NBER Summer Institute

What's New in Econometrics – Time Series

Lecture 12

July 16, 2008

Forecasting and Macro Modeling

with Many Predictors, Part II

Outline

4) Other High-Dimensional Forecasting Methods

Recall the introductory discussion of optimal forecasting with many orthogonal predictors, in which the frequentist problem was shown to be closely linked to the Bayes problem:

Frequentist: $\min_{\tilde{\delta}} r_{G_n}(\tilde{d}) = \kappa \int E(\tilde{d} - d)^2 dG_n(d)$ cdf of d_i

Bayes: $\min_{\tilde{\delta}} r_G(\tilde{d}) = \kappa \int E(\tilde{d} - d)^2 dG(d)$ subjective prior

Empirical Bayes: $\min_{\tilde{\delta}} r_{\hat{G}}(\tilde{d}) = \kappa \int E(\tilde{d} - d)^2 d\hat{G}(d)$ estimated "prior"

So far we have focused on a setup – the DFM – in which the DFM imposed structure on the coefficients in

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

The DFM said that, if P_t are the principal components, then only the first r of them matter – the rest of the coefficients are exactly zero.

High-dimensional methods, ctd.

The DFM implication that only the first r elements of δ are nonzero is an intriguing conjecture, but it might be false, or (more usefully) might not provide a good approximation.

The methods we will discuss now address the possibility that the remaining $n - r$ (= 135 – 4 = 131, say) principal components matter – or equivalently, all the X's enter separately with some small but useful weight.

This problem of prediction with many predictors has received a lot of attention in the stats literature so we will draw on it heavily:

- Empirical Bayes (parametric and nonparametric)
- Bayesian model averaging (BMA)
- Bagging, Lasso, etc
- Hard threhsholding methods including false discovery rate (FDR) (which is closely linked to Empirical Bayes, see Efron (2003))

High-dimensional methods, ctd.

We will focus on methods for orthogonal regressors (some generalize to non-orthogonal, some don't)

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

where $\mathbf{P}'\mathbf{P}/T = I_n$ (e.g. $\mathbf{P}$ = principal components)

Some (of many) methods:

1. ***Optimal Bayes*** estimator under the assumption $\delta_i = d_i/\sqrt{T}$, d_i i.i.d. G;
   The d_i i.i.d G model is the opposite extreme from a DFM (exchangeability: ordering i doesn't matter)
2. ***Hard thresholding*** (i.e. using a fixed t-statistic cutoff).
3. ***Information criteria*** AIC, BIC: here these reduced to hard thresholding with a cutoff c_T, where $c_T \to \infty$ (but not too quickly)

High-dimensional methods, ctd.

4. ***False discovery rate (FDR) methods***. In this problem, FDR turns into hard thresholding, except using the t-statistic is compared to a cutoff c_T that depends on the full distribution of t-statistics, $t_1,\ldots, t_n$. Used in genomics (10 million probes on a chip, pick out the sites that have unusual characteristics controlling the false positive rate, not the false negative rate as in testing).

5. ***Bootstrap aggregation ("bagging")*** (Breiman (1996), Bühlmann and Yu (2002); Inoue and Kilian (2008)).

6. ***Bayesian model averaging (BMA)***.

- ***References***
  - o Leamer (1978); Min and Zellner (1990); Fernandez, Ley, and Steele (2001a,b), Koop and Potter (2004)
  - o Surveys: Hoeting, Madiga, Raftery, and Volinsky (1999), Geweke and Whiteman (2004)
- ***Basic idea***: there are many possible models (submodels); assign them prior probability and compute posterior means.
- ***The BMA setup*** (notation: using X_t, not P_t – this doesn't need orthogonalized regressors in theory).

$Y_{t+1} \mid X_t$ is given by one of K models, denoted by $M_1,\ldots, M_K$.

Models are linear, so M_k lists variables in model k

$\pi(M_k)$ = prior probability of model k

D_t denotes the data set through date t

BMA, ctd.

The ***predictive density*** is the density of Y_{T+1} given the past data – the priors and the model are integrated out:

$$f(Y_{T+1}|D_T) = \sum_{k=1}^{K} f_k(Y_{T+1} | D_T)\Pr(M_k | D_T),$$

where $f_k(Y_{T+1}|D_T) = k^{\text{th}}$ predictive density

The ***posterior probability*** of model k is:

$$\Pr(M_k|D_T) = \frac{\Pr(D_T | M_k)\pi(M_k)}{\sum_{i=1}^{K}\Pr(D_T | M_i)\pi(M_i)},$$

where

$$\Pr(D_T|M_k) = \int \Pr(D_T | \theta_k, M_k)\pi(\theta_k | M_k)d\theta_k$$

θ_k = parameters in model k

$\pi(\theta_k|M_k)$ = prior for θ_k in model k

BMA, ctd.

Under quadratic loss, optimal forecast is the mean of the predictive density, which is the weighted average of the forecasts you would make under each model, weighted by the posterior probability of that model:

$$\tilde{Y}_{T+1|T} = \sum_{k=1}^{K} \Pr(M_k \mid D_T)\tilde{Y}_{M_k,T+1|T},$$

where $\tilde{Y}_{M_k,T+1|T}$ = posterior mean of Y_{T+1} for model M_k.

Comments

- Akin to forecast combining – where there are K forecasts
- How many models are there? How many distinct subsets of 135 variables can you make?
- fun for computational Bayesians (MCMC, etc)
- This simplifies with orthogonal regressors however…

BMA, ctd.

