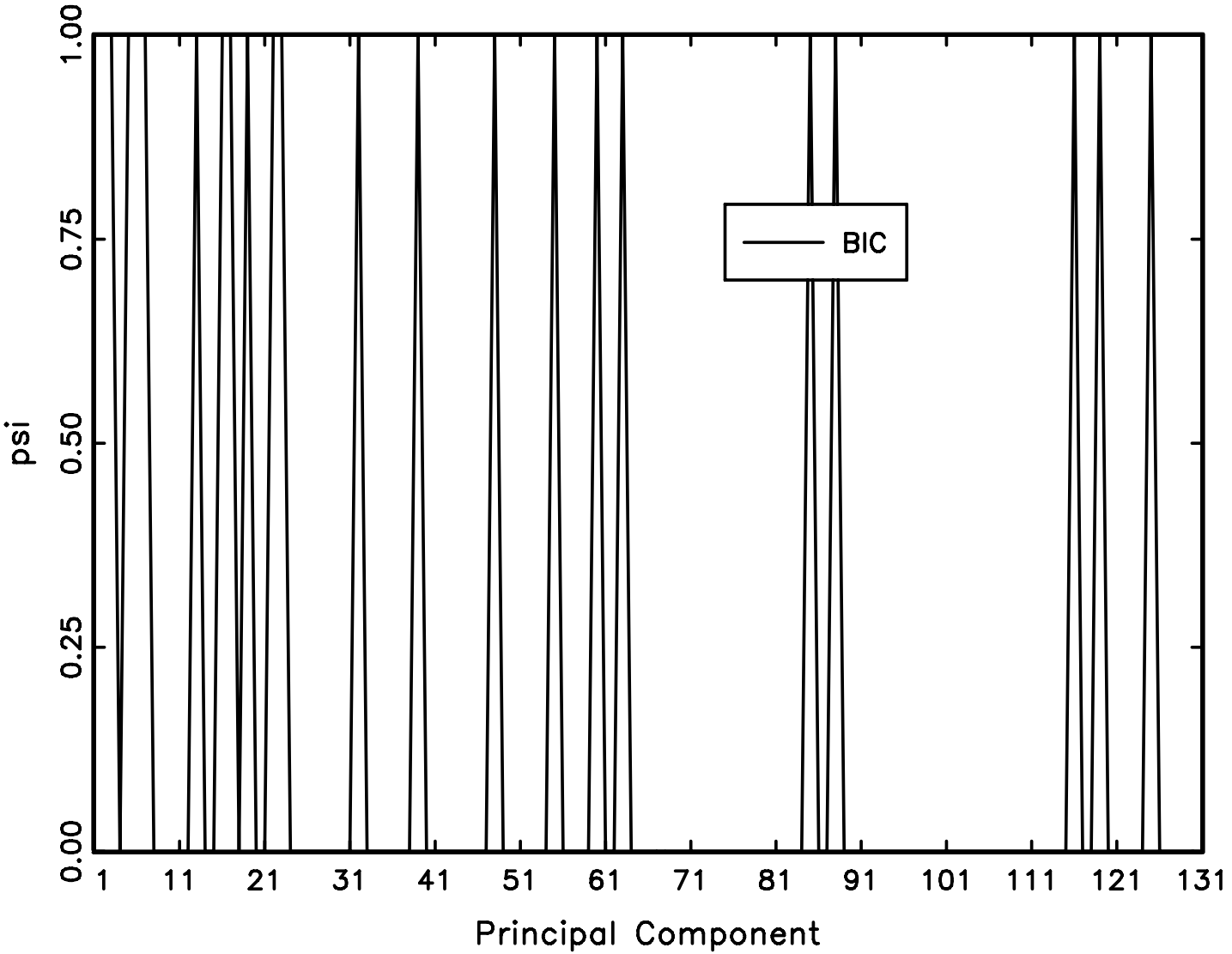
| | |
|---|---|
| BMA($1/n^2$,0.5) | BMA using $X$ , $p_Y = 4$, $g = 1/n^2$, $p = 0.5$ |
| BMA(1,0.5) | BMA using $X$ , $p_Y = 4$, $g = 1$, $p = 0.5$ – informative prior |
| BMA-PC($1/n^2$,0.5) | BMA using PC, $p_Y = 4$, $g = 1/n^2$, $p = 1/2$ – uninformative |
| BMA-PC(1,0.5) | BMA using PC, $p_Y = 4$, $g = 1$, $p = ½$ – informative |
| PEB-PC | BMA using PC, $p_Y = 4$, EB estimates of $g$ and $p$: estimate $g = .03$ (wide spread), $p = .03$ (rare) for $h$=6 |
| SNP | simple nonparametric empirical Bayes (kernel estimator of the score of $m$) |
| BIC-PC | PC using BIC selection, $p_Y = 4$ |
| Bagging-PC | Bagging using PC, $c = 1.96$ with Newey-West $t$-statistics, $p_Y = 4$ (will discuss PLS-estimated $c$ also) |

