Dynamic factor analysis: a likelihood-based approach with applications in economics and finance

Siem Jan Koopman

Department of Econometrics, Vrije Universiteit Amsterdam

Tinbergen Institute

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

CEF 09 / ERCIM 09 – Limassol, Cyprus – Oct 29-31, 2009

CFE 09 / ERCIM 09

All organisers, but especially,

Erricos John Kontoghiorghes

THANK YOU !

My presentation is joint work

Presentation is based on joint work with

Borus Jungbacker

Morgan-Stanley, London

Michel van der Wel

Econometric Institute, Erasmus School of Economics,
Erasmus University Rotterdam,
ECARES, University of Aarhus

Outline

- Introduction and discussion of dynamic factor models.
- The likelihood-based approach
- Generalizations
- Computational issues
- Illustration 1: world economic model
- Illustration 2: macroeconomic analysis
- Illustration 3: modelling yield curves (interest rates)
- Conclusions

The dynamic factor model: a basic example

We consider a basic vesion of the dynamic factor model. Denote the $N \times 1$ observation vector by y_t and the $r \times 1$ vector of factors by f_t.

$$y_t = \begin{pmatrix} y_{1t} \\ \vdots \\ \vdots \\ y_{Nt} \end{pmatrix}, \qquad f_t = \begin{pmatrix} f_{1t} \\ \vdots \\ f_{rt} \end{pmatrix}, \qquad N >> r.$$

A basic dynamic factor model can represented as

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

where

$$\varepsilon_t \sim IID(0, \Sigma_\varepsilon), \qquad \zeta_t \sim IID(0, \Sigma_\zeta).$$

For identification purposes, we have $vec(\Sigma_\zeta) = (I - \Phi \otimes \Phi)^{-1} vec(I)$. In other words, factors in f_t are standardized.

Dynamic factor models

- Dynamic factor models have the typical form

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

  where f_t represents a low-dimensional dynamic factor process that loads on y_t via matrix Λ for $t = 1, \ldots, T$.

- Key contributions for far more general models 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 more ...

- Estimation treatments are based on frequency domain methods, principal components, static factor analysis, etc: likelihood-based treatments are usually dismissed on computational grounds...

First some results

We first present some results from Jungbacker & Koopman (2008) that lead to computationally efficient methods for:

- signal extraction and forecasting;
- likelihood evaluation;
- parameter estimation via maximum likelihood;
- Bayesian estimation using MCMC.

The results enable us to carry out these tasks as a matter of routine

We return to basic example

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 have $N >> r$ but it also possible to have $N >> T$.

Estimation concentrates on Λ, Σ_ε and Φ.

First we concentrate on signal extraction of f_t and likelihood evaluation for **given** values of Λ, Σ_ε and Φ.

Signal extraction

Model

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

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 (for given Λ, Σ_ε and Φ). 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 : (

Kalman filter

For example, the Kalman filter evaluates

$$\widetilde{f}_t = \mathsf{E}(f_t|y_1,\ldots,y_{t-1}), \qquad P_t = \mathsf{Var}(f_t|y_1,\ldots,y_{t-1}),$$

and for our basic dynamic factor model

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

the Kalman filter is given by the recursive equations

$$\begin{aligned}
v_t &= y_t - \Lambda\widetilde{f}_t, && N\times 1\\
V_t &= \Lambda P_t\Lambda' + \Sigma_\varepsilon, && N\times N\\
K_t &= \Phi P_t\Lambda' V_t^{-1}, && r\times N\\
\\
\widetilde{f}_{t+1} &= \Phi\widetilde{f}_t + K_t v_t, && r\times 1\\
P_{t+1} &= \Phi P_t\Phi' - K_t V_t K_t' + \Sigma_\zeta, && r\times r \qquad \text{for } t=1,\ldots,T.
\end{aligned}$$

Kalman filter modification

Key problem for Kalman filter when N is large: computation of inverse of the $N \times N$ prediction error variance matrix V_t in

$$V_t = \Lambda P_t \Lambda' + \Sigma_\varepsilon, \qquad K_t = \Phi P_t \Lambda' V_t^{-1}.$$

Since P_t has dimension $r \times r$ and $N >> r$, we can apply the "inversion lemma" to compute V_t^{-1}.