BMA with orthogonal regressors

Clyde, Desimone, and Parmigiani (1996), Clyde (1999):

- Variable j is in the model with probability π (coin flip)
- Given the model, the coefficients are distributed with a conjugate "g-prior" – and you get a closed form expression for posteriors

Comments:

1. Link to forecast combination – Bates and Granger (1969)
2. If the parameters of the prior (the "hyperparameters") are estimated, then this is parametric empirical Bayes.
3. All the theory and setup of BMA is for the cross-sectional case – the theoretical Bayes justification doesn't go through with predetermined regressors, nor for multistep forecasts. So its motivation is by analogy to to the i.i.d./exogenous regressor case.

Digression: shrinkage representations

All the estimators based on the regression

$Y_{t+1} = \delta' P_t + \varepsilon_{t+1},\ t = 1,\ldots, T, \quad P'P/T = I_n$ (e.g. principal components)

except FDR are shrinkage estimators (remember James-Stein) and produce forecasts of the form (at least, asymptotically as n, $T \to \infty$),

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

where ψ is a function that depends on the estimator (typically $0 \leq \psi(x) \leq 1$) and t_i is the t-statistic testing whether $\delta_i = 0$.

The shrinkage expression (1) also has a forecast combination interpretation: $\hat{\delta}_i P_{it}$ is the forecast made using the i^{th} predictor

Shrinkage representations, ctd

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

Here are some ψ functions:

Optimal Bayes estimator under the assumption $\delta_i = d_i/\sqrt{T}$, d_i i.i.d. ~ G;

$$\psi^B(u) = 1 + \frac{\ell(t)}{t}$$

where ℓ is the score of the marginal distribution of $\hat{\delta}_i$

Hard thresholding (i.e. using a fixed t-statistic cutoff).

$$\psi(t) = \mathbf{1}(|t_i| > c),\ c \text{ is some cutoff}$$

Information criteria AIC, BIC: here these reduced to hard thresholding with $\psi(t) = \mathbf{1}(|t_i| > c_T)$, where $c_T \to \infty$ (but not too quickly)

7. ***Large VARs***

De Mol, Giannone, and Reichlin (2006)

- Strong priors and estimated hyperparameter (EB implementation)
- Also consider Lasso (another high-dimensional method from statistics)

Outline

1) Why Might You Want To Use Hundreds of Series?
2) Dimensionality: From Curse to Blessing
3) Dynamic Factor Models: Specification and Estimation
4) Other High-Dimensional Forecasting Methods
5) **Empirical Performance of High-Dimensional Methods**
6) SVARs with Factors: FAVAR
7) Factors as Instruments
8) DSGEs and Factor Models

5) Empirical Performance of High-Dimensional Methods

(a) Data selection and preparation issues
(b) Comparisons among factor estimation methods
(c) Comparisons among many-predictor forecasting methods
(d) Empirical evidence on in-sample fit of DFM model
(e) Many-predictor methods vs. the world

Disclaimer: There now is a large literature and considerable practitioner experience with empirical DFMs, and a smaller but also substantial literature examining other many-predictor methods. This discussion is informed by this body of empirical knowledge but does not pretend to be a survey. See the survey and meta-analysis by Eichmeier and Ziegler (2006) for a bibliography.

(a) Data selection and preparation issues

Bear in mind that…

- The factors you get out depend on the data you put in.

- More variables do not always mean more information, for example putting in CND, CD, CS *and* total consumption doesn't make sense (aggregation identity).

- Judgment should be exercised about the balance between various categories of data; if most of the data are production and output, your dominant factor will be an output factor

(b) Comparisons among factor estimation methods

Discussed above. Empirical evidence suggests estimation method is not a first order issue although there is limited evidence on MLE (2-step or full) to date.

(c) Comparisons among many-predictor forecasting methods

Papers include Inoue and Kilian (2008), Koop and Potter (2004), Bańbura, Gianonne, and Reichlin (2008), Stock and Watson (2006a, 2006b).

- DFMs generally outperform the many-predictor statistical methods.
- Results from Stock and Watson (2006b) are consistent with this literature and make the point. SW consider forecasts in the shrinkage family,

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

  Forecasting methods: basic DFM (4 factors), bagging, BMA with fixed hyperparameters, Empirical Bayes (parametric, nonparametric), BIC hard thresholding; US monthly data. Figures are:

  (1) the ψ functions for the different procedures

  (2) the resulting $\psi(t)$ weights on the ordered principle components

Fig. 1. Shrinkage factors for PC forecasting model (unemployment)