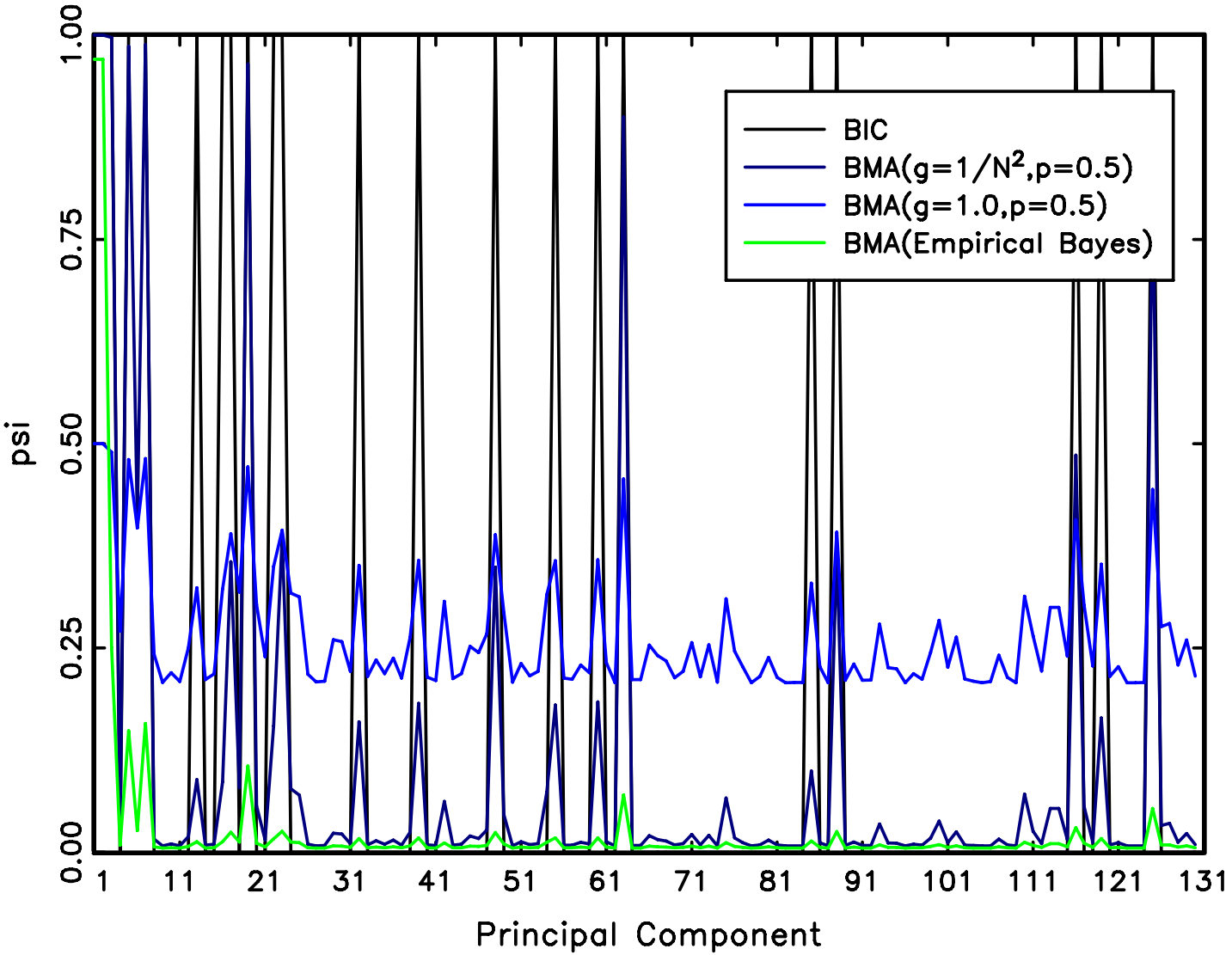
- $h = 1, 3, 6, 12$ months; 9 series forecasts
- ***All results are MSFEs, relative to Combined ADL Mean***

# Shrinkage Factors for PC Forecasting Models



# Shrinkage factors for each PC: Unemployment Rate




BIC
BMA(g=1/N2,p=0.5)
BMA(g=1.0,p=0.5)
BMA(Empirical Bayes)
psi
0.00
0.25
0.50
0.75
1.00
1
11
21
31
41
51
61
71
81
91
101
111
121
131
Principal Component

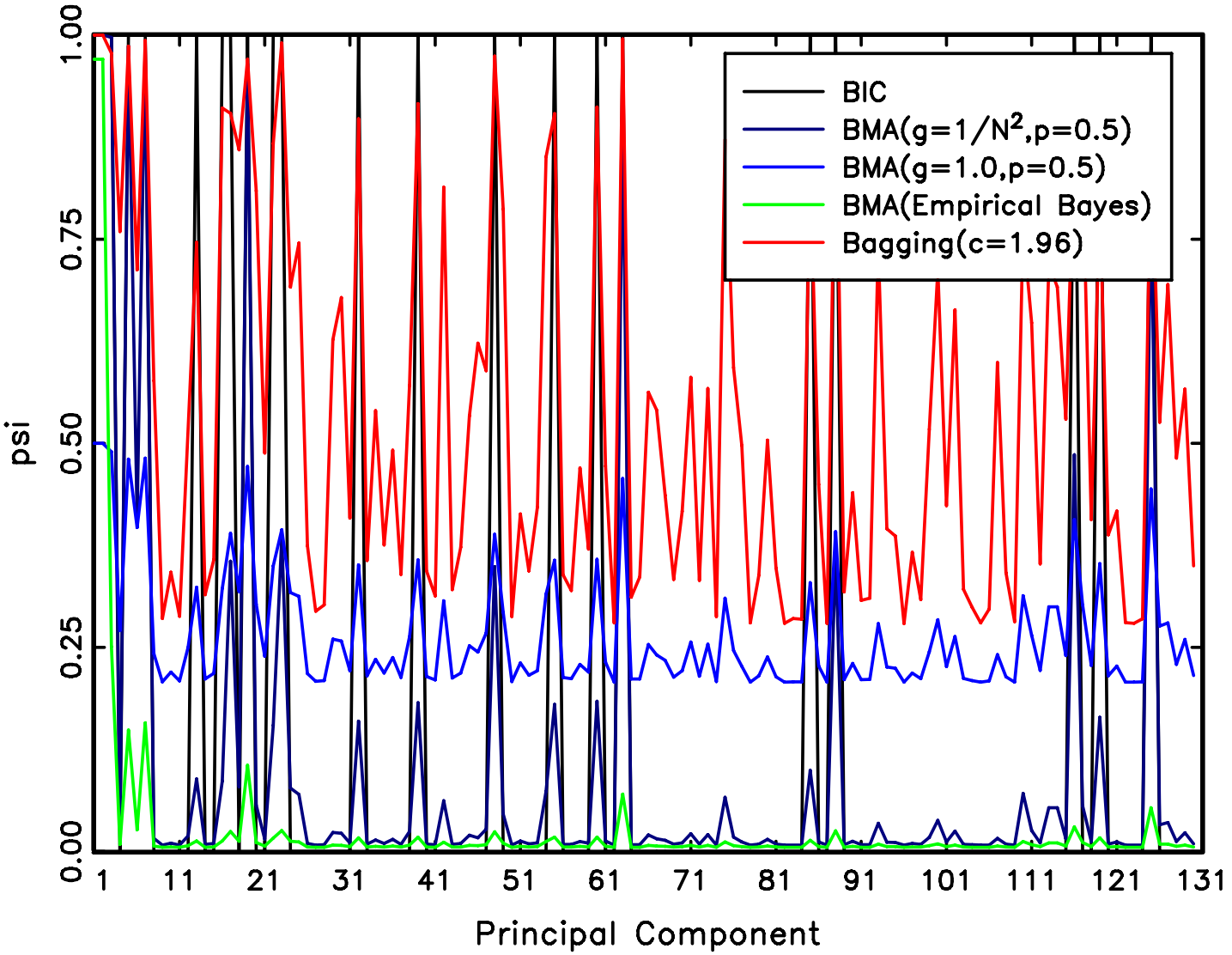

BIC
BMA(g=1/N2,p=0.5)
BMA(g=1.0,p=0.5)
BMA(Empirical Bayes)
Bagging(c=1.96)
psi
0.00
0.25
0.50
0.75
1.00
1
11
21
31
41
51
61
71
81
91
101
111
121
131
Principal Component