Positive:
This will lead to computational savings...

Negative:
We need to modify Kalman filter, standard software cannot be used.

Also, Lemma requires inverse of P_t but does not necessarily exist.

The stationary VARMA(p, q) model for f_t is the typical example, see Ansley and Kohn (1985).

Therefore, we propose a general and more convenient method next.

Transformation by regression

We transform the observation model in a lower dimension. The DFM is

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

for **given** values of Λ, Σ_ε and Φ.
Transform data: base it on GLS with "covariate" Λ, for every t,

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

Then, construct model for $\widehat{f}_t = P y_t$, that is

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

with $P\Lambda = I$ and $e_t = P\varepsilon_t \sim IID\{0, \sigma_\varepsilon^2 (\Lambda' \Sigma_\varepsilon^{-1} \Lambda)^{-1}\}$. We will show 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.

Kalman filter

Kalman filter still evaluates

$$\widetilde{f}_t = \mathsf{E}(f_t|y_1,\ldots,y_{t-1}), \qquad P_t = \mathsf{Var}(f_t|y_1,\ldots,y_{t-1}),$$

but for our transformed model

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

the Kalman filter becomes

$$\begin{aligned} v_t &= \widehat{f}_t - \widetilde{f}_t, \\ V_t &= P_t + (\Lambda'\Sigma_\varepsilon^{-1}\Lambda)^{-1}, \\ K_t &= \Phi P_t V_t^{-1}, \qquad \Leftarrow\ V_t \text{ is now } r\times r \text{ !!!} \end{aligned}$$

$$\begin{aligned} \widetilde{f}_{t+1} &= \Phi\widetilde{f}_t + K_t v_t, \\ P_{t+1} &= \Phi P_t\Phi' - K_t V_t K_t' + \Sigma_\zeta, \qquad \text{for } t = 1,\ldots,T. \end{aligned}$$

Two-step method

DFM 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 and known variance matrices.

Signal extraction for f_t and likelihood evaluation by two-step method:

1. Cross-section "GLS" step: compute

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

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

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

How does it work ?

DFM model $y_t = \Lambda f_t + \varepsilon_t$.

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 $\log |A|^T$.

$$A = \begin{bmatrix} P \\ R \end{bmatrix}, \qquad y_t^+ = \begin{pmatrix} \widehat{f}_t \\ u_t \end{pmatrix}, \qquad \begin{array}{ll} A & N \times N, \\ P & r \times N, \\ R & N - r \times N, \end{array}$$

where $\widehat{f}_t = Py_t$ is $r \times 1$ and $u_t = Ry_t$ is $N - r \times 1$.
Choose A s.t. $P\Lambda$ has full-column rank r, $R\Lambda = 0$ and $P\Sigma_\varepsilon R' = 0$.
Then

$$\widehat{f}_t = P\Lambda f_t + e_t, \qquad\qquad Ry_t = u_t,$$

$$\begin{pmatrix} e_t \\ u_t \end{pmatrix} \sim \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} P\Sigma_\varepsilon P' & 0 \\ 0 & R\Sigma_\varepsilon R' \end{bmatrix} \right\}.$$

Conditions for transformation

A suitable matrix $A = [P' \, , \, R']'$ needs to fulfill the following conditions:

1. A full rank (prevents any loss of information);
2. $P\Sigma_\varepsilon R' = 0$ (ensures that two equations are independent);
3. Row$\{R\}$ = Col$\{\Lambda\}^\perp$ (such that Ry_t does not depend on f_t).

LEMMA 1:

Matrix A satisfies these conditions if and only if

$$P = C\Lambda'\Sigma_\varepsilon^{-1},$$

for any nonsingular $r \times r$ matrix C and full-column rank $N \times r$ matrix Λ.

Conditions can be weakened somewhat (column space of Λ is key).