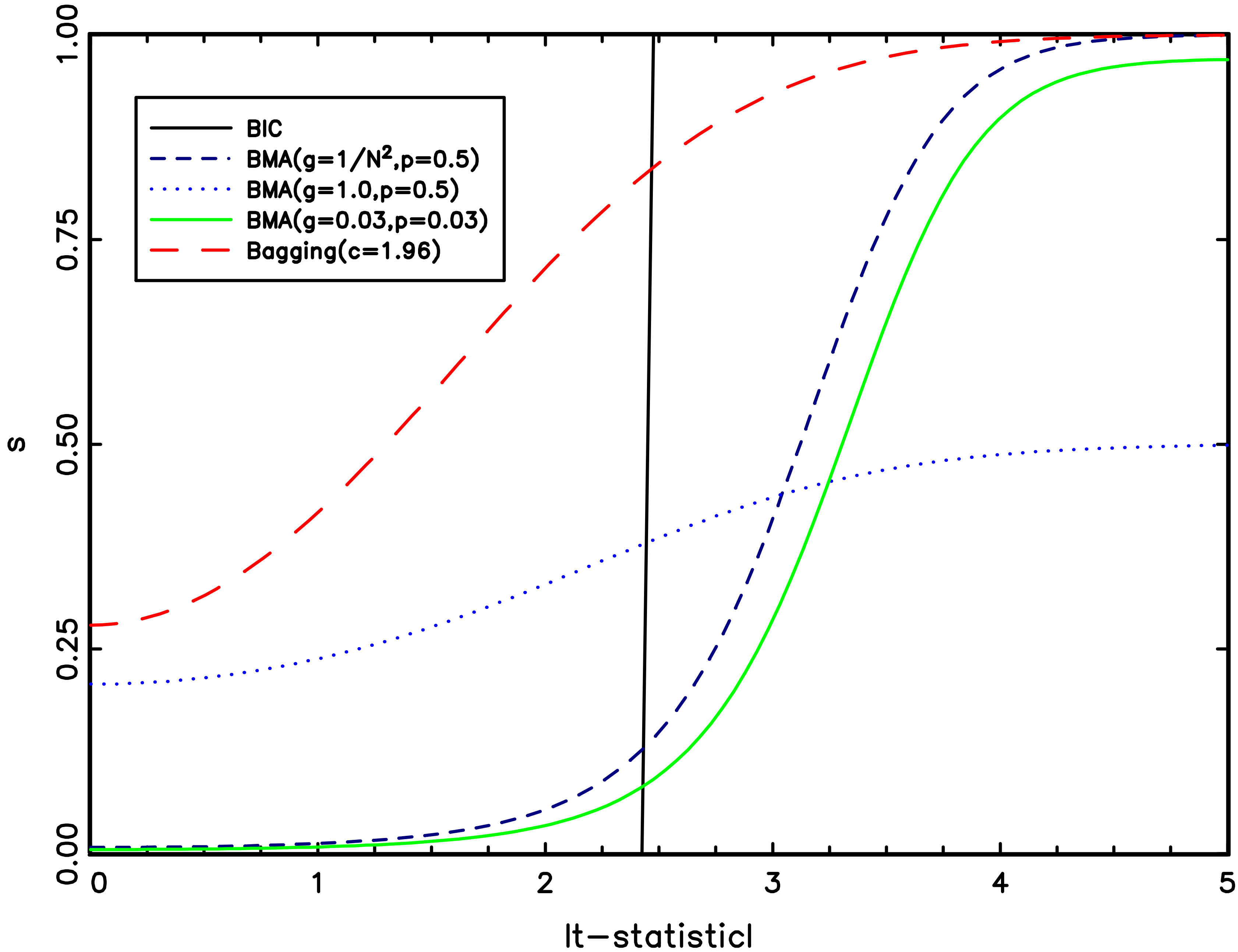
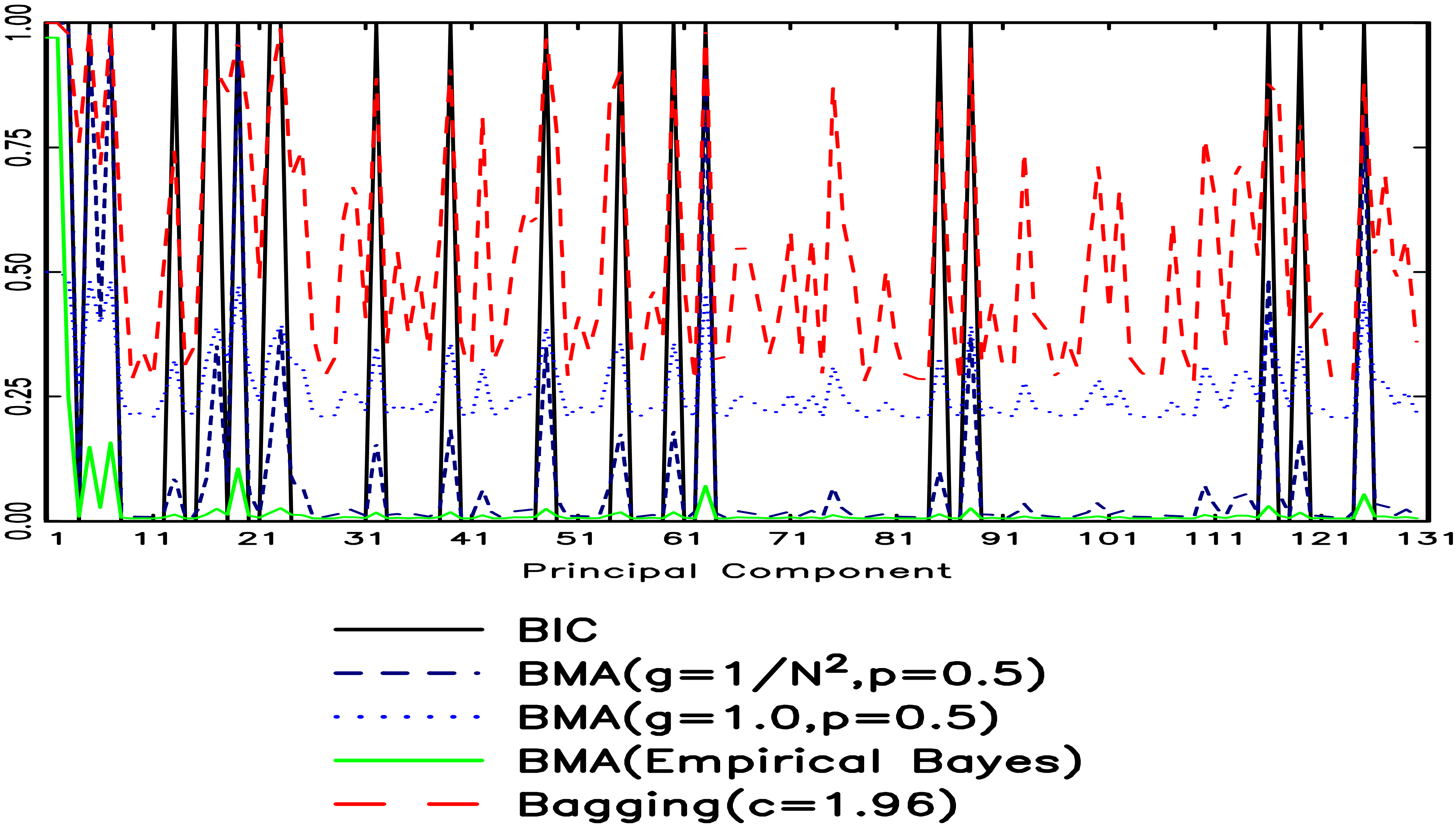
