Likelihood-based analysis for dynamic factor models

Borus Jungbacker and Siem Jan Koopman

Department of Econometrics, Vrije Universiteit Amsterdam

Tinbergen Institute, Amsterdam

`http://staff.feweb.vu.nl/koopman`

Sveriges Riksbank – Conference on “... state space models”

Stockholm – October 17-18, 2008

Dynamic factor models

- Dynamic factor models have the typical form

$$y_{it} = \mu_i + \sum_{j=0}^{m} \lambda'_{ij} f_{t-j} + u_{it}, \qquad u_{it} = \phi_i u_{it} + \varepsilon_{it},$$

  where f_t is a lower-dimensional (unobserved) dynamic process that may also enter y_{it} in lags for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.
- Key contributions are given by Joreskog (1969), Sargent and Sims (1977), Geweke (1977), Reinsel (1983), Connor and Korajczyk (1986,1988,1993), Harvey, Fernandez-Macho and Stock (1987), Forni, Hallin, Lippi and Reichlin (2000, 2002), Stock and Watson (2002,2005), Bai and Ng (2002,2004,2007), Marcellino, Stock and Watson (2003), Breitung (2005), Doz, Giannone and Reichlin (2006) and many interesting empirical contributions recently.
- Estimation treatments are based on frequency domain methods, principal components, static factor analysis, etc: “exact” treatments are usually dismissed on computational grounds...

Dynamic factor model with regression effects

We consider the slightly more general form

$$y_{it} = \mu_i + x_{it}\beta + \sum_{j=0}^{m} \lambda'_{ij} f_{t-j} + u_{it}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T,$$

where

- y_{it} denotes the observed value for the i^{th} time series at time t;
- μ_i is a fixed and unknown constant;
- x_{it} is a $1 \times K$ vector of covariates;
- β is a $K \times 1$ vector of regression coefficients;
- f_t is an $r \times 1$ vector of stationary common factors;
- λ_{ij} is an $r \times 1$ vector of fixed and unknown loadings for f_{t-j};
- u_{it} is the stationary idiosyncratic autoregressive component.

Panel time series models

The inclusion of constants (fixed effects) and covariates (exogenous) variables into the dynamic factor model allows the extension of the results towards panel time series models that may also include unobserved dynamic effects.

Vector representation

The DFM model in vector form is

$$y_t = \bar{\mu} + \bar{X}_t\beta + \Lambda(L)f_t + u_t,$$

$$\Phi(L)f_t = \Theta(L)\zeta_t, \qquad \Psi(L)u_t = \varepsilon_t, \qquad t = 1, \ldots, T,$$

where ...

Vector representation

The DFM model in vector form is

$$y_t = \bar{\mu} + \bar{X}_t \beta + \Lambda(L) f_t + u_t,$$

$$\Phi(L) f_t = \Theta(L)\zeta_t, \qquad \Psi(L) u_t = \varepsilon_t, \qquad t = 1, \dots, T,$$

where $y_t = (y_{1t}, \dots, y_{Nt})'$, $u_t = (u_{1t}, \dots, u_{Nt})'$,
$\bar{\mu} = (\mu_1, \dots, \mu_N)'$, $\bar{X}_t = (x'_{1t}, \dots, x'_{Nt})'$;

- $\Lambda(L) = \Lambda_0 + \sum_{j=1}^{m} \Lambda_j L^j$ with $\Lambda_j = (\lambda_{1j}, \dots, \lambda_{Nj})'$ for $j = 0, \dots, m$;
- $\Phi(L) f_t = \Theta(L)\zeta_t$, with innovations ζ_t, Var$(\zeta_t) = \Sigma_\zeta$, and lag polynomials $\Phi(L) = I - \sum_{j=1}^{q_\Phi} \Phi_j L^j$ and $\Theta(L) = I + \sum_{j=1}^{q_\Theta} \Theta_j L^j$;
- $\Psi(L) u_t = \varepsilon_t$, with innovations ε_t, Var$(\varepsilon_t) = \Sigma_\varepsilon$, and lag polynomial is $\Psi(L) = I - \sum_{j=1}^{q_\Psi} \Psi_j L^j$.

Parameters are in ψ. Parameters excluding $\bar{\mu}$ and β are in θ, that is

$$\psi = (\bar{\mu}' \ , \ \beta' \ , \ \theta')'.$$

DFM model assumptions

1. For all $|z| \leq 1$, we have $|\Phi(z)| \neq 0$ and $|\Psi(z)| \neq 0$.
2. By taking $\mathcal{F}_t$ as the σ-algebra generated by $y_1, \ldots, y_t$, we have $\mathbb{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0, \quad \mathbb{E}(\zeta_t|\mathcal{F}_{t-1}) = 0, \quad \mathbb{E}(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) = \Sigma_\varepsilon, \quad \mathbb{E}(\zeta_t\zeta_t'|\mathcal{F}_{t-1}) = \Sigma_\zeta,$ for $t = 1, \ldots, T$ and with $\mathcal{F}_0 = \emptyset$.
3. Sequences $\{\varepsilon_t\}$ and $\{\zeta_t\}$ are uncorrelated and have finite fourth moments.
4. Sequence $\{\bar{X}_t\}$ is independent of $\{\varepsilon_t\}$ and $\{\zeta_t\}$.
5. Matrix $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \bar{X}_t\bar{X}_{t+j}' = \Gamma_X(j)$ exists and is finite for every integer j.
6. The permissible parameter space S_ψ is a compact subset of the Euclidean space. The true parameter ψ_0 is an interior point of S_ψ.
7. For ψ as an element of S_ψ and $\psi \neq \psi_0$, $t = 0, 1, \ldots$ and $\psi^* \in S_\psi$, let $\Gamma_y(t; \psi^*) = \mathbb{E}_{\psi^*}(y_1 y_{1+t}')$ denote the autocovariance function. Then $\Gamma_y(s; \psi) \neq \Gamma_y(s; \psi_0)$ for at least one value of s.
   Generally, all parameters in $\bar{\Lambda}$, $\Phi_1, \ldots, \Phi_{q_\Phi}$, $\Theta_1, \ldots, \Theta_{q_\Theta}$ and Σ_ζ cannot be identified: it is necessary to restrict matrices depending on model.
8. The process $y_t - \bar{\mu} - \bar{X}_t\beta$ can be written as a VAR process $\Pi(L)y_t = e_t$ where $\Pi(z) = I - \sum_{i=1}^{\infty} \Pi_i z^i$, $\mathbb{E}(e_t|\mathcal{F}_{t-1}) = 0$ and the elements of $\Pi_1, \Pi_2 \ldots$ are absolutely summable.

Contributions

We present new results that lead to computationally efficient methods for a likelihood-based analysis of high-dimensional dynamic factor models:

- Signal extraction;
- Likelihood evaluation; marginal likelihood (not here);
- Parameter estimation via maximum likelihood;
- Forecasting (not here);
- Bayesian estimation using MCMC (not here).

Illustration 1: We treat a panel of 132 time series from which four or seven dynamic factors are extracted. Such models require the estimation of more than 1000 parameters.
Illustration 2: A World Yield Curve...

The new results of this paper enable us
to estimate this set of parameters as a matter of routine

A basic example

Consider the basic dynamic factor model, for $N \times 1$ observation vector y_t and $r \times 1$ latent factor vector f_t, as given by

$$y_t = \Lambda f_t + \varepsilon_t, \qquad f_t = \Phi f_{t-1} + \zeta_t, \qquad t = 1, \ldots, T,$$

where $\varepsilon_t \sim IID(0, \Sigma_\varepsilon)$ and $\zeta_t \sim IID(0, \Sigma_\zeta)$. For identification purposes, we have $vec(\Sigma_\zeta) = (I - \Phi \otimes \Phi)^{-1} vec(I)$. In other words, factors f_t are standardized.