Connect Lemma 1 with well-known regression for $y = X\beta + \varepsilon$ with "hat" matrix $P = (X'X)^{-1}X'$ and "residual" matrix $M = I - XP$: P' full-column rank k, rank M is $n - k$, $MX = 0$ and $MP' = 0$.

Illustration

Consider the one-factor model

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

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

For this choice of C, the scalar $\widehat{f}_t$ 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

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

The model for the univariate time series $\widehat{f}_t$ is then given by

$$\widehat{f}_t = f_t + e_t, \qquad \mathbb{E}(e_t e_t'|\mathcal{F}_{t-1}) = C, \qquad t = 1, \ldots, T.$$

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. $R\Sigma_\varepsilon P' = 0$, ensures that both equations are independent;
3. Row$\{R\} =$ Col$\{\Lambda\}^\perp$ implies that u_t does not depend on α_t;
4. $|R\Sigma_\varepsilon R'| = 1$, is not restrictive, it simplifies various calculations.

From the fourth condition, it follows that the Jacobian term becomes

$$|A|^2 = |\Sigma_\varepsilon|^{-1}|A\Sigma_\varepsilon A'| = |\Sigma_\varepsilon|^{-1}|P\Sigma_\varepsilon P'||R\Sigma_\varepsilon R'| = |\Sigma_\varepsilon|^{-1}|\Sigma_e|.$$

where $\Sigma_e =$ Var$(P\varepsilon_t) = P\Sigma_\varepsilon P'$ and can be evaluated conveniently by Kalman filter for the low-dimensional model for $\widehat{f_t} = P\Lambda f_t + e_t$.

This condition is particularly convenient for likelihood evaluation, next.

Likelihood evaluation

Gaussian likelihood (GL) based on transformation via A is

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

where y^P is stack of $Py_t = \widehat{f}_t$ and y^R is stack of $Ry_t = u_t$.
The first term $\ell(y^P;\psi)$ is evaluated by the Kalman filter for $\widehat{f}_t$ model.
The second term is

$$\ell(y^R;\psi) = -\frac{(N-r)T}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T} u_t'(R\Sigma_\varepsilon R')^{-1}u_t,$$

since $|R\Sigma_\varepsilon R'| = 1$ from condition 4.

LEMMA 2:

$$u_t'(R\Sigma_\varepsilon R')^{-1}u_t = a_t'\Sigma_\varepsilon^{-1}a_t,$$

where $a_t = y_t - \Lambda\widehat{f}_t$ is the "GLS" residual for y_t.
Result applies for any choice of C. Result is not trivial !

Proof Lemma 2

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

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

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

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

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

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

It follows that for the evaluation of the loglikelihood, computation of matrix R and vectors u_t, 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 $|\Sigma_e|$ is covered by KFS.

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

Doz, Giannone, and Reichlin (2006) show, under mild conditions, that likelihood-based estimates are consistent for the true factors when $T \to \infty$ and $N \to \infty$, even if the model is misspecified.

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

Parameter estimation by maximum likelihood

The DFM model in vector form is

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

Coefficients in Λ, Φ and Σ_ε are collected in vector ψ (can be large).

For estimation of ψ, no apparent miracles, just labour:

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

Main computational work for both EM (Watson and Engle, 1983) and analytical score (Koopman and Shephard, 1992) relies on Kalman filter methods; we take full advantage of these and results given here.

Comment: don't shy away from maximizing a likelihood function in a 1000-dimensional space: analytical scores available and scores are informative.

Extension 1: regression effects

The DFM model with regression effects can be represented as

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

Within our dimension reduction approach, the estimation of constant vector μ and β is treated within the same transformation and at no additional computational cost of any significance.

Problem is that selection of μ and β appears in both equations:
$\widehat{f_t} = Py_t = P\mu + PX_t\beta + P\Lambda f_t + e_t$ and $Ry_t = R\mu + RX_t\beta + u_t$.
How to get a joint estimate for μ and β ?
But it can be done, even in the same two-step procedure!

Perhaps not so surprising given the linear modelling framework.
However:

the devil is in the detail...