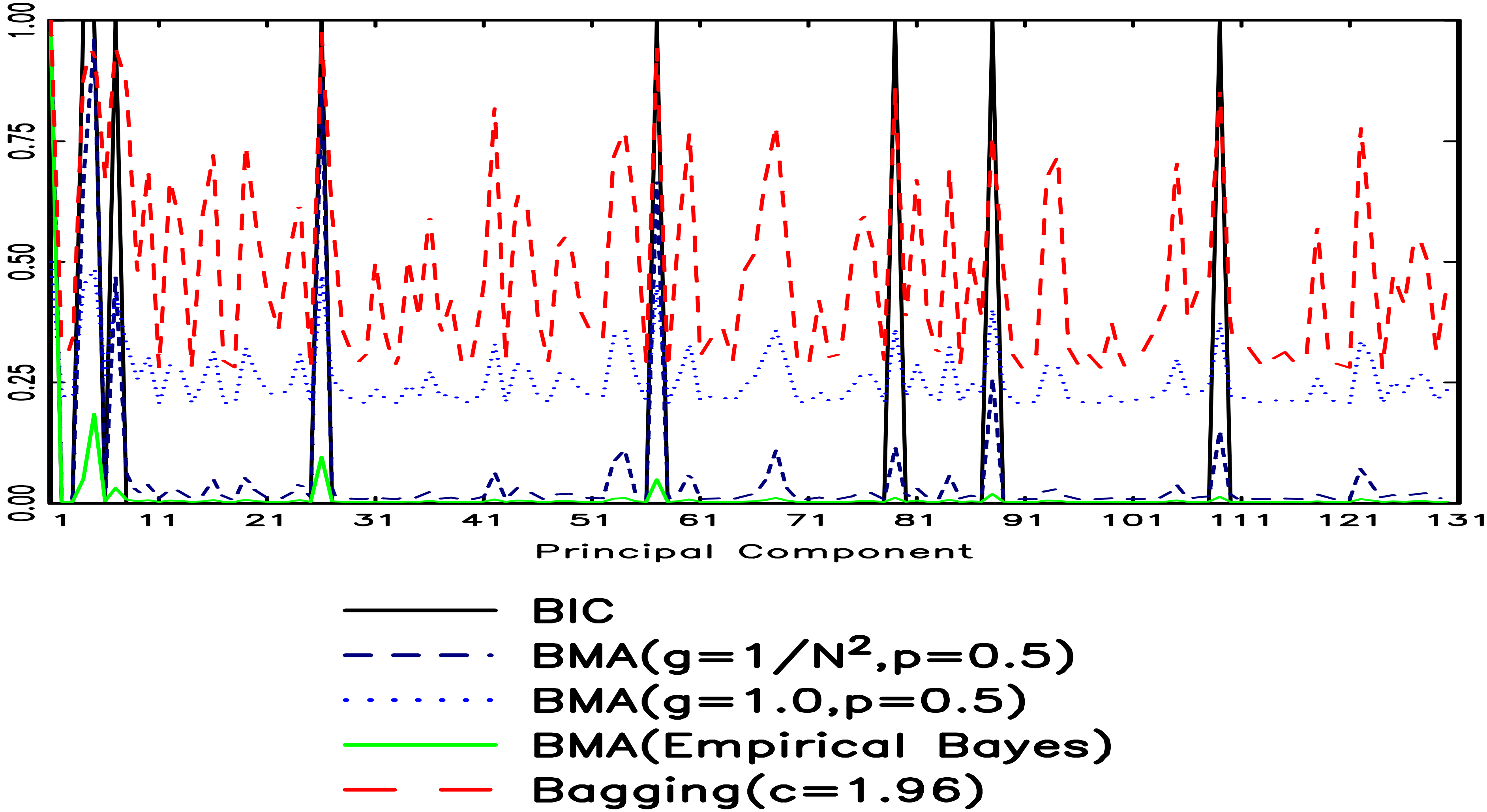