A basic example (cont.)

For $N \times 1$ data vector y_t and r factors in f_t, the basic DFM is

$$y_t = \Lambda f_t + \varepsilon_t, \qquad f_t = \Phi f_{t-1} + \zeta_t, \qquad t = 1, \ldots, T.$$

Cross-section dimension N is typically high and time series length T is moderate.

- We are interested in $N >> T$.
- Estimation concentrates on Λ, Σ_ε and Φ.
- However, first we concentrate on
  - signal extraction of f_t,
  - likelihood evaluation,

  for **given** values of Λ, Σ_ε and Φ.

Signal extraction

Model

$$y_t = \Lambda f_t + \varepsilon_t, \qquad f_{t+1} = \Phi f_t + \zeta_t, \qquad t = 1, \ldots, T.$$

can be viewed as a state space model with f_t as the state vector.

Likelihood evaluation is based on predicion error decomposition

$$\ell = p(y_1) \prod_{t=2}^{T} p(y_t | y_1, \ldots, y_{t-1}),$$

and is routinely computed by the Kalman filter. Evaluation of

$$\widetilde{f}_t = E(f_t | y_1, \ldots y_s), \qquad Var(f_t | y_1, \ldots y_s), \qquad s = t-1, \ldots, T,$$

for $t = 1, \ldots, T$ is carried out by Kalman filter and related methods.

Kalman filter methods often dismissed as N becomes very large :(

Transformation by regression

However, huge computational gains can be obtained as follows :)
Model

$$y_t = \Lambda f_t + \varepsilon_t, \qquad f_t = \Phi f_{t-1} + \zeta_t, \qquad t = 1, \ldots, T.$$

Carry out OLS based on “covariate” matrix Λ (is given), for every t:

$$\widehat{f}_t = P y_t, \qquad \text{where} \quad P = (\Lambda'\Lambda)^{-1}\Lambda'.$$

Then, transform model for y_t to a model for $\widehat{f}_t$, that is

$$\widehat{f}_t = f_t + e_t, \qquad f_t = \Phi f_{t-1} + \zeta_t, \qquad t = 1, \ldots, T,$$

with $e_t = P\varepsilon_t \sim IID\{0, \sigma_\varepsilon^2(\Lambda'\Lambda)^{-1}\}$. It can be shown that

$$\widetilde{f}_t = E(f_t | y_1, \ldots y_s) = E(f_t | \widehat{f}_1, \ldots \widehat{f}_s), \qquad t, s = 1, \ldots, T.$$

It implies that observation equation dimension N reduces to r.

Two-step method

Model

$$y_t = \Lambda f_t + \varepsilon_t, \qquad f_t = \Phi f_{t-1} + \zeta_t, \qquad t = 1, \ldots, T,$$

for known Λ, Φ, Σ_ε.

Signal extraction for f_t is carried out in two steps:

1. Cross-section step: OLS (or GLS later ...)

$$\widehat{f}_t = (\Lambda'\Lambda)^{-1}\Lambda' y_t.$$

2. Time series step: use Kalman filter methods to evaluate $\widetilde{f}_t = E(f_t | y_1, \ldots y_s)$ based on low-dimensional model

$$\widehat{f}_t = f_t + e_t, \qquad f_t = \Phi f_{t-1} + \zeta_t, \qquad e_t \sim IID\{0, \sigma_\varepsilon^2 (\Lambda'\Lambda)^{-1}\}.$$

It turns out that **all** inference can be based on this model for $\widehat{f}_t$, including the evaluation of the likelihood function.

Towards a more general DFM model: static form

Static form of the DFM model $y_t = \bar{\mu} + \bar{X}_t\beta + \Lambda(L)f_t + u_t$ with VARMA factors $\Phi(L)f_t = \Theta(L)\zeta_t$ and VAR disturbances $u_t = \Psi(L)\varepsilon_t$ is given by

$$y_t = \mu + d_t + X_t\beta + \Lambda F_t + \varepsilon_t, \qquad F_t = (f_t', f_{t-1}', \ldots, f_{t-s}')',$$

where $\mu = \Psi(I)\bar{\mu}$, $d_t = \sum_{j=1}^{q_\Psi} \Psi_j y_{t-j}$, $X_t = \Psi(L)\bar{X}_t$ and $\Lambda = (\Lambda_0^*, \Lambda_1^*, \ldots, \Lambda_s^*) \ldots$ with Λ_i^* from $\Psi(L)\Lambda(L)$ and $s = \max(m, q_\Psi)$.

Towards a more general DFM model: state space form

Static form of the DFM model $y_t = \bar{\mu} + \bar{X}_t\beta + \Lambda(L)f_t + u_t$ with VARMA factors $\Phi(L)f_t = \Theta(L)\zeta_t$ and VAR disturbances $u_t = \Psi(L)\varepsilon_t$ is given by

$$y_t = \mu + d_t + X_t\beta + \Lambda F_t + \varepsilon_t, \qquad F_t = (f_t', f_{t-1}', \ldots, f_{t-s}')',$$

where $\quad \mu = \Psi(I)\bar{\mu}, \quad d_t = \sum_{j=1}^{q_\Psi} \Psi_j y_{t-j}, \quad X_t = \Psi(L)\bar{X}_t \quad$ and $\Lambda = (\Lambda_0^*, \Lambda_1^*, \ldots, \Lambda_s^*) \ldots \quad$ with Λ_i^* from $\Psi(L)\Lambda(L)$ and $s = \max(m, q_\Psi)$.

State space form is given by the observation equation

$$y_t = \mu + d_t + X_t\beta + Z\alpha_t + \varepsilon_t,$$

where $\alpha_t = (f_t', f_{t-1}', \ldots, f_{t-\kappa}')'$ with $\kappa = \max(s, q_\Phi, q_\Theta + 1)$ such that we obtain the companion form for the VARMA factors (state equation)

$$\alpha_t = H\alpha_{t-1} + R\zeta_t, \qquad t = 1, \ldots, T.$$

Since F_t is a sub-set of α_t, we have $Z = \Lambda G, \qquad G = (I_{r\cdot s}, 0)$.

Transforming the observation vector

Consider model $y_t = \mu + d_t + X_t\beta + Z\alpha_t + \varepsilon_t$ with $Z = \Lambda G$.

Transform $y_t^+ = Ay_t$, for $t = 1, \ldots, T$, for some non-singular matrix A: MMSLEs are not affected and loglikelihood function differs only by the Jacobian term $\log |A|^T$.

$$A = \begin{bmatrix} A^L \\ A^H \end{bmatrix}, \qquad y_t^+ = \begin{pmatrix} y_t^L \\ y_t^H \end{pmatrix},$$

where $\quad y_t^L = A^L(y_t - \mu - d_t - X_t\beta), \qquad y_t^H = A^H(y_t - \mu - d_t - X_t\beta)$.
Choose A s.t.

$$y_t^L = A^L Z\alpha_t + e_t^L, \qquad\qquad y_t^H = e_t^H,$$

$$\begin{pmatrix} e_t^L \\ e_t^H \end{pmatrix} \sim \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \Sigma_L & 0 \\ 0 & \Sigma_H \end{bmatrix} \right\}, \qquad \text{with} \quad \begin{aligned} \Sigma_L &= A^L \Sigma_\varepsilon A^{L\prime}, \\ \Sigma_H &= A^H \Sigma_\varepsilon A^{H\prime}. \end{aligned}$$

Conditions for transformation

A suitable matrix A needs to fulfill the following conditions:

1. A is full rank, prevents any loss of information;
2. $A^H \Sigma_\varepsilon A^{L\prime} = 0$, ensures that both equations are independent;
3. Row$\{A^H\}$ = Col$\{Z\}^\perp$ with $Z = \Lambda G$, implies that y_t^H does not depend on α_t (can be weakened);

LEMMA 1:

Matrix A satisfies these conditions if and only if

$$A^L = C\Lambda^{\dagger\prime}\Sigma_\varepsilon^{-1},$$

for some nonsingular $r^\dagger \times r^\dagger$ matrix C and for some $N \times r^\dagger$ matrix $\Lambda^\dagger$; the $r^\dagger$ columns of $\Lambda^\dagger$ form a basis for the column space of Λ.