**Table A. MSFEs relative to simple combination: Unemployment Rate**

| | $h = 1$ | $h = 3$ | $h = 6$ | $h = 12$ |
|---|---|---|---|---|
| AR | 1.07 | 1.13 | 1.20 | 1.21 |
| OLS | 1.83 | 1.53 | 1.75 | 2.07 |
| Combined-SSR | 0.90 | 0.90 | 1.01 | 1.05 |
| FAAR-OLS | 0.86 | 0.85 | 0.93 | 0.99 |
| FAAR-GLS | 0.95 | 0.84 | 0.84 | 0.93 |
| FAAR-WLS | 0.86 | 0.86 | 0.92 | 0.92 |
| BMA($1/n^2$,0.5) | 0.88 | 0.88 | 1.19 | 1.47 |
| BMA(1,0.5) | 0.87 | 0.84 | 1.01 | 1.27 |
| BMA-PC($1/n^2$,0.5) | 0.87 | 0.83 | 0.97 | 1.05 |
| BMA-PC(1,0.5) | 0.97 | 0.91 | 0.95 | 1.04 |
| PEB-PC | 0.86 | 0.82 | 0.93 | 0.99 |
| BIC-PC | 0.97 | 0.96 | 1.15 | 1.45 |
| Bagging-PC ($c$=1.96) | 1.16 | 1.07 | 1.23 | 1.54 |

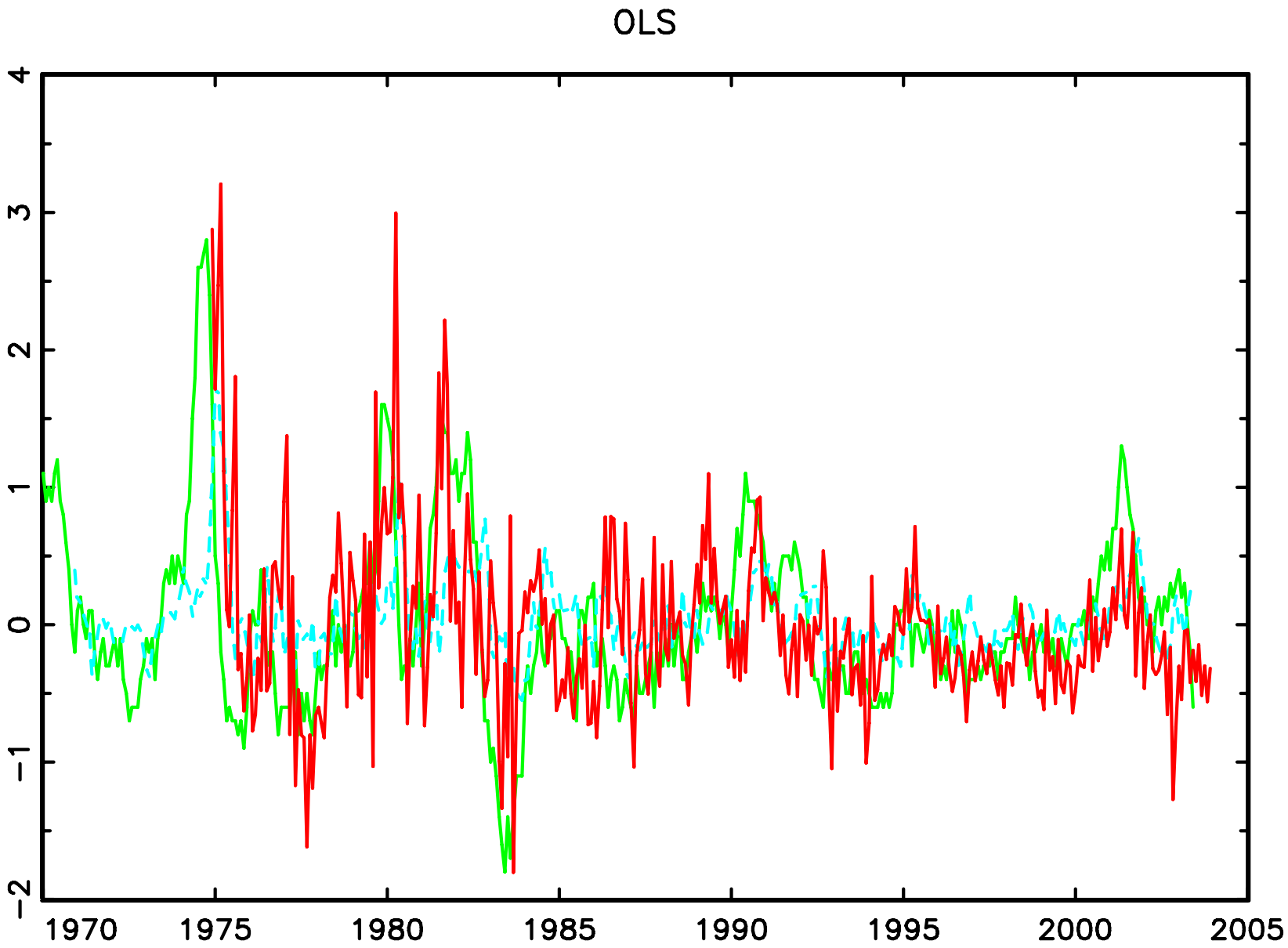
cf. Boivin-Ng (2005) – other PC methods, standard PC works well