Consequently, the derivations are lengthy, see paper.

Extension 2: including lags, generalized DFM

The generalized DFM model (without regression effects) is

$$y_t = \Lambda(L) f_t + \xi_t, \qquad \Psi(L)\xi_t = \varepsilon_t, \qquad \Theta(L) f_t = \Phi(L)\zeta_t,$$

where $A(L)$ represents a matrix polynomial. The coefficient matrices in autoregressive polynomial for ξ_t is diagonal (cross-sectional uncorr).

In state space, we can handle the general DFM by rewriting it as

$$y_t = \sum \Psi_j y_{t-j} + \Lambda^\dagger F_t + \varepsilon_t, \qquad F_t = Z_t \alpha_t, \qquad \alpha_t = T\alpha_{t-1} + R\eta_t,$$

where "signal" F_t is the stack of f_t, f_{t-1}, etc. and is sufficiently large to incorporate the necessary lags implied by polynomials $\Lambda(L)$ and $\Psi(L)$.

Loading matrix $\Lambda^\dagger$ include coefficients from matrices Λ_j and Ψ_j.
The VARMA process for f_t can be put in state space form.

Note: even more generality can be handled *but* make sure that all coefficients can be empirically identified from the data set.

Extension 3: simulation smoothing

In case a Bayesian analysis of the DFM is preferred, Markov chain Monte Carlo (MCMC) methods may typically be employed.

In a linear Gaussian context, MCMC sampling schemes for the DFM class of models have been designed, are straightforward and similar to those for standard state space models, see Fruhwirth-Schnatter (1994), Carter and Kohn (1994) and de Jong and Shephard (1995).

In particular, sampling from $p(f|y)$, with f stack of f_t and y stack of y_t, for $t = 1, \ldots, T$, is the key challenge: *simulation smoothing*.

Durbin and Koopman (2002) simulation smoother is based on an "only mean" version of the Kalman filter and smoother. Since KFS applied to f can be based on the low-dimensional model, simulation from $p(f|y)$ can also be based on the low-dimensional model:

$$f^i \sim p(f|y;\psi^{i-1}) \equiv p(f|\widehat{f};\psi^{i-1}).$$

MCMC needs multiple simulated paths of f, comp. efficiency is key.

Extension 4: non-Gaussian nonlinear panels

In DFM:

observation eq: $y_t = \Lambda f_t + \varepsilon_t \qquad or \quad y_t \sim p(y_t|f_t) = \mathsf{N}(\Lambda f_t, \Sigma_\varepsilon)$

state eq: $f_t = \Phi f_{t-1} + \zeta_t$

We can also consider observation density $p(y_t|f_t)$ to be a non-Gaussian density or nonlinear function.

In case y_t is a panel of count data time series, we may have

$$p(y_t|f_t) = \mathsf{Poisson}(\lambda_t), \qquad \lambda_t = \exp(c + \Lambda f_t).$$

This framework can be generalized in various directions.

The main motivation of this work can be found in credit risk analysis (modeling defaults), criminology (crime trends), durations, etc.

See work by Koopman and Lucas (2008, JBES), Koopman, Lucas, Ooms, van Montfort and van der Geest (2008, Statistica Neerlandica) and Koopman, Lucas and Schwaab (2009, wp).

Extension 5: unbalanced panels (missing values)

Kalman filter and smoother methods can deal with missings in y_t. What is the issue ?

Consider DFM with AR(1) disturbances

$$y_t = \Lambda f_t + \xi_t, \qquad f_t = \Phi f_{t-1} + \eta_t, \qquad \xi_t = \Psi \xi_{t-1} + \varepsilon_t,$$

and with diagonal Ψ. In an unbalanced panel, where random entries of y_t are missing, the representation

$$y_t = \Psi y_{t-1} + (\Lambda\,,\, -\Psi\Lambda)(f_t'\,,\, f_{t-1}')' + \varepsilon_t,$$

is not appropriate when missing entries appear in y_t and y_{t-1}.

The representation with ξ_t in the state vector is computationally brutal !