For this matrix A^L, we can always find a matrix A^H that satisfies 1,2,3. However, we will not need A^H in our treatment.

Illustration

Consider the one-factor model

$$y_t = \Lambda f_t + \varepsilon_t, \qquad f_t = \phi f_{t-1} + \zeta_t + \theta \zeta_{t-1}, \qquad t = 1, \ldots, T.$$

Apply transformation based on $A^L = C\Lambda' \Sigma_\varepsilon^{-1}$ with $C = \left(\Lambda' \Sigma_\varepsilon^{-1} \Lambda\right)^{-1}$.

For this choice of C, the scalar y_t^L is effectively the generalised least squares (GLS) estimator of f_t in the "regression model" $y_t = \Lambda f_t + \varepsilon_t$, for a given t. We have

$$y_t^L = \left(\Lambda' \Sigma_\varepsilon^{-1} \Lambda\right)^{-1} \Lambda' \Sigma_\varepsilon^{-1} y_t, \qquad t = 1, \ldots, T.$$

The model for the univariate time series y_t^L is then given by

$$y_t^L = f_t + e_t^L, \qquad \mathbb{E}(e_t^L e_t^{L\,\prime} | \mathcal{F}_{t-1}) = C, \qquad t = 1, \ldots, T.$$

An additional condition for convenience

A suitable matrix A needs to fulfill the following conditions:

1. A is full rank, prevents any loss of information;
2. $A^H \Sigma_\varepsilon A^{L\prime} = 0$, ensures that both equations are independent;
3. Row$\{A^H\}$ = Col$\{Z\}^\perp$ with $Z = \Lambda G$, implies that y_t^H does not depend on α_t (can be weakened);
4. $|\Sigma_H| = 1$ where $\Sigma_H = A^H \Sigma_\varepsilon A^{H\prime}$

The additional fourth condition is not restrictive, it is about scaling and it simplifies various calculations.

For example, from the fourth condition, it follows that

$$|A|^2 = |\Sigma_\varepsilon|^{-1} |A \Sigma_\varepsilon A'| = |\Sigma_\varepsilon|^{-1} |A^L \Sigma_\varepsilon A^{L\prime}||A^H \Sigma_\varepsilon A^{H\prime}| = |\Sigma_\varepsilon|^{-1} |\Sigma_L|.$$

Particularly convenient for likelihood evaluation, next.

Likelihood evaluation

Gaussian likelihood (GL) based on transformation via A is

$$\ell(y;\psi) = \ell(y^L;\psi) + \ell(y^H;\psi) + T\log|A|, \quad |A|^2 = |\Sigma_\varepsilon|^{-1}|\Sigma_L|.$$

The first term $\ell(y^L;\psi)$ can be evaluated by the Kalman filter.
The second term is

$$\ell(y^H;\psi) = -\frac{(N-m)T}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T} y_t^{H\,\prime}\Sigma_H^{-1}y_t^H,$$

as the log-determinental term vanishes since $|\Sigma_H| = 1$ (condition 4).

LEMMA 2:

$$y_t^{H\,\prime}\Sigma_H^{-1}y_t^H = e_t'\Sigma_\varepsilon^{-1}e_t,$$

where $e_t = y_t - \Lambda^\dagger\left(\Lambda^{\dagger\,\prime}\Sigma_\varepsilon^{-1}\Lambda^\dagger\right)^{-1}\Lambda^{\dagger\,\prime}\Sigma_\varepsilon^{-1}y_t$ is the GLS residual for data-vector y_t, covariate matrix $\Lambda^\dagger$ and variance matrix Σ_ε.
Choice of C is irrelevant.

Sketch of Proof Lemma 2

$$\begin{aligned} y_t^{H\prime}\Sigma_H^{-1}y_t^H &= (y_t - d_t)' A^{H\prime}(A^H\Sigma_\varepsilon A^{H\prime})^{-1}A^H(y_t - d_t) \\ &= (y_t - d_t)' J^H \Sigma_\varepsilon^{-1}(y_t - d_t), \end{aligned}$$

where $J^H \stackrel{def.}{=} A^{H\prime}(A^H\Sigma_\varepsilon A^{H\prime})^{-1}A^H\Sigma_\varepsilon$ is the projection matrix for a GLS with covariate matrix $A^{H\prime}$ and variance matrix Σ_ε^{-1}. Similarly, define $J^L \stackrel{def.}{=} A^{L\,\prime}(A^L\Sigma_\varepsilon A^{L\,\prime})^{-1}A^L\Sigma_\varepsilon$ as the GLS projection matrix for covariate matrix $A^{L\prime}$ and variance matrix Σ_ε^{-1}.

Since A is full rank and $A^L\Sigma_\varepsilon A^{H\prime} = 0$, we must have $J^H = I - J^L$. The definition of A^L implies that $J^H = I - \Sigma_\varepsilon^{-1}\Lambda^\dagger(\Lambda^{\dagger\,\prime}\Sigma_\varepsilon^{-1}\Lambda^\dagger)^{-1}\Lambda^{\dagger\,\prime}$ and

$$J^{H\prime} = \Sigma_\varepsilon A^{H\prime}(A^H\Sigma_\varepsilon A^{H\prime})^{-1}A^H = I - \Lambda^\dagger(\Lambda^{\dagger\,\prime}\Sigma_\varepsilon^{-1}\Lambda^\dagger)^{-1}\Lambda^{\dagger\,\prime}\Sigma_\varepsilon^{-1} \stackrel{def.}{=} M_\Lambda.$$

Proof is completed by $J^H\Sigma_\varepsilon^{-1} = J^H\Sigma_\varepsilon^{-1}J^{H\prime}$ and $e_t \stackrel{def.}{=} M_\Lambda(y_t - d_t)$.

Likelihood evaluation

Gaussian likelihood (GL) can now be expressed as

$$\ell(y;\psi) = c + \ell(y^L;\psi) - \frac{T}{2}\log\frac{|\Sigma_\varepsilon|}{|\Sigma_L|} - \frac{1}{2}\sum_{t=1}^{T} e_t'\Sigma_\varepsilon^{-1}e_t,$$

where c is a constant independent of both y and ψ.

It follows that for the evaluation of the loglikelihood, computation of matrix A^H and vectors y_t^H, for $t = 1,\ldots,T$, is not required.

Matrix Σ_ε is oftentimes treated as diagonal or has other strong structure (blocks, bands, spatial). Term $|\Sigma_L|$ is delivered by KFS.

This GL expression is instrumental for a computationally feasible approach to a quasi-likelihood based analysis of the dynamic factor model.

Regression part

The DFM model in vector form is

$$y_t = \bar{\mu} + \bar{X}_t\beta + \Lambda(L)f_t + u_t,$$

$$\Phi(L)f_t = \Theta(L)\zeta_t, \qquad \Psi(L)u_t = \varepsilon_t, \qquad t = 1, \ldots, T,$$

Within this approach, the estimation of constant vector $\bar{\mu}$ and β is treated within the same transformation and at no additional computational cost of any significance.

The calculations can be done in the same two-step procedure.

This may seem obvious given the linear model settings.
However:

the devil is in the detail...

Consequently, the derivations are lengthy.