Fig. 2. Weights $\psi(t_i)$ on the ordered principle components

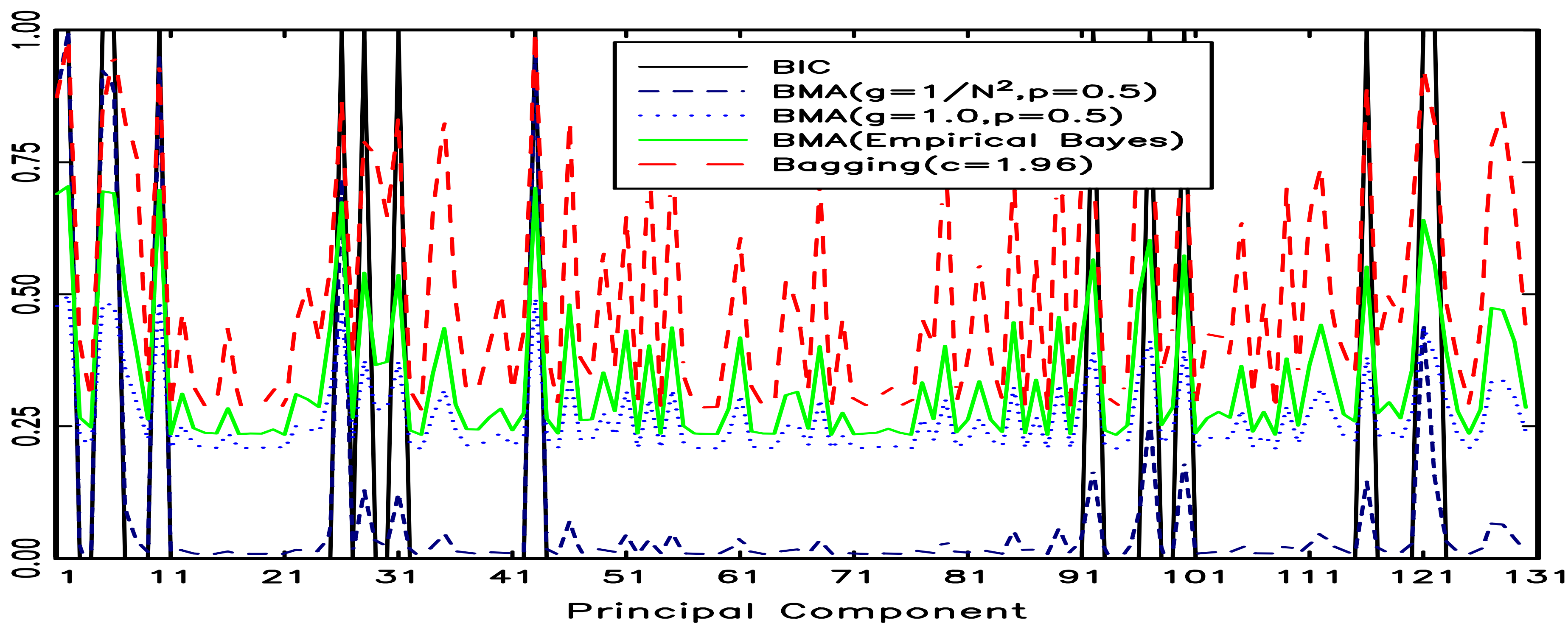
(a) Unemployment rate



(b) CPI inflation rate



(c) 10-Year T-bond Rate



BIC

BMA($g=1/N^2$, $p=0.5$)

BMA($g=1.0$, $p=0.5$)

BMA(Empirical Bayes)

Bagging($c=1.96$)

(d) Empirical evidence on in-sample fit of DFMs

Applications to US and EU data find that the first few PCs explain a large fraction of the data.

Watson (2004) comment on Giannone, Reichlin and Sala (2004)

Watson (2004) comment on Giannone, Reichlin and Sala (2004)

Table 1
Fraction of variance explained by one- and two-factor models

	Sargent and Sims[1]		Giannone, Reichlin, and Sala[2]	
Series	1 factor	2 factors	1 factor	2 factors
Average weekly hours	0.77	0.80	0.49	0.61
Layoffs	0.83	0.85	0.72	0.82
Employment	0.86	0.88	0.85	0.91
Unemployment	0.77	0.85	0.74	0.82
Industrial production	0.94	0.94	0.88	0.93
Retail sales	0.46	0.69	0.33	0.47
New orders durables	0.67	0.86	0.65	0.74
Sensitive material prices	0.19	0.74	0.53	0.60
Wholesale prices	0.20	0.69	0.34	0.67
M1	0.16	0.20	0.15	0.30

1. From Table 21 of Sargent and Sims (1977).
2. From Appendix 6.2.

Empirical evidence on in-sample fit of DFMs, ctd

Stock and Watson (2005)

- Test exact DFM restrictions, find large fraction of rejections in U.S. quarterly data
- But the rejections are all very small in a R^2 sense.
- The approximate DFM seems to be a good description of the data

(e) Many-predictor methods vs. the world

Generally speaking, it depends on the application

Eichmeier and Ziegler (2006) (limitations)

Inflation

U.S. survey by Stock and Watson (2008)

Output

US, EU – generally find substantial improvements (especially US) over other models

Outline

1) Why Might You Want To Use Hundreds of Series?
2) Dimensionality: From Curse to Blessing
3) Dynamic Factor Models: Specification and Estimation
4) Other High-Dimensional Forecasting Methods
5) Empirical Performance of High-Dimensional Methods
6) **SVARs with Factors: FAVAR**
7) Factors as Instruments
8) DSGEs and Factor Models

6) SVARs with Factors: FAVAR

Challenges & critiques of standard SVAR modeling include:

- The Rudebush (1998) critique of SVARs with short-run timing identification: Fed uses more information than is in a standard VAR
- The invertibility problem in SVARs: is $Ru_t = \varepsilon_t$, $\varepsilon_t = R^{-1}u_t$ plausible?
- Including more variables in the VAR might improve forecast efficiency and provide an internally consistent set of forecasts for a large number of variables – but confronts the n^2p parameter problem

Bernanke, Boivin, and Eliasz's (2005) (BBE) idea is to use factors as a way to solve this problem: in a DFM, factors summarize all the relevant information on the economy. The result is the BBE Factor Augmented VAR (FAVAR).