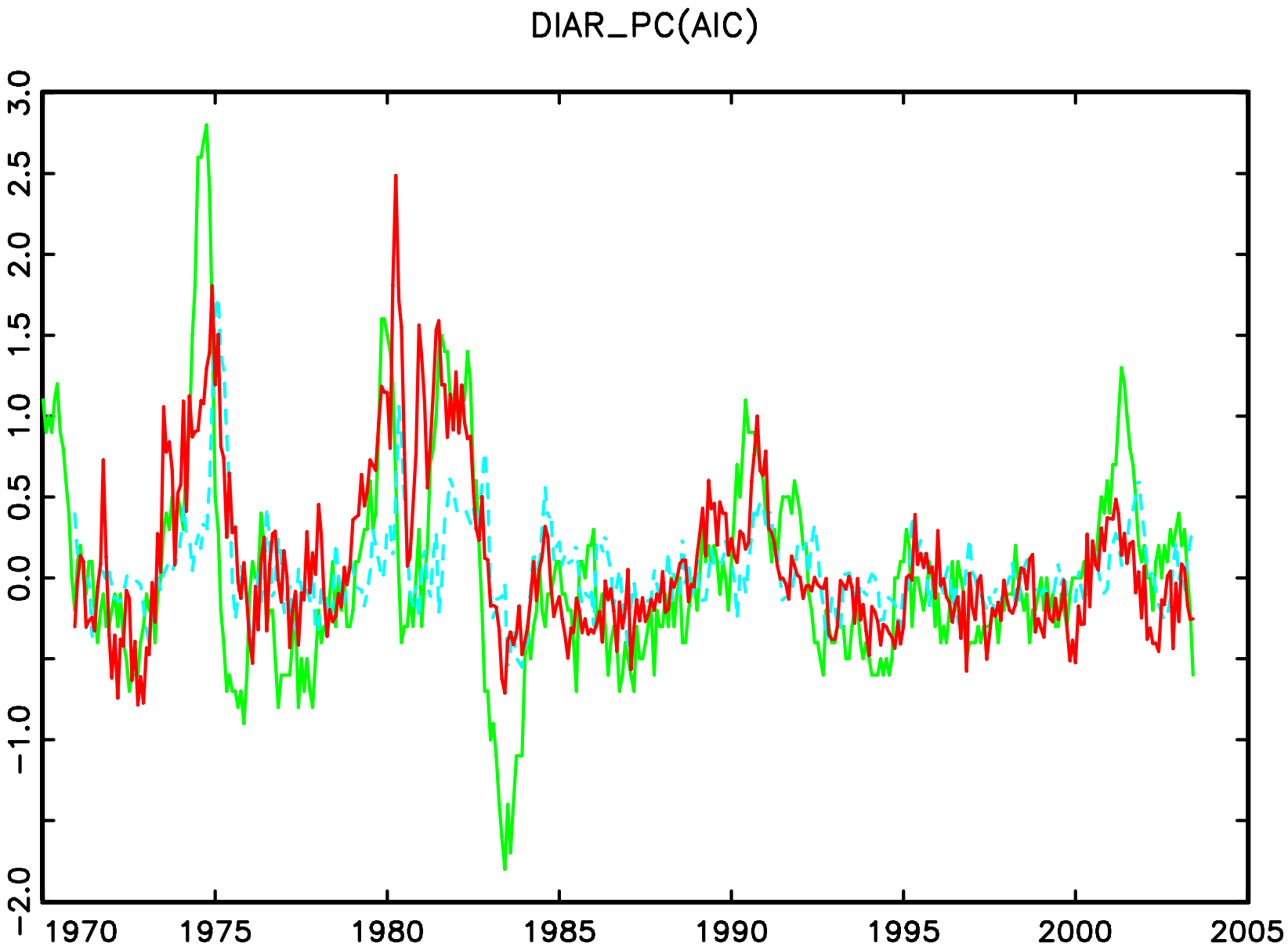
## Plots of unemployment rate forecasts

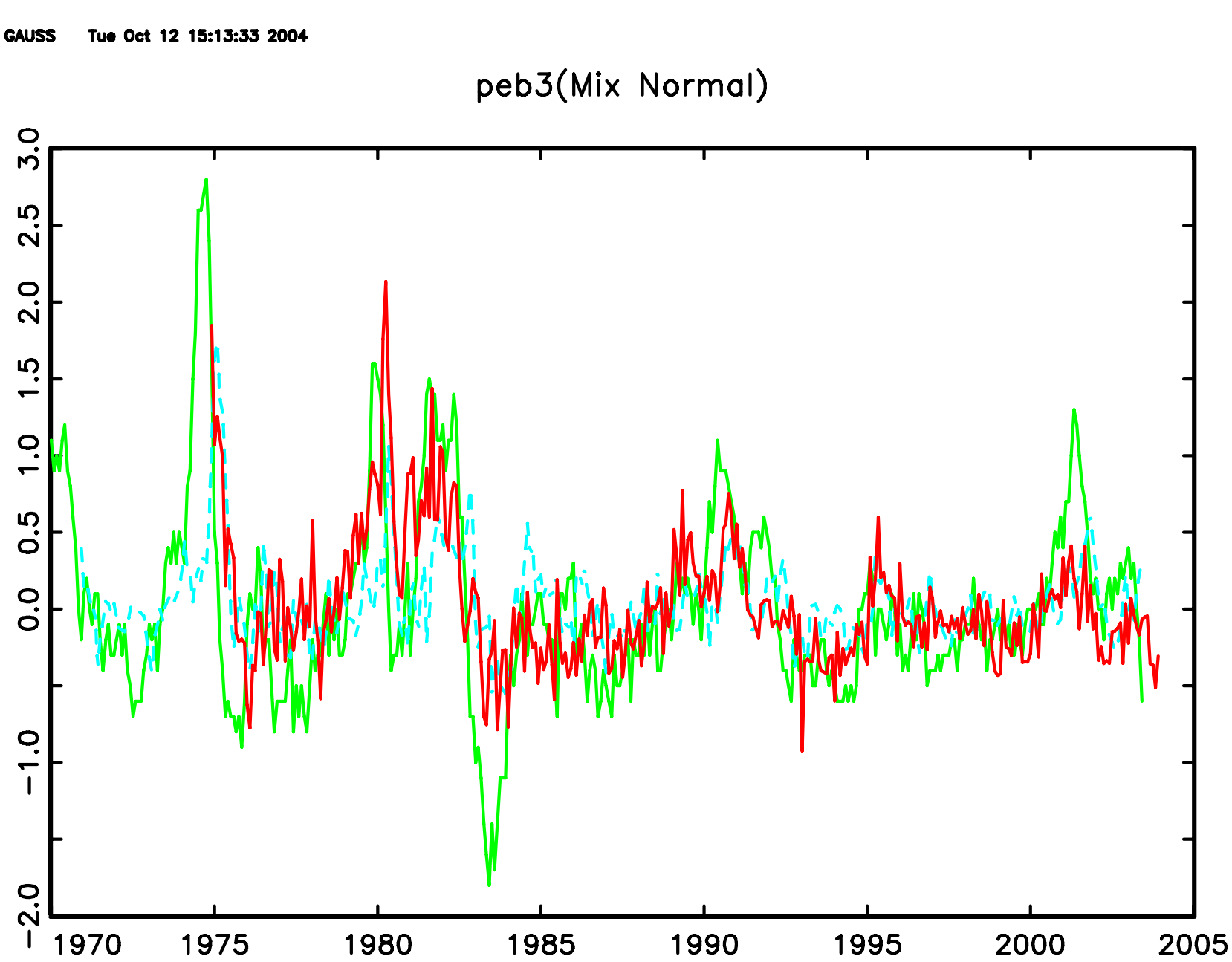
Green = unemployment rate

Blue = AR(AIC) **6-month** ahead forecast

Red = Candidate **6-month** ahead forecast


OLS
4
3
2
1
0
−1
−2
1970
1975
1980
1985
1990
1995
2000
2005


DIAR_PC(AIC)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−1.0
−2.0
1970
1975
1980
1985
1990
1995
2000
2005