Computational gains

The two panels below present the gains in computing time when evaluating the loglikelihood respectively the diffuse loglikelihood functions of two types of dynamic factor models. Model A is of the form $y_{it} = \lambda_i' f_t + \varepsilon_{it}$ and model B of the form $y_{it} = \mu_i + \lambda_i' f_t + \varepsilon_{it}$, where f_t is a VAR(1), $\varepsilon_{it} \sim IID(0, \sigma^2)$, for some positive scalar σ and μ_i is a scalar. The ratio d_1/d_2 is reported: d_1 is the CPU time for the standard (diffuse) Kalman filter and d_2 is CPU time for our new algorithms. The ratios are reported for different panel dimensions N and different state vector dimensions p.

	Model A					Model B				
$N \backslash p$	1	5	10	25	50	1	5	10	25	50
10	2.0	1.3	–	–	–	10.4	2.3	–	–	–
50	5.7	4.7	3.1	1.5	–	50.6	40.0	18.0	3.4	–
100	6.7	7.5	5.6	2.5	1.5	55.0	62.0	47.2	13.5	3.2
250	8.7	14.8	12.4	5.5	3.0	79.0	82.2	82.9	63.6	22.6
500	12.5	15.9	21.2	10.2	5.4	107.5	108.9	109.5	108.7	69.7

Maximum likelihood estimation

The DFM model in vector form is

$$y_t = \bar{\mu} + \bar{X}_t\beta + \Lambda(L) f_t + u_t,$$

$$\Phi(L) f_t = \Theta(L)\zeta_t, \qquad \Psi(L) u_t = \varepsilon_t, \qquad t = 1, \ldots, T.$$

All coefficients in $\Lambda(L)$, $\Phi(L)$, $\Theta(L)$ and $\Psi(L)$ and Σ_ε are collected in ψ which can be potentially large.

No miracles here, just hard work:

- EM algorithm;
- direct likelihood maximization based on analytical score for ψ.

Main computational work for both EM (Watson and Engle) and analytical score (Koopman and Shephard) relies on Kalman filter methods and can take advantage of the results given here.

Final comment: don't shy away from maximizing a likelihood function in a 1000-dimensional space: repeated score evaluations are informative.

Illustration 1: a macro-economic panel

We consider the "sims.xls" data file, used in Stock and Watson (2005) “Implication of DFM for VAR”, obtained from Mark W. Watson website.

The data file consists of $N = 132$ variables and we used the balanced sample of Sept 1960 – Dec 2003 ($T = 515$).

The model is given by

$$y_t = \mu + \Lambda f_t + u_t, \qquad f_t = \Phi f_{t-1} + \zeta_t, \qquad t = 1, \ldots, T,$$

where

$$u_{i,t} = \rho_i u_{i,t-1} + \varepsilon_{it}, \qquad i = 1, \ldots, N,$$

with $\varepsilon_{it} \sim IID(0, \Sigma_\varepsilon)$, $\zeta_t \sim IID(0, \Sigma_\zeta)$, $\text{vec}(\Sigma_\zeta) = (I - \Phi \otimes \Phi)^{-1}\text{vec}(I)$.

The number of factors are (i) $r = 7$ with $\rho_i = 0$; (ii) $r = 4$ with $\rho_i \neq 0$.

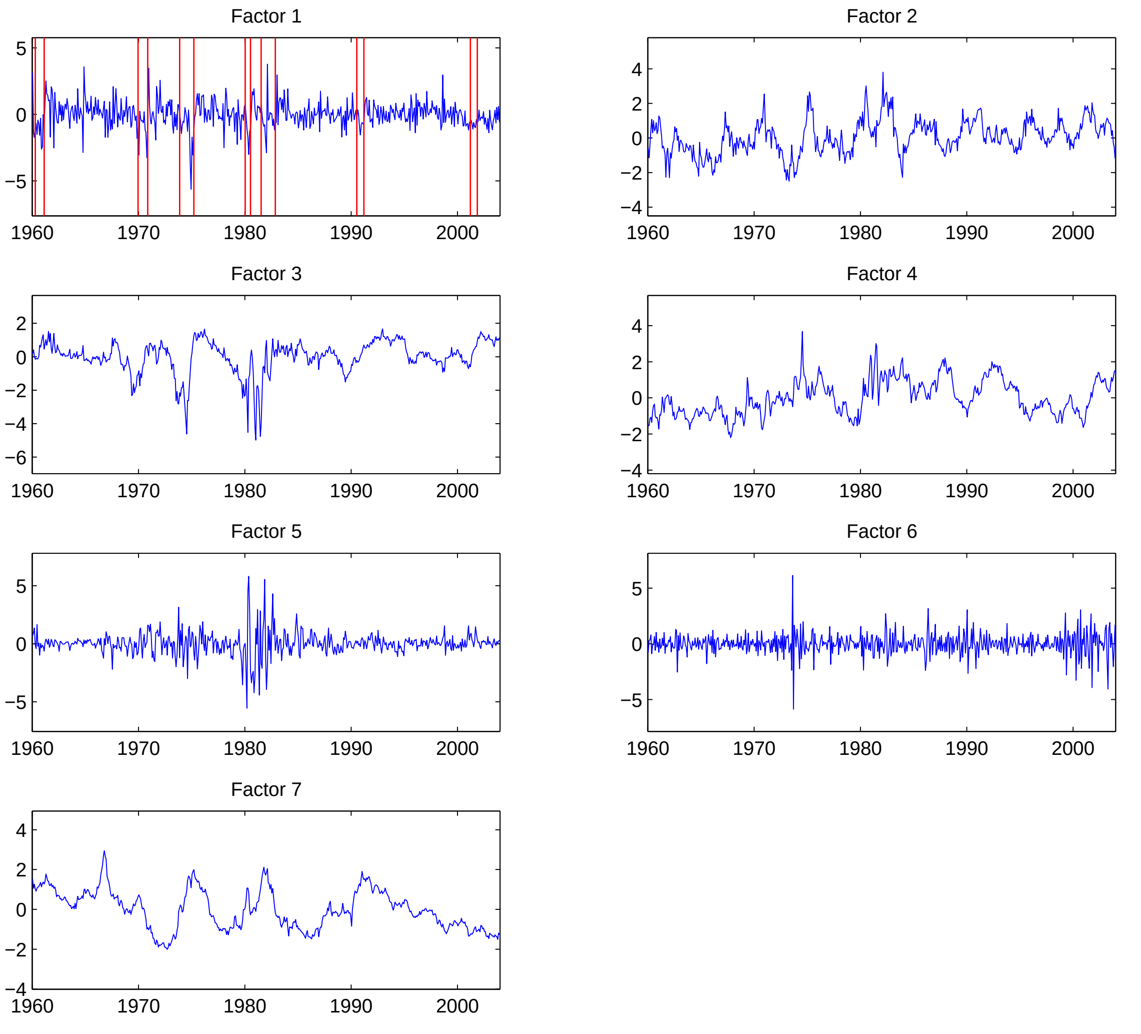
Further, Λ full, Φ full, Σ_ε diagonal: more than 1000 parameters ...

Estimation has taken place using the new methods as presented.