Jungbacker, Koopman and van der Wel (2009) have solved the problem by appropriate mixing the two representations and by putting missing y_t's temporarily in the state vector. This is a practical solution.

Illustration 1: DFM for World economy

Consider gross domestic product (GDP) of 168 countries worldwide.

The data is an unbalanced panel, many missing entries.

We aim to signal extract the world business cycle together with region-specific factors and country-specific factors.

Specific structure is placed on Λ.

Maximum likelihood estimation is feasible: estimation of basic model takes 10 minutes; estimation of a more elaborate model can take 30-40 minutes.

CFE09 presentation yesterday by *Stefano Grassi* based on working paper of Grassi, Jungbacker and Koopman (2009).

Illustration 2: 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 $\rho_i \neq 0$.

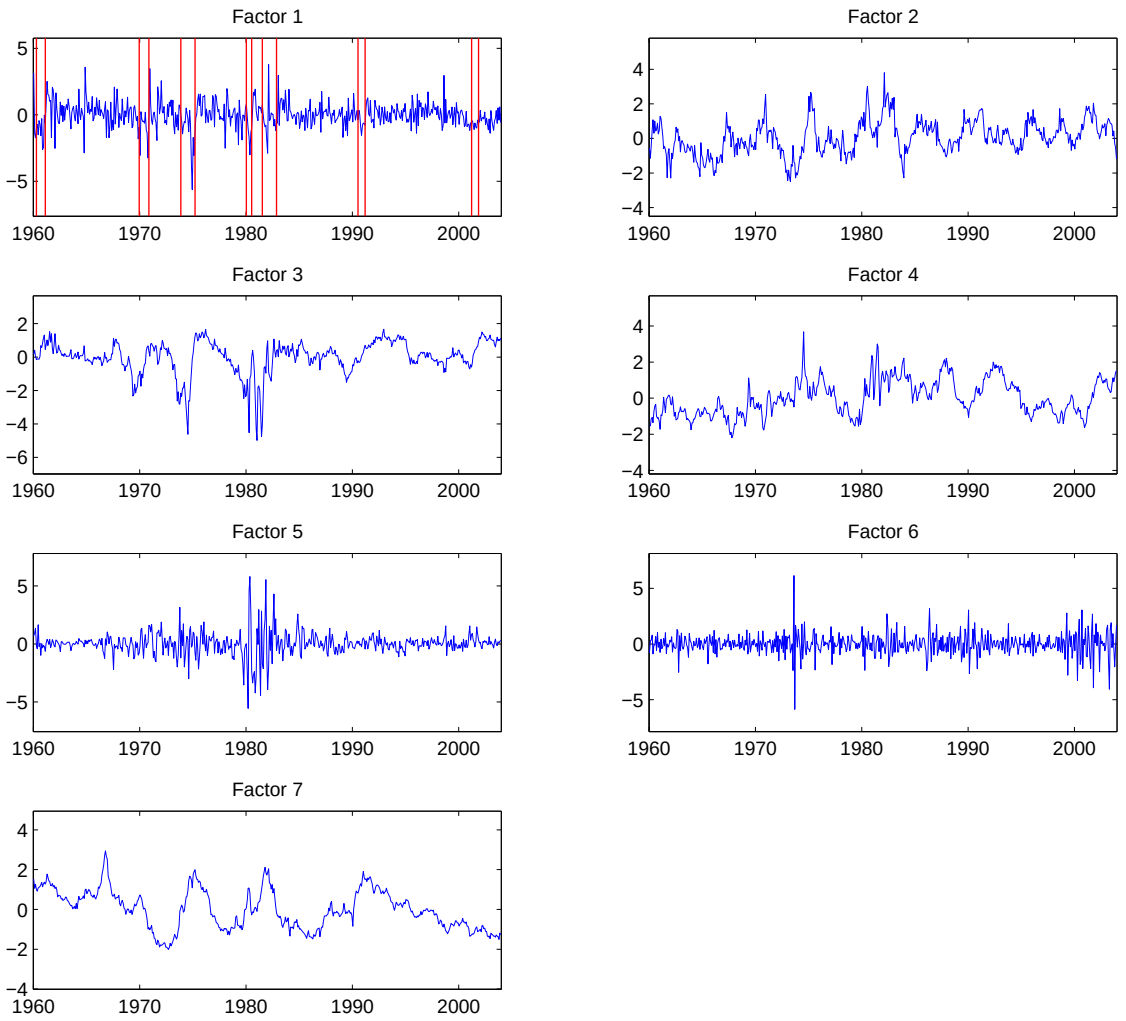
Further, Λ full, Φ full, Σ_ε diagonal: more than 1000 parameters ...