GAUSS Tue Oct 12 15:13:33 2004
peb3(Mix Normal)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−1.0
−2.0
1970
1975
1980
1985
1990
1995
2000
2005

# Summary for all 9 series (simple combining = 1)

| Method | Average Rel. MSFE (Fraction Rel. MSFE < 1) | | |
|---|---|---|---|
| | Full OOS Period | Split Out-of-Sample Period | |
| | | First Half | Second Half |
| AR | **1.10 (0.00)** | 1.12 (0.00) | 1.07 (0.03) |
| OLS | **2.16 (0.00)** | 2.44 (0.00) | 2.02 (0.00) |
| Combined-SSR | 1.05 (0.39) | 1.01 (0.50) | 1.14 (0.22) |
| FAAR-OLS | **0.96 (0.81)** | 0.96 (0.67) | 1.00 (0.69) |
| FAAR-GLS | **0.98 (0.61)** | 0.94 (0.67) | 1.14 (0.44) |
| FAAR-WLS | **0.96 (0.75)** | 0.95 (0.64) | 1.02 (0.67) |
| BMA(1/$n^2$,0.5) | **1.16 (0.31)** | 1.13 (0.33) | 1.31 (0.17) |
| BMA(1,0.5) | **1.23 (0.28)** | 1.17 (0.31) | 1.49 (0.17) |
| BMA-PC(1/$n^2$,0.5) | **1.07 (0.39)** | 1.01 (0.53) | 1.24 (0.22) |
| BMA-PC(1,0.5) | **1.08 (0.44)** | 1.07 (0.47) | 1.16 (0.31) |
| PEB-PC | **1.06 (0.42)** | 1.04 (0.42) | 1.15 (0.33) |
| BIC-PC | **1.34 (0.17)** | 1.33 (0.25) | 1.51 (0.06) |
| Bagging-PC | **1.54 (0.00)** | 1.61 (0.11) | 1.63 (0.03) |

# Empirical Bayes estimates of $p$, $g$

| | Forecast Horizon | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Series | 1 | | 3 | | 6 | | 12 | |
| | $\hat{p}$ | $\hat{g}$ | $\hat{p}$ | $\hat{g}$ | $\hat{p}$ | $\hat{g}$ | $\hat{p}$ | $\hat{g}$ |
| PI | 0.01 | 0.06 | 0.01 | 0.05 | 0.08 | 0.14 | 0.09 | 0.15 |
| IP | 0.19 | 0.15 | 0.13 | 0.09 | 0.10 | 0.05 | 0.07 | 0.04 |
| UR | 0.02 | 0.04 | 0.04 | 0.04 | **0.03** | **0.03** | 0.27 | 0.04 |
| EMP | 0.01 | 0.03 | 0.10 | 0.08 | 0.13 | 0.09 | 0.12 | 0.06 |
| TBILL | 0.07 | 0.11 | 0.05 | 0.07 | 0.08 | 0.07 | 0.07 | 0.08 |
| TBOND | 0.37 | 1.00 | 0.41 | 0.63 | 0.48 | 0.42 | 0.24 | 0.22 |
| PPI | 0.60 | 1.36 | 0.04 | 0.13 | 0.01 | 0.04 | 0.06 | 0.10 |
| CPI | 0.46 | 0.28 | 0.01 | 0.03 | 0.01 | 0.02 | 0.04 | 0.04 |
| PCED | 0.22 | 0.46 | 0.02 | 0.09 | 0.01 | 0.04 | 0.05 | 0.08 |

*small p: nonzero coefficients are rare*

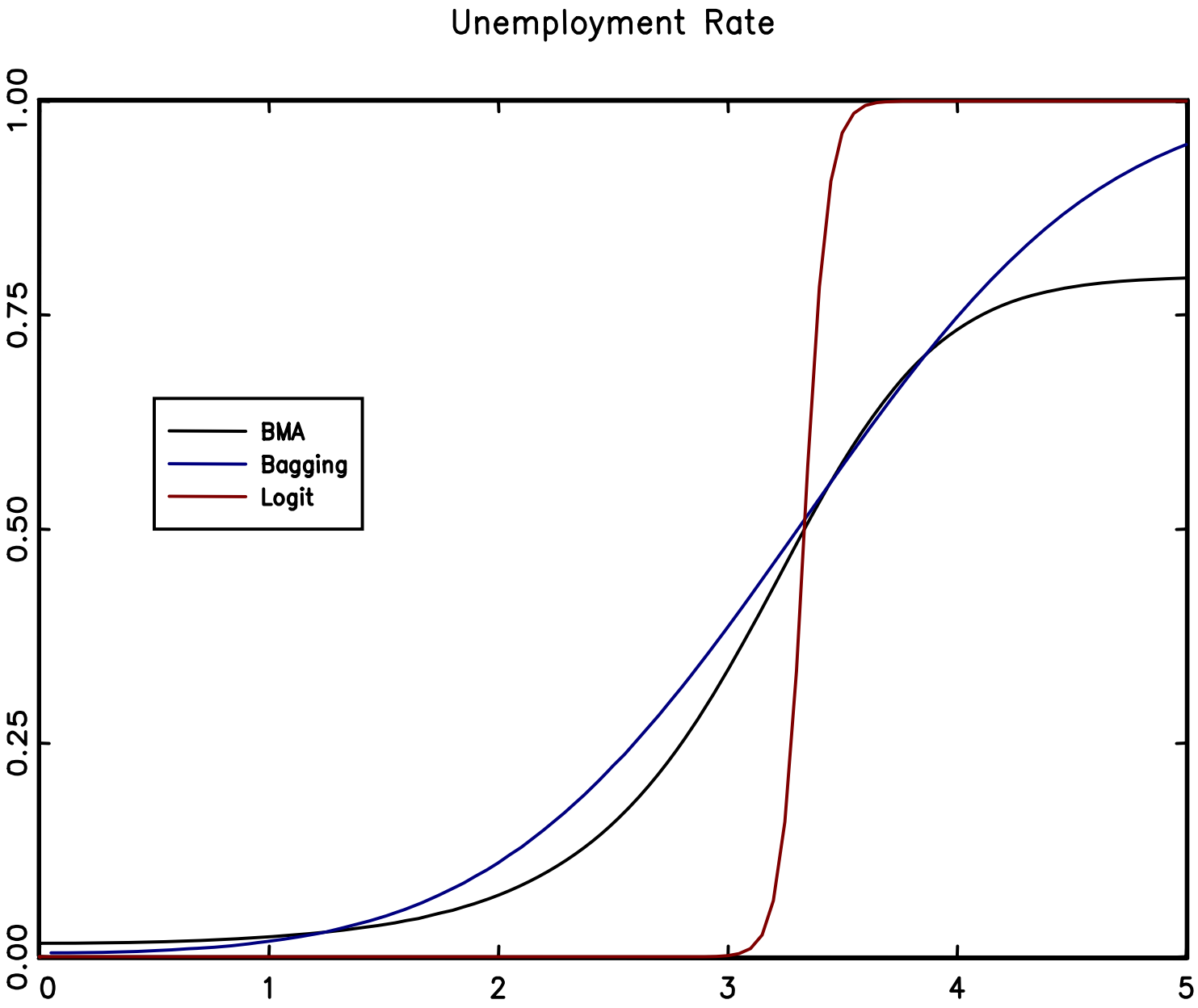
*small g: wide spread of prior for $\delta$, if it is nonzero*