FAVAR, ctd

There are a number of ways FAVAR can be implemented, the following papers use related approaches but differ in the details:

Bernanke, B.S., and J. Boivin (2003), Bernanke, Boivin, and Eliasz (2005) (BBE), Favero and Marcellino (2001), Favero, Marcellino, and Neglia (2004); also see Giannone, Reichlin, and Sala (2004) on the invertibility issue.

Here we follow the spirit of BBE (2005) although some technical details (but not identification ideas) are different – this development follows Stock and Watson (2005).

One approach would be simply to put factors into a SVAR, however the factors themselves are not identified so making any identification assumptions about their innovations is difficult.

FAVAR, ctd.

VAR form of the exact DFM

DFM with first order dynamics from above:

$$F_t = \Phi F_{t-1} + \mathrm{G}\eta_t$$

$$X_t = \Lambda F_t + e_t$$

$$e_t = De_{t-1} + \zeta_t$$

where D is diagonal. Quasi-difference X_t:

$$(I - D\mathrm{L})X_\mathrm{t} = (I - D\mathrm{L})\Lambda F_t + \zeta_t = \Lambda F_t - D\Lambda F_{t-1} + \zeta_t$$

Substitute in $F_t = \Phi F_{t-1} + \mathrm{G}\eta_t$:

$$(I - D\mathrm{L})X_\mathrm{t} = \Lambda(\Phi F_{t-1} + \mathrm{G}\eta_t) - D\Lambda F_{t-1} + \zeta_t$$

Rearrange:

$$X_\mathrm{t} = (\Lambda\Phi - D\Lambda)F_{t-1} + DX_{t-1} + \Lambda\mathrm{G}\eta_t + \zeta_t$$

Putting the F_t and X_t equations together yields,

VAR form of the DFM, ctd.

$$\begin{pmatrix} F_t \\ X_t \end{pmatrix} = \begin{pmatrix} \Phi & 0 \\ \Lambda\Phi - D\Lambda & D \end{pmatrix} \begin{pmatrix} F_{t-1} \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} G & 0 \\ \Lambda G & I \end{pmatrix} \begin{pmatrix} \eta_t \\ \zeta_t \end{pmatrix}$$

Writing the reduced form VAR as $A(L)X_t = u_t$, the VAR innovations are $u_t = X_t - \text{Proj}(u_t|F_{t-1}, F_{t-2},\ldots, X_{t-1}, X_{t-2},\ldots) = \Lambda G\eta_t + \zeta_t$, where we are treating the F's as observed (this is justified by large n asymptotics).

The ζ's are disturbances to the idiosyncratic process. What we are interested in is the response of X_t to structural shocks, which affect all the variables. The structural shocks ε_t are related to the innovations in the dynamic factors:

$$R\eta_t = \varepsilon_t$$

FAVAR

reduced form:
$$\begin{pmatrix} F_t \\ X_t \end{pmatrix} = \begin{pmatrix} \Phi & 0 \\ \Lambda\Phi - D\Lambda & D \end{pmatrix} \begin{pmatrix} F_{t-1} \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} G & 0 \\ \Lambda G & I \end{pmatrix} \begin{pmatrix} \eta_t \\ \zeta_t \end{pmatrix}$$

structure:
$$\underset{q\times q}{R}\ \underset{q\times q}{\eta_t} = \underset{q\times q}{\varepsilon_t}$$

The structural IRF is the distributed lag of X_t on ε_t. Now

$$X_t = \Lambda F_t + e_t$$

and
$$F_t = \Phi F_{t-1} + G\eta_t = \Phi F_{t-1} + GR^{-1}\varepsilon_t,$$

so
$$X_t = \Lambda(I - \Phi L)^{-1}GR^{-1}\varepsilon_t + e_t$$

so the structural IRF is $\Lambda(I - \Phi L)^{-1}GR^{-1}$.

FAVAR, ctd.

Comments:

1. ***Lags***. These formulas are for first order dynamics – with higher order dynamics the expression above becomes,

$$\begin{pmatrix} F_t \\ X_t \end{pmatrix} = \begin{pmatrix} \Phi(L) & 0 \\ \Lambda\Phi(L) - D(L)\Lambda & D(L) \end{pmatrix} \begin{pmatrix} F_{t-1} \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} G & 0 \\ \Lambda G & I \end{pmatrix} \begin{pmatrix} \eta_t \\ \zeta_t \end{pmatrix}$$

structure: $$\underset{q\times q}{R}\ \underset{q\times q}{\eta_t} = \underset{q\times q}{\varepsilon_t}$$

2. ***Identification***. The identification problem is finding R, where $R\eta_t = \varepsilon_t$. This is now amenable to applying the SVAR identification toolkit:
   - Timing scheme (BBE: slow/policy/fast, see Lecture #7)
   - long run restrictions
   - sign restrictions (see Ahmadi and Uhlig (2007))
   - heteroskedasticity

3. ***Structural shocks.*** The η_t shocks are the shocks to the *dynamic* factors: $F_t = \Phi F_{t-1} + \mathrm{G}\eta_t$. These are *not* the residuals from a VAR estimated using F_t: the number of static factor innovations $r \geq q$. Implementation involves estimating the space of dynamic factor shocks, which in turn entails (i) estimating the number of dynamic factors q, and (ii) reduced rank regressions to estimate η_t.

4. ***Many impulse responses.*** The structural IRF is $\Lambda(I - \Phi L)^{-1}GR^{-1}$, which yields IRFs for all the X's in the system!