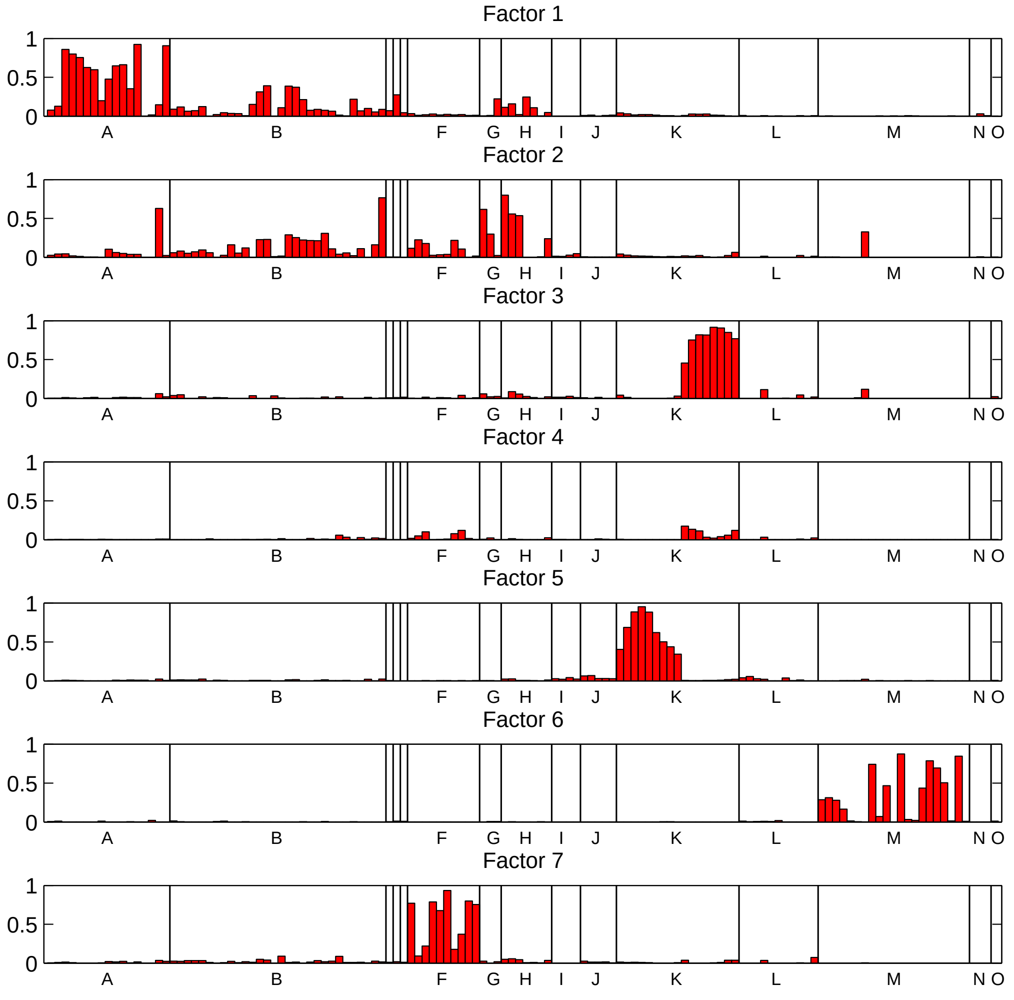
Data-set

Code	Description	Number of Time Series
A	Real Output and Income	17
B	Employment and Hours	30
C	Real Retail	1
D	Manufacturing and Trade Sales	1
E	Consumption	1
F	Housing Starts and Sales	10
G	Real Inventories	3
H	Orders	7
I	Stock Prices	4
J	Exchange Rates	5
K	Interest Rates and Spreads	17
L	Money and Credit Quantity Aggregates	11
M	Price Indexes	21
N	Average Hourly Earnings	3
O	Miscellanea	1

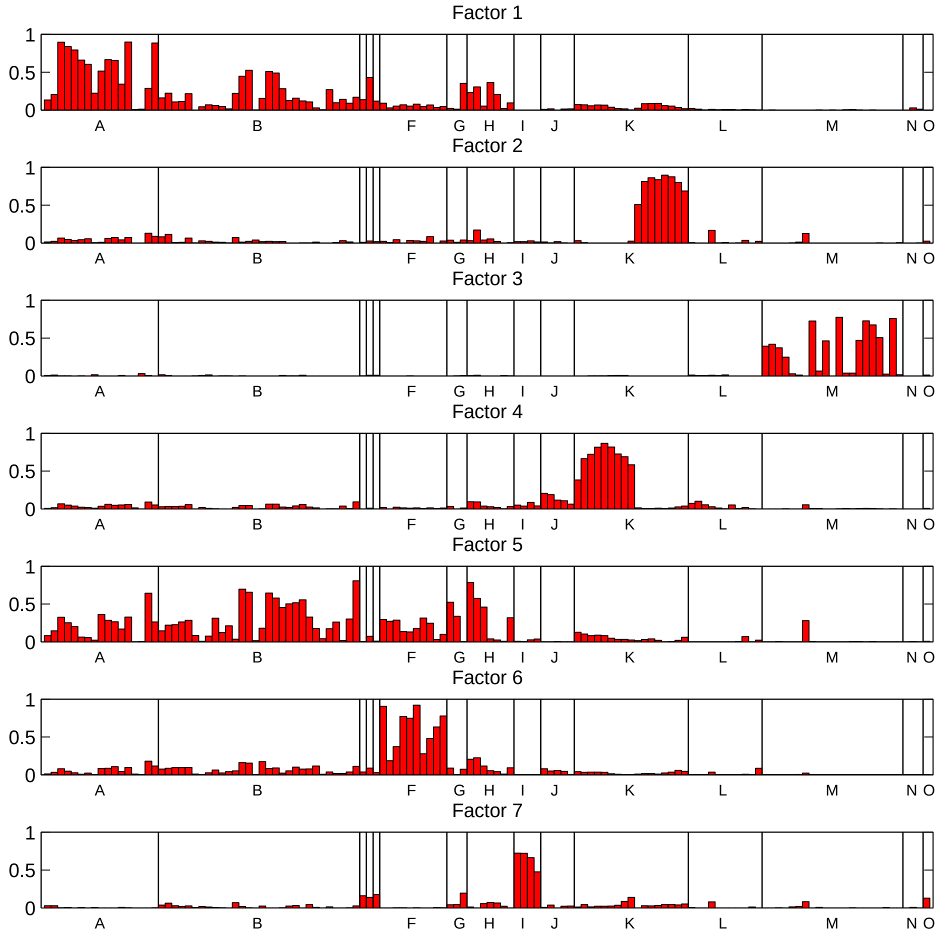
Estimated seven factors (without AR errors)