# PLS estimation of BMA $p$, $g$; of bagging threshold $c$; and of logistic $\psi$ function $\beta_0$, $\beta_1$ – all 9 series

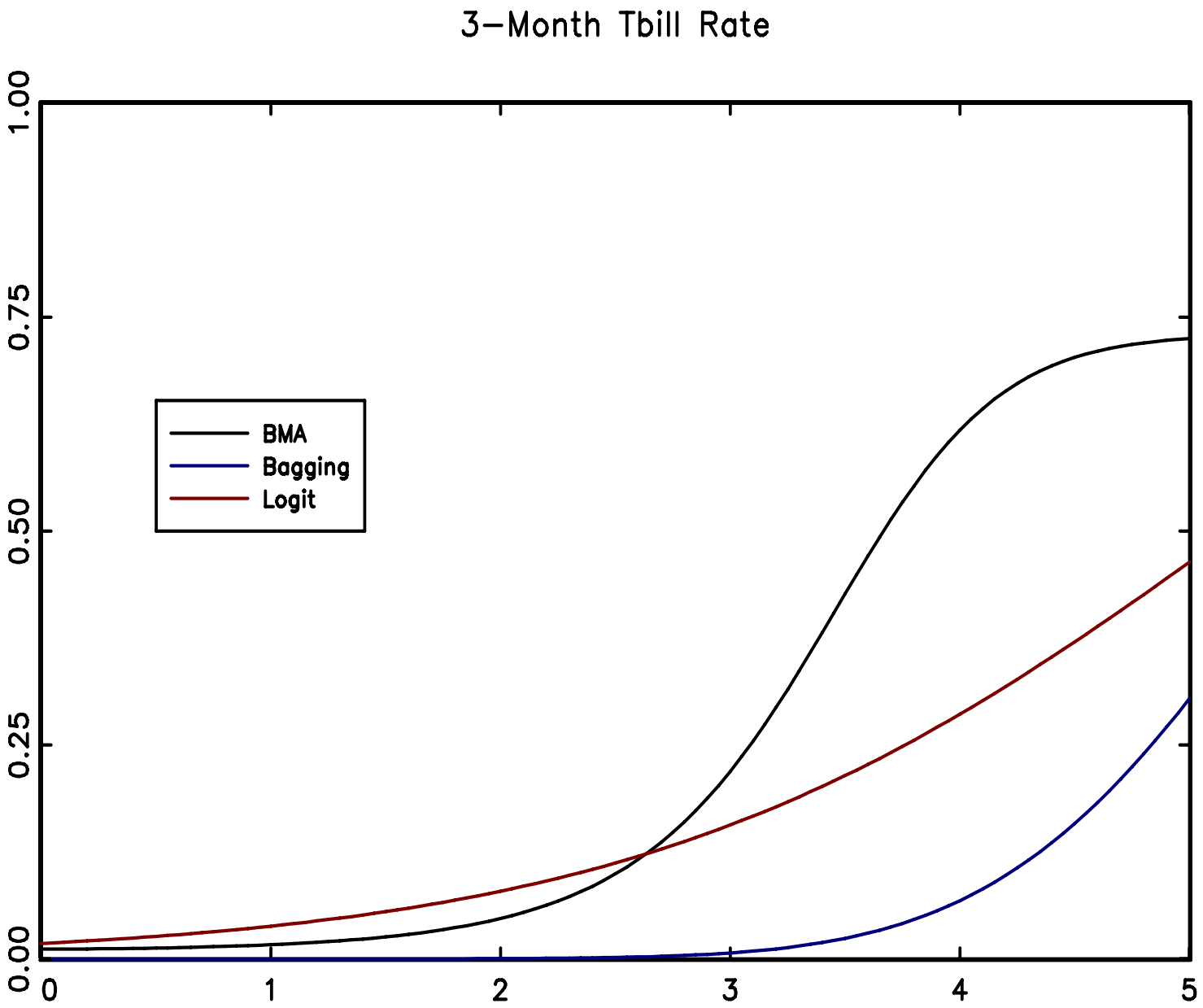
| | RMSFE | | | | Fraction of forecast variance coming from first 4 factors | | | |
|---|---|---|---|---|---|---|---|---|
| Series | First 4 | BMA | Bagging | Logistic | First 4 | BMA | Bagging | Logistic |
| PI | 1.042 | 0.919 | 0.927 | **0.905** | 1.00 | 0.78 | 0.87 | 0.90 |
| IP | **0.769** | 0.840 | 0.891 | 0.841 | 1.00 | 0.59 | 0.93 | 0.65 |
| Unemp | **0.710** | 0.784 | 0.805 | 0.790 | 1.00 | 0.80 | 0.84 | 0.82 |
| EMP | **0.880** | 0.910 | 0.977 | 0.914 | 1.00 | 0.46 | * | 0.31 |
| Tbill | 0.871 | **0.839** | 0.856 | 0.840 | 1.00 | 0.73 | 0.89 | 0.89 |
| Tbond | 1.018 | **0.978** | 0.996 | **0.978** | 1.00 | * | * | * |
| PPI | 1.053 | **0.999** | 1.000 | **0.999** | 1.00 | * | * | * |
| CPI | 0.969 | 0.947 | 0.978 | **0.942** | 1.00 | 0.17 | * | 0.55 |
| PCED | 1.174 | **0.985** | 0.995 | **0.985** | 1.00 | * | * | * |

- Estimated bagging is comparable to PEB-BMA
- Forecasts are heavily driven by first four factors