5.***Overidentification.*** These systems move from being exactly identified SVARs to potentially heavily overidentified. Consider the BBE fast/slow identification idea: the slow identification restriction now applies to a huge block of variables, specifically, ε_t^r should not load on any of the slow moving variables. Let u_t^S be the VAR innovations to the slow-moving variables, $u_t^S = X_t^S - \text{Proj}(X_t^S | F_{t-1}, F_{t-2}, \ldots, X_{t-1}, X_{t-2}, \ldots)$. Under the fast/slow identification scheme, $\text{Proj}(u_t^S | \varepsilon_t^r)$ should be zero. These many overidentifying restrictions are testable.

Outline

1) Why Might You Want To Use Hundreds of Series?
2) Dimensionality: From Curse to Blessing
3) Dynamic Factor Models: Specification and Estimation
4) Other High-Dimensional Forecasting Methods
5) Empirical Performance of High-Dimensional Methods
6) SVARs with Factors: FAVAR
7) **Factors as Instruments**
8) DSGEs and Factor Models

7) Factors as Instruments

Independently developed by Kapetanios and Marcellino (Oct. 2006, revised 2008) and Bai and Ng (Oct. 2006, revised 2007b)

Remember the weak instrument problem…

- Using factors might be a way to use more information, without the pitfalls of the many instrument problem!
- The instruments $\hat{F}_t$ are linear combinations of the X_t's, but the key insight is that the coefficients of that linear combination are estimated separately, not in the first-stage regression (the X's don't enter the moment conditions explicitly).
- The mathematics is essentially the same as the math used to show that $\hat{F}_t$ can be used in a forecasting regression without a generated regressor problem.

Factors as instruments, ctd.

Main result: under conditions like those above (the approximate DFM conditions), and the "usual" large-*n* rate condition $N^2/T \to \infty$, and a strong instrument assumption,

$$\sqrt{T}\,(\hat{\beta}^{TSLS}(F_t) - \hat{\beta}^{TSLS}(\hat{F}_t)) \xrightarrow{p} 0 \qquad (2)$$

where $\hat{F}_t$ is the PC estimator of the factors. So IV is as efficient if the factors are known as if they are not when N is large.

Simulation results in Kapetanios and Marcellino (2008) and Bai and Ng (2007b) are promising concerning the finite-sample validity of (2) under strong instruments.

<u>Factors as instruments, ctd.</u>

Additional comments

1. The idea of using principal components as instruments is old (Kloek and Mennes (1960), Amemiya (1966)) – what is new is proving optimality results using the DFM as the conceptual framework.
2. Not all the individual *X*'s need to be valid instruments – the *e*'s could be correlated with the included endogenous regressor, what matters is that the *F*'s are not correlated.
3. If there isn't a factor structure, then the PC estimates are going to random linear combinations of the *X*'s. But if the *X*'s are all valid instruments, the $\hat{F}_t$'s remain valid instruments even without a factor structure (details in Bai and Ng (2007)).
4. If the instruments (*F*'s) are weak, then weak instrument asymptotics kicks in. (The original hope is that weak instruments will be less of a problem using the *F*'s.)

Outline

1) Why Might You Want To Use Hundreds of Series?
2) Dimensionality: From Curse to Blessing
3) Dynamic Factor Models: Specification and Estimation
4) Other High-Dimensional Forecasting Methods
5) Empirical Performance of High-Dimensional Methods
6) SVARs with Factors: FAVAR
7) Factors as Instruments
8) **DSGEs and Factor Models**

8) DSGEs and Factor Models

"Reduced form" DFM with first order dynamics from above:

$$F_t = \Phi F_{t-1} + G\eta_t$$

$$X_t = \Lambda F_t + e_t$$

$$e_t = De_{t-1} + \zeta_t$$

Boivin and Giannoni (2006b) replace the reduced form state space model with a linearized DSGE:

$$\tilde{F}_t = \tilde{\Phi} \tilde{F}_{t-1} + \tilde{G} \tilde{\eta}_t \quad (3)$$

$$X_t = \tilde{\Lambda} \tilde{F}_t + e_t, \quad (4)$$

$$e_t = De_{t-1} + \zeta_t \quad (5)$$

where ~ means that (3) is a structural model (DSGE), cf. Sargent (1989), Boivin-Giannoni (2006b).

DSGEs and factor models, ctd.

$$\tilde{F}_t = \tilde{\Phi}\tilde{F}_{t-1} + \tilde{G}\tilde{\eta}_t$$

$$X_t = \tilde{\Lambda}\tilde{F}_t + e_t,$$

$$e_t = De_{t-1} + \zeta_t$$

The DSGE implies restrictions on $\tilde{\Lambda}$ that identify $\tilde{F}_t$:

- The elements of $\tilde{F}_t$ correspond to "output gap" (x_t), "inflation" (π_t), "the interest rate" (r_t), "hours worked", etc. In the example DSGE in lecture 8, $\tilde{F}_t = (x_t, \pi_t, r_t)'$.
- The meanings of the elements $\tilde{F}_t$ within the DSGE imply restrictions on $\tilde{\Lambda}$ that identify $\tilde{F}_t$
- The system, with restrictions on $\tilde{\Lambda}$ imposed, is in SS form and the KF can be used to compute the likelihood. Estimation is a combination of DFM MLE and DSGE MLE with a small number of variables:
  - initial values using PC estimates of the factors
  - modified Jungbacker-Koopman (2008) speedup?