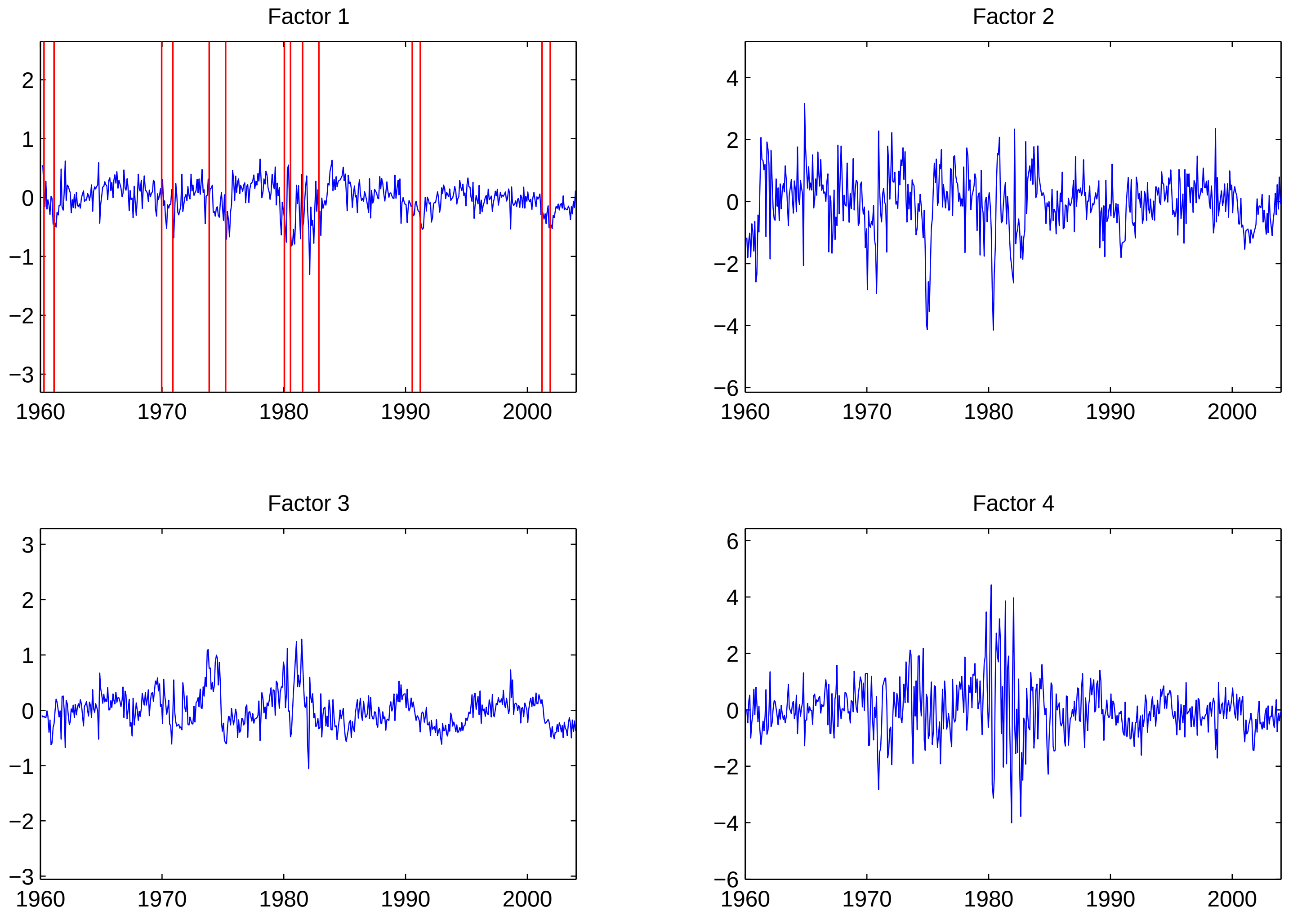
R^2 for seven factors



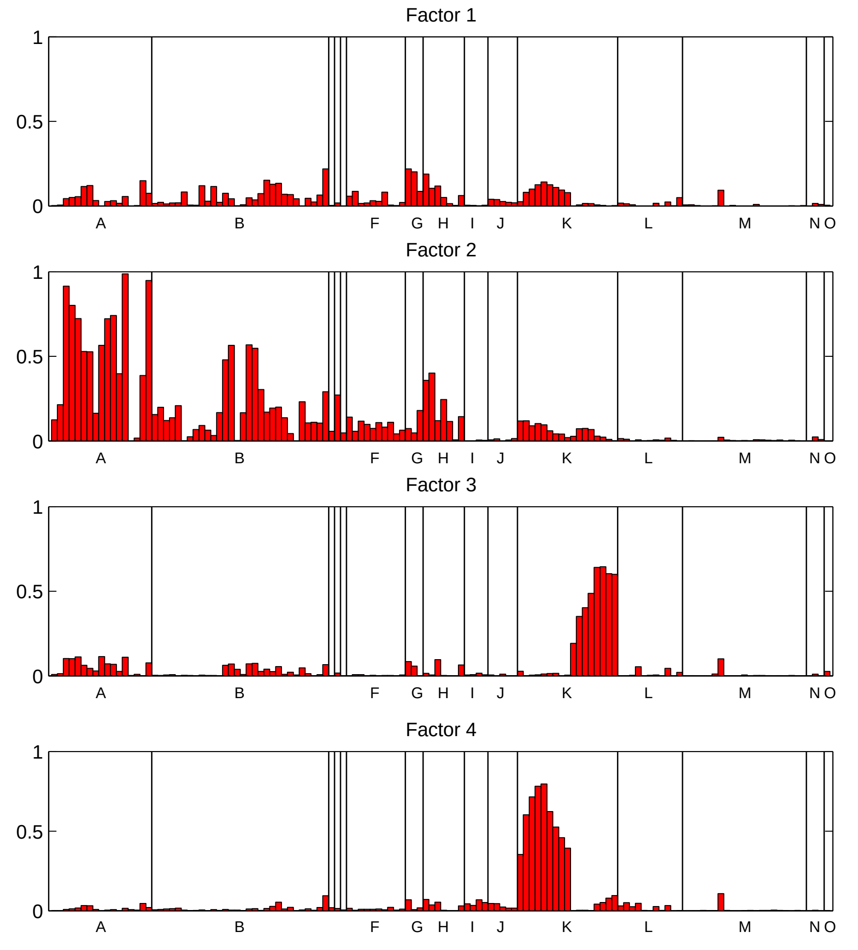
R^2 for seven PCs from Stock-Watson procedure



Estimated four factors (with AR errors)