# Est'd $\psi$ functions (PLS) – unemployment rate



# Est'd $\psi$ functions (PLS) – 3-month T-bill



## Summary of Main Findings

1. The DFM seems to fit US data well, with a moderate number of factors (we estimate 7)

2. BMA and other methods can be used in time series applications, including multiperiod forecasts, in the context of "pseudo-BMA" – their behavior is the same whether or not the modeling assumptions (i.i.d. Gaussianity + strict exogeneity) hold

3. Empirical comparisons with other methods including empirical Bayes BMA indicate that DFM forecasts with a small number of factors are difficult to beat – there does not seem to be linearly exploitable information beyond the first few factors

# Correlation of forecasts: Averages across series and horizon

| | Method | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | Combined-Mean | 1.00 | . | . | . | . | . | . | . | . | . | . | . | . | . |
| **2** | AR | ***0.94*** | 1.00 | . | . | . | . | . | . | . | . | . | . | . | . |
| **3** | OLS | 0.43 | 0.36 | 1.00 | . | . | . | . | . | . | . | . | . | . | . |
| **4** | Combined-SSR | 0.75 | 0.65 | 0.47 | 1.00 | . | . | . | . | . | . | . | . | . | . |
| **5** | FAAR-OLS | 0.77 | 0.65 | 0.50 | 0.77 | 1.00 | . | . | . | . | . | . | . | . | . |
| **6** | FAAR-GLS | 0.73 | 0.61 | 0.53 | 0.73 | **0.86** | 1.00 | . | . | . | . | . | . | . | . |
| **7** | FAAR-WLS | 0.77 | 0.65 | 0.50 | 0.78 | ***0.98*** | **0.86** | 1.00 | . | . | . | . | . | . | . |
| **8** | BMA(1/$n^2$,0.5) | 0.65 | 0.56 | 0.59 | 0.79 | 0.77 | 0.73 | 0.77 | 1.00 | . | . | . | . | . | . |
| **9** | BMA(1,0.5) | 0.60 | 0.50 | **0.80** | 0.71 | 0.73 | 0.73 | 0.73 | **0.86** | 1.00 | . | . | . | . | . |
| **10** | BMA-PC(1/$n^2$,0.5) | 0.68 | 0.57 | 0.63 | 0.78 | **0.82** | 0.77 | **0.82** | **0.82** | **0.83** | 1.00 | . | . | . | . |
| **11** | BMA-PC(1,0.5) | 0.71 | 0.65 | **0.87** | 0.74 | 0.72 | 0.71 | 0.72 | 0.79 | **0.88** | **0.87** | 1.00 | . | . | . |
| **12** | PEB-PC | 0.67 | 0.57 | 0.66 | 0.77 | **0.80** | 0.75 | **0.80** | **0.80** | **0.82** | ***0.94*** | **0.87** | 1.00 | . | . |
| **13** | BIC-PC | 0.57 | 0.48 | 0.70 | 0.66 | 0.70 | 0.66 | 0.69 | 0.74 | **0.80** | **0.88** | **0.85** | **0.84** | 1.00 | . |
| **14** | Bagging-PC | 0.52 | 0.44 | ***0.96*** | 0.59 | 0.62 | 0.63 | 0.62 | 0.70 | **0.87** | 0.78 | ***0.94*** | 0.78 | **0.84** | 1.00 |

# More Results (see the paper)…

**a. $h$ = 1**

| | PI | IP | UR | EMP | TBILL | TBOND | PPI | CPI | PCED |
|---|---|---|---|---|---|---|---|---|---|
| Combined-Mean Root-MSFE | 6.51 | 7.54 | 0.17 | 2.16 | 0.55 | 0.34 | 5.48 | 2.52 | 1.93 |
| *MSFE Relative to Combined-Mean* | | | | | | | | | |
| AR | 1.04 | 1.09 | 1.07 | 1.07 | 1.03 | 1.02 | 1.04 | 1.06 | 1.03 |
| OLS | 1.87 | 1.94 | 1.83 | 2.41 | 1.73 | 1.43 | 2.70 | 2.33 | 2.04 |
| Combined-SSR | 0.88 | 0.91 | 0.90 | 0.93 | 1.03 | 0.93 | 1.06 | 1.07 | 1.02 |
| FAAR-OLS | 0.96 | 0.91 | 0.86 | 0.92 | 0.86 | 0.92 | 1.00 | 0.97 | 0.98 |
| FAA-GLS | 1.00 | 1.05 | 0.95 | 1.23 | 0.93 | 0.96 | 1.04 | 1.01 | 1.05 |
| FAAR-WLS | 0.96 | 0.90 | 0.86 | 0.95 | 0.86 | 0.93 | 1.01 | 0.95 | 0.98 |
| BMA(1/$n^2$,0.5) | 0.92 | 0.89 | 0.88 | 1.02 | 1.08 | 0.94 | 1.07 | 1.10 | 1.02 |
| BMA(1,0.5) | 0.94 | 0.83 | 0.87 | 1.10 | 1.05 | 0.99 | 1.21 | 1.21 | 1.22 |
| BMA-PC(1/$n^2$,0.5) | 0.95 | 0.90 | 0.87 | 0.91 | 0.89 | 0.90 | 1.08 | 1.13 | 1.02 |
| BMA-PC(1,0.5) | 0.99 | 0.92 | 0.97 | 0.96 | 0.99 | 0.95 | 1.14 | 1.15 | 1.09 |
| PEB-PC | 0.97 | 0.99 | 0.86 | 0.99 | 0.91 | 0.96 | 1.29 | 1.18 | 1.11 |
| BIC-PC | 1.05 | 0.96 | 0.97 | 1.02 | 1.07 | 1.02 | 1.28 | 1.26 | 1.22 |
| Bagging-PC | 1.22 | 1.15 | 1.16 | 1.44 | 1.28 | 1.14 | 1.60 | 1.54 | 1.42 |