Boivin-Giannoni (2006b) identification: Setup: let $\tilde{\lambda}$ denote a nonzero entry (not all the same – just dropping subscripts)

$$\begin{bmatrix} \text{output gap series \#1} \\ \vdots \\ \text{output gap series \#n}_{\text{Y}} \\ \text{inflation series \#1} \\ \vdots \\ \text{inflation series \#n}_{\text{C}} \\ \vdots \\ ---- \\ \text{Information series \#1} \\ \vdots \\ \text{Information series \#n}_{\text{info}} \end{bmatrix} = \begin{bmatrix} \tilde{\lambda} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{\lambda} & 0 & \cdots & 0 \\ 0 & \tilde{\lambda} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \tilde{\lambda} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ & & & \\ \tilde{\lambda} & \tilde{\lambda} & \cdots & \tilde{\lambda} \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{\lambda} & \tilde{\lambda} & \cdots & \tilde{\lambda} \end{bmatrix} \begin{bmatrix} x_t \\ \pi_t \\ \vdots \\ \tilde{F}_{last,t} \end{bmatrix}, \text{ where } \tilde{F}_t = \begin{bmatrix} x_t \\ \pi_t \\ \vdots \\ \tilde{F}_{last,t} \end{bmatrix}$$

Or

$$\begin{bmatrix} X_{\text{sensor},t} \\ X_{\text{info},t} \end{bmatrix} = \begin{bmatrix} \tilde{\Lambda}_{\text{sensor}} \\ \tilde{\Lambda}_{\text{info}} \end{bmatrix} \tilde{F}_t, \text{ where } \tilde{F}_t = \tilde{\Phi}(\text{L})\tilde{F}_t + \tilde{\varepsilon}_{t}t$$

In general the information series can have weights on expectations of future F_t (e.g. term spreads) but by the VAR structure of the factors plus the DFM assumptions those are projected back on F_t.

Results from Boivin-Giannone (they use Bayes methods)

Case A: 7 variables

Case B: 14 variables

Case C: 91 variables

Table 1: Priors and estimates of structural parameters

	Prior Distribution			SW	Case A	Case B	Case C
	Type	Mean	St.Err.				
φ	Normal	4	1.5	5.36	5.88	6.17	3.81
				(0.88)	(1.11)	(1.13)	(1.04)
σ_c	Normal	1	0.375	1.54	1.45	1.79	1.63
				(0.24)	(0.23)	(0.44)	(0.44)
h	Beta	0.7	0.1	0.71	0.75	0.54	0.50
				(0.07)	(0.07)	(0.27)	(0.27)
σ_L	Normal	2	0.75	2.34	2.18	2.42	2.41
				(0.60)	(0.65)	(0.69)	(0.68)
ϕ	Normal	1.25	0.125	1.42	1.24	1.37	1.26
				(0.08)	(0.07)	(0.07)	(0.07)
$1/\psi$	Normal	0.2	0.075	0.32	0.27	0.26	0.27
				(0.06)	(0.06)	(0.06)	(0.06)
ξ_ω	Beta	0.75	0.05	0.81	0.77	0.78	0.82
				(0.02)	(0.03)	(0.04)	(0.03)
ξ_p	Beta	0.75	0.05	0.88	0.90	0.88	0.86
				(0.01)	(0.02)	(0.01)	(0.02)
γ_ω	Beta	0.5	0.15	0.39	0.45	0.43	0.48
				(0.12)	(0.14)	(0.14)	(0.14)
γ_p	Beta	0.5	0.15	0.66	0.72	0.50	0.36
				[illegible]	[illegible]	[illegible]	[illegible]

				(0.12)	(0.14)	(0.14)	(0.14)
γ_p	Beta	0.5	0.15	0.66	0.72	0.50	0.36
				(0.08)	(0.19)	(0.15)	(0.14)
ρ	Beta	0.75	0.1	0.76	0.67	0.72	0.70
				(0.02)	(0.05)	(0.04)	(0.03)
$r_{\pi 0}$	Normal	1.8	0.1	1.78	1.81	1.72	1.66
				(0.08)	(0.10)	(0.10)	(0.09)
$r_{\pi 1}$	Normal	-0.3	0.1	-0.22	-0.22	-0.30	-0.39
				(0.09)	(0.12)	(0.10)	(0.09)
r_{y0}	Normal	0.188	0.05	0.22	0.23	0.24	0.22
				(0.03)	(0.03)	(0.03)	(0.03)
r_{y1}	Normal	-0.063	0.05	-0.13	-0.11	-0.12	-0.12
				(0.03)	(0.03)	(0.04)	(0.03)
Implied parameters							
pseudo EIS: $\frac{1-h}{(1+h)\sigma_c}$				0.110	0.099	0.167	0.204
slope of PC: $\frac{(1-\beta\xi_p)(1-\xi_p)}{(1+\beta\gamma_p)\xi_p}$				0.011	0.007	0.012	0.018

The parameter estimates are given by the median of the posterior distribution

Results are based on 100 000 replications. Standard errors are reported in ().

Misc. concluding DFM comments

1. Everything in this lecture has applied to variables with short-run dependence. There is a fair amount of work extending DFMs to handle unit roots and cointegration, for one of several papers in the literature see Bai and Ng (2004) (and see their references).

2. We also have ignored TVP and structural breaks in DFMs. DFMs have a certain robustness to TVP and structural breaks, however the only published work with any TVP aspect in DFMs is Stock and Watson (2002) and Phillips and Sul (1997). Recent unpublished work includes Stock and Watson (2007) and Banergjee, Marcellino, and Masten (2007).

Summary

1. The quest for exploiting large data sets has made considerable advances
2. Large *n is* a blessing – turning the principle of parsimony on its head ($N^2/T \to \infty$ results)
3. State of knowledge of DFM estimation and factor extraction is pretty advanced: it doesn't seem to make a lot of difference what method you use if *n* is large, but this said the MLE (two-step seems to be enough) has some nice properties theoretically and in initial applications.
4. Applications to forecasting are well advanced and implemented in real time. Applications to SVARs (FAVAR), IV estimation, and DSGE estimation are promising.