R^2 for four factors



Box-Ljung Q(5) statistics

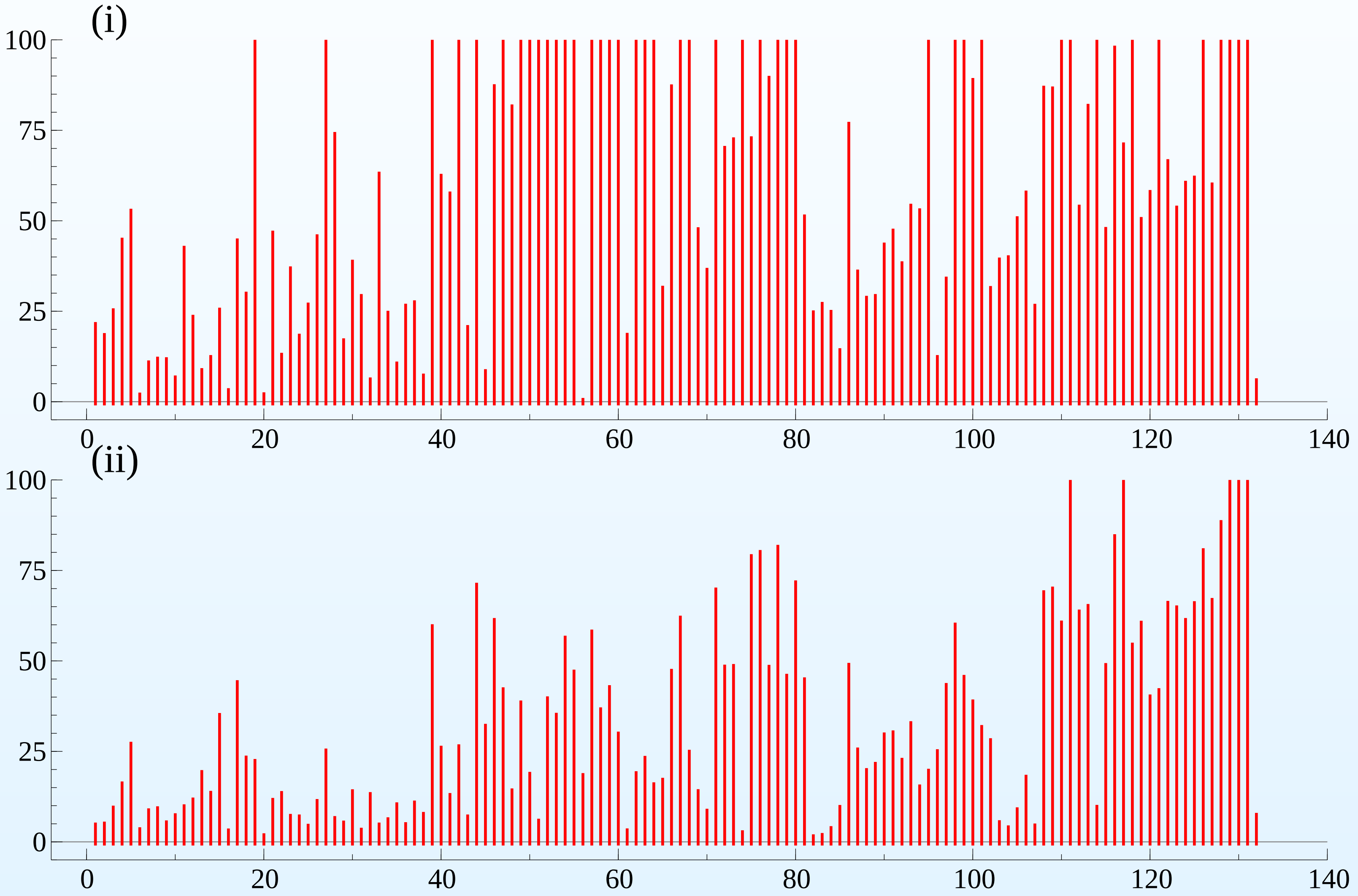


Illustration 2: A World Yield Curve

The US yield curve (Fama-Bliss)

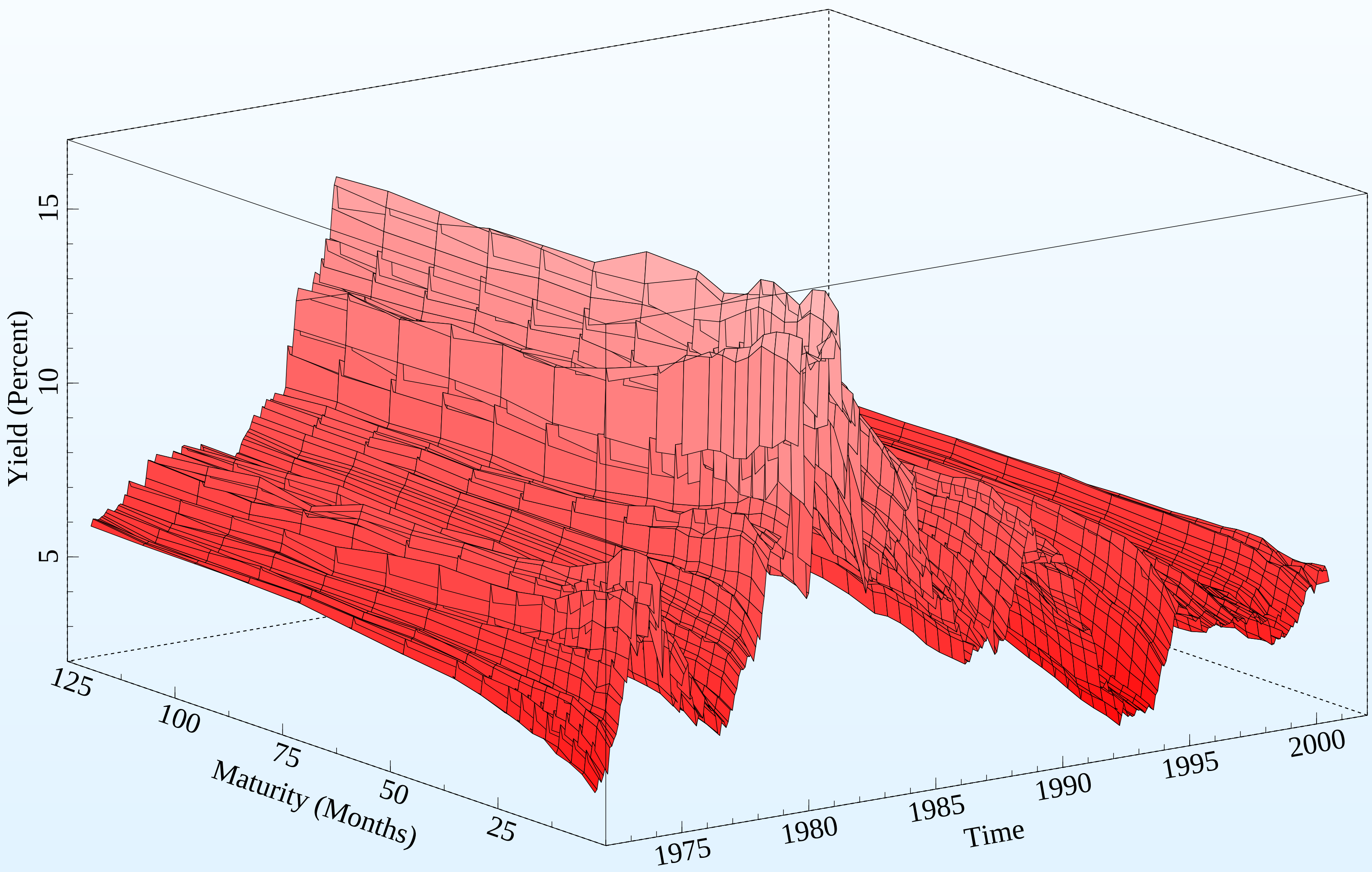


Illustration 2: A dynamic Yield Curve model

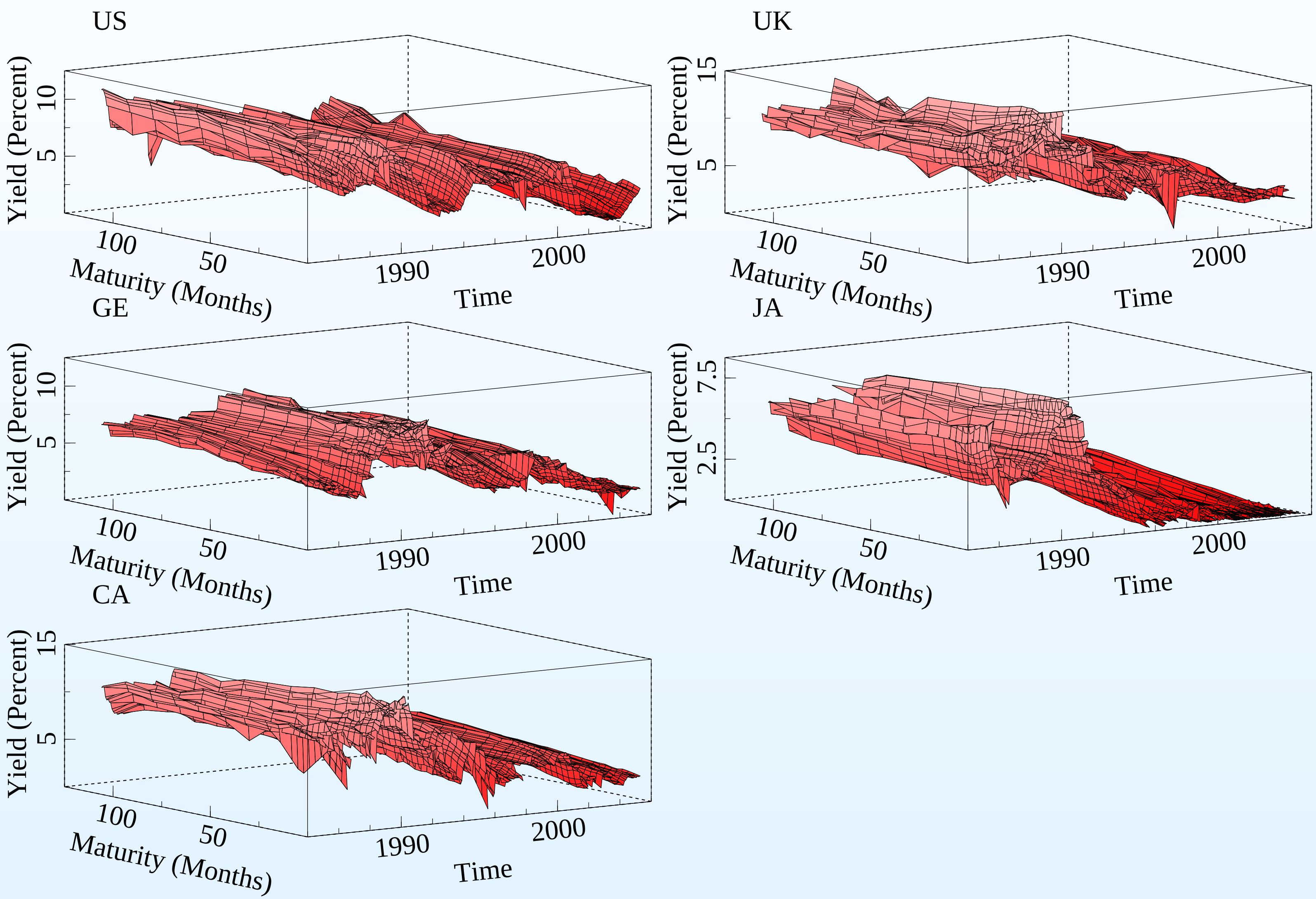
The dynamic Nelson-Siegel model of Diebold, Rudebusch and Aruoba (2006) is given by the dynamic 3-factor model:

$$
\begin{aligned}
y_t &= \Gamma(\lambda) f_t + \varepsilon_t, & \varepsilon_t &\sim NID(0, \Sigma_\varepsilon), \\
f_{t+1} &= (I - \Phi)\mu + \Phi f_t + \eta_t, & \eta_t &\sim NID(0, \Sigma_\eta),
\end{aligned}
$$

with y_t containing typically 17 maturities ($N = 17$),

$$
\begin{aligned}
y_t &= (y_t(\tau_1), \ldots, y_t(\tau_N))', \\
\Gamma_{ij}(\lambda) &= \begin{cases} 1, & j = 1, \\ \left(1 - e^{-\lambda \cdot \tau_i}\right) \ / \ \lambda \cdot \tau_i, & j = 2, \\ \left(1 - e^{-\lambda \cdot \tau_i} - \lambda \cdot \tau_i e^{-\lambda \cdot \tau_i}\right) \ / \ \lambda \cdot \tau_i, & j = 3, \end{cases} \\
f_t &= (f_{1t}, f_{2t}, f_{3t})', \qquad (f_{1t} = \text{level} \ , \ f_{2t} = \text{slope} \ , \ f_{3t} = \text{curvature}) \\
\varepsilon_t &= (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})', \\
\eta_t &= (\eta_{1t}, \eta_{2t}, \eta_{3t})'.
\end{aligned}
$$

Yield Curve Data for more countries



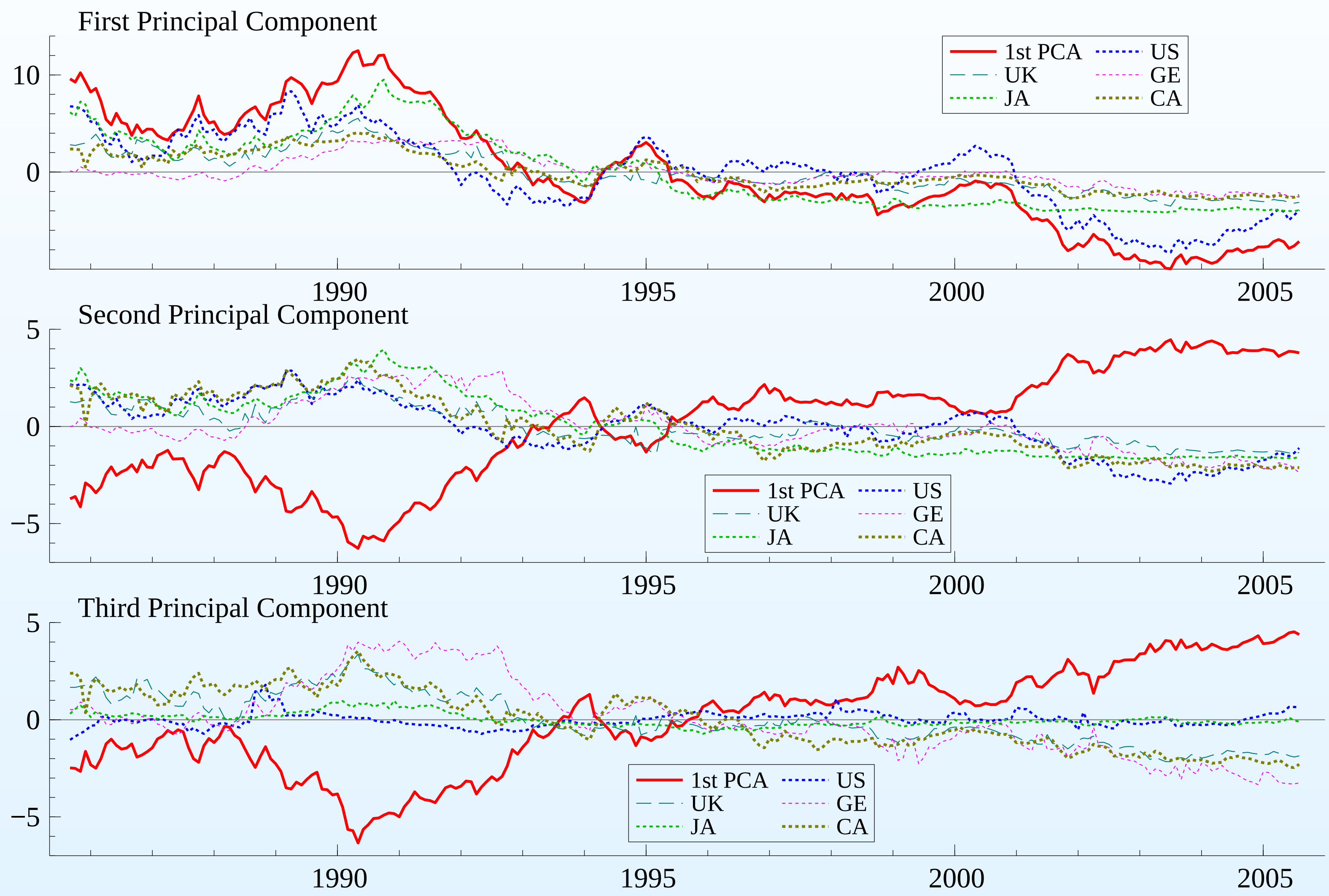
A World Yield Curve

Diebold, Li and Yue (2007) propose a dynamic factor analysis for obtaining a World Yield Curve. Due to the "curse of dimensionality" problem, they resort to MCMC analysis.

In joint work with Michel van der Wel (Rotterdam, Aarhus) and Borus Jungbacker, we consider

- ML estimation (based on presented results);
- Different structures for introducing the "factors" of "factors":
  - level factors of all countries can be described by one or two common level factors;
  - the same can be applied to the slope and curvature factors but using different specifications.
- "unrestricted" dynamic factor analysis can be considered as well.
- specifications can be properly tested by likelihood ratio tests (at least within the same number of latent factors).

Dynamic factor estimates: preliminary results



Likelihood-based analysis for dynamic factor models

Borus Jungbacker and Siem Jan Koopman

Department of Econometrics, VU University Amsterdam

Tinbergen Institute, Amsterdam

`http://staff.feweb.vu.nl/koopman`

Thank you !