**b. $h$ = 3**

| | PI | IP | UR | EMP | TBILL | TBOND | PPI | CPI | PCED |
|---|---|---|---|---|---|---|---|---|---|
| Combined-Mean Root-MSFE | 3.40 | 5.56 | 0.32 | 1.76 | 1.26 | 0.75 | 3.92 | 1.97 | 1.43 |
| *MSFE Relative to Combined-Mean* | | | | | | | | | |
| AR | 1.07 | 1.14 | 1.13 | 1.11 | 1.03 | 1.02 | 1.07 | 1.14 | 1.08 |
| OLS | 2.00 | 1.57 | 1.53 | 1.36 | 1.33 | 1.26 | 2.71 | 2.72 | 2.35 |
| Combined-SSR | 1.04 | 1.00 | 0.90 | 0.93 | 0.87 | 0.92 | 1.22 | 1.24 | 1.07 |
| FAAR-OLS | 0.98 | 0.85 | 0.85 | 0.91 | 0.91 | 0.96 | 0.98 | 0.91 | 0.95 |
| FAA-GLS | 0.99 | 0.96 | 0.84 | 1.05 | 0.88 | 0.94 | 1.01 | 0.92 | 0.98 |
| FAAR-WLS | 0.99 | 0.84 | 0.86 | 0.92 | 0.89 | 0.93 | 1.01 | 0.92 | 0.97 |
| BMA(1/$n^2$,0.5) | 0.96 | 0.96 | 0.88 | 0.98 | 0.88 | 0.95 | 1.30 | 1.31 | 1.09 |
| BMA(1,0.5) | 0.97 | 0.82 | 0.84 | 0.92 | 0.94 | 1.02 | 1.56 | 1.41 | 1.23 |
| BMA-PC(1/$n^2$,0.5) | 1.01 | 0.86 | 0.83 | 0.92 | 0.81 | 0.87 | 1.30 | 1.26 | 1.13 |
| BMA-PC(1,0.5) | 0.99 | 0.83 | 0.91 | 0.82 | 0.84 | 0.88 | 1.38 | 1.39 | 1.18 |
| PEB-PC | 0.96 | 0.83 | 0.82 | 0.89 | 0.82 | 0.91 | 1.42 | 1.30 | 1.19 |
| BIC-PC | 1.23 | 1.01 | 0.96 | 1.10 | 0.99 | 0.98 | 1.56 | 1.48 | 1.40 |
| Bagging-PC | 1.40 | 1.04 | 1.07 | 1.00 | 1.04 | 1.03 | 1.92 | 1.89 | 1.57 |

**c. $h$ = 6**

| | PI | IP | UR | EMP | TBILL | TBOND | PPI | CPI | PCED |
|---|---|---|---|---|---|---|---|---|---|
| Combined-Mean Root-MSFE | 2.39 | 4.15 | 0.50 | 1.64 | 1.66 | 1.06 | 3.08 | 1.71 | 1.19 |
| *MSFE Relative to Combined-Mean* | | | | | | | | | |
| AR | 1.08 | 1.18 | 1.20 | 1.06 | 1.06 | 1.02 | 1.07 | 1.17 | 1.10 |
| OLS | 2.58 | 2.64 | 1.75 | 2.07 | 1.35 | 1.47 | 3.07 | 2.32 | 2.69 |
| Combined-SSR | 1.16 | 1.12 | 1.01 | 0.90 | 0.90 | 0.98 | 1.24 | 1.24 | 1.14 |
| FAAR-OLS | 1.10 | 1.21 | 0.93 | 1.04 | 0.86 | 0.99 | 0.99 | 0.85 | 0.97 |
| FAA-GLS | 0.97 | 1.00 | 0.84 | 1.09 | 0.88 | 0.96 | 1.01 | 0.86 | 0.97 |
| FAAR-WLS | 1.13 | 1.20 | 0.92 | 1.04 | 0.80 | 0.99 | 0.99 | 0.84 | 0.97 |
| BMA(1/$n^2$,0.5) | 1.22 | 1.46 | 1.19 | 1.38 | 0.86 | 1.08 | 1.35 | 1.24 | 1.19 |
| BMA(1,0.5) | 1.19 | 1.39 | 1.01 | 1.35 | 0.87 | 1.13 | 1.69 | 1.41 | 1.43 |
| BMA-PC(1/$n^2$,0.5) | 1.12 | 1.09 | 0.97 | 1.12 | 0.79 | 0.99 | 1.41 | 1.20 | 1.17 |
| BMA-PC(1,0.5) | 1.07 | 0.99 | 0.95 | 1.03 | 0.86 | 0.98 | 1.47 | 1.27 | 1.26 |
| PEB-PC | 1.04 | 1.06 | 0.93 | 1.04 | 0.80 | 1.03 | 1.35 | 1.15 | 1.16 |
| BIC-PC | 1.46 | 1.30 | 1.15 | 1.57 | 0.98 | 1.15 | 1.93 | 1.38 | 1.68 |
| Bagging-PC | 1.71 | 1.71 | 1.23 | 1.64 | 1.07 | 1.25 | 2.19 | 1.71 | 1.94 |

**d. $h = 12$**

| | PI | IP | UR | EMP | TBILL | TBOND | PPI | CPI | PCED |
|---|---|---|---|---|---|---|---|---|---|
| Combined-Mean Root-MSFE | 1.86 | 3.40 | 0.84 | 1.66 | 2.12 | 1.57 | 2.73 | 1.59 | 1.12 |
| *MSFE Relative to Combined-Mean* | | | | | | | | | |
| AR | 1.05 | 1.25 | 1.21 | 1.15 | 1.12 | 1.02 | 1.14 | 1.28 | 1.19 |
| OLS | 2.28 | 2.83 | 2.07 | 2.30 | 2.15 | 1.75 | 3.16 | 2.62 | 3.44 |
| Combined-SSR | 1.14 | 1.16 | 1.05 | 1.07 | 1.11 | 0.97 | 1.28 | 1.17 | 1.11 |
| FAAR-OLS | 1.10 | 1.34 | 0.99 | 0.99 | 0.90 | 1.02 | 0.97 | 0.76 | 0.96 |
| FAA-GLS | 1.03 | 1.04 | 0.93 | 1.08 | 0.97 | 1.03 | 1.03 | 0.84 | 0.98 |
| FAAR-WLS | 1.14 | 1.38 | 0.92 | 1.01 | 0.93 | 1.01 | 0.95 | 0.74 | 0.99 |
| BMA(1/$n^2$,0.5) | 1.41 | 1.80 | 1.47 | 1.61 | 1.21 | 1.33 | 1.41 | 1.29 | 1.17 |
| BMA(1,0.5) | 1.41 | 1.64 | 1.27 | 1.55 | 1.50 | 1.33 | 1.86 | 1.51 | 1.67 |
| BMA-PC(1/$n^2$,0.5) | 1.19 | 1.22 | 1.05 | 1.21 | 1.22 | 1.15 | 1.47 | 1.23 | 1.40 |
| BMA-PC(1,0.5) | 1.04 | 1.12 | 1.04 | 1.15 | 1.25 | 1.05 | 1.50 | 1.28 | 1.36 |
| PEB-PC | 1.11 | 1.14 | 0.99 | 1.12 | 1.02 | 1.12 | 1.43 | 1.09 | 1.28 |
| BIC-PC | 1.57 | 1.88 | 1.45 | 1.58 | 1.61 | 1.52 | 1.93 | 1.60 | 2.02 |
| Bagging-PC | 1.76 | 2.14 | 1.54 | 1.83 | 1.85 | 1.57 | 2.28 | 1.80 | 2.42 |