Estimation has taken place using the new methods as presented.

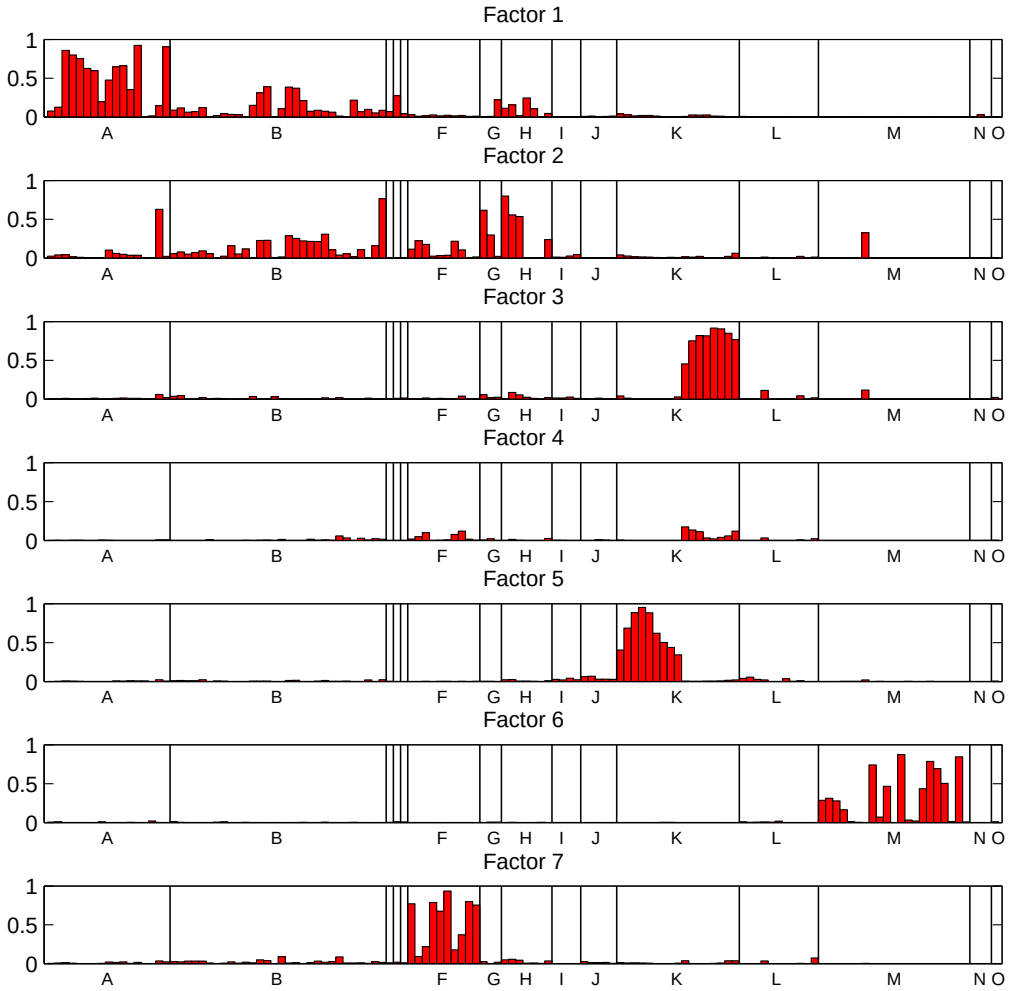
Data-set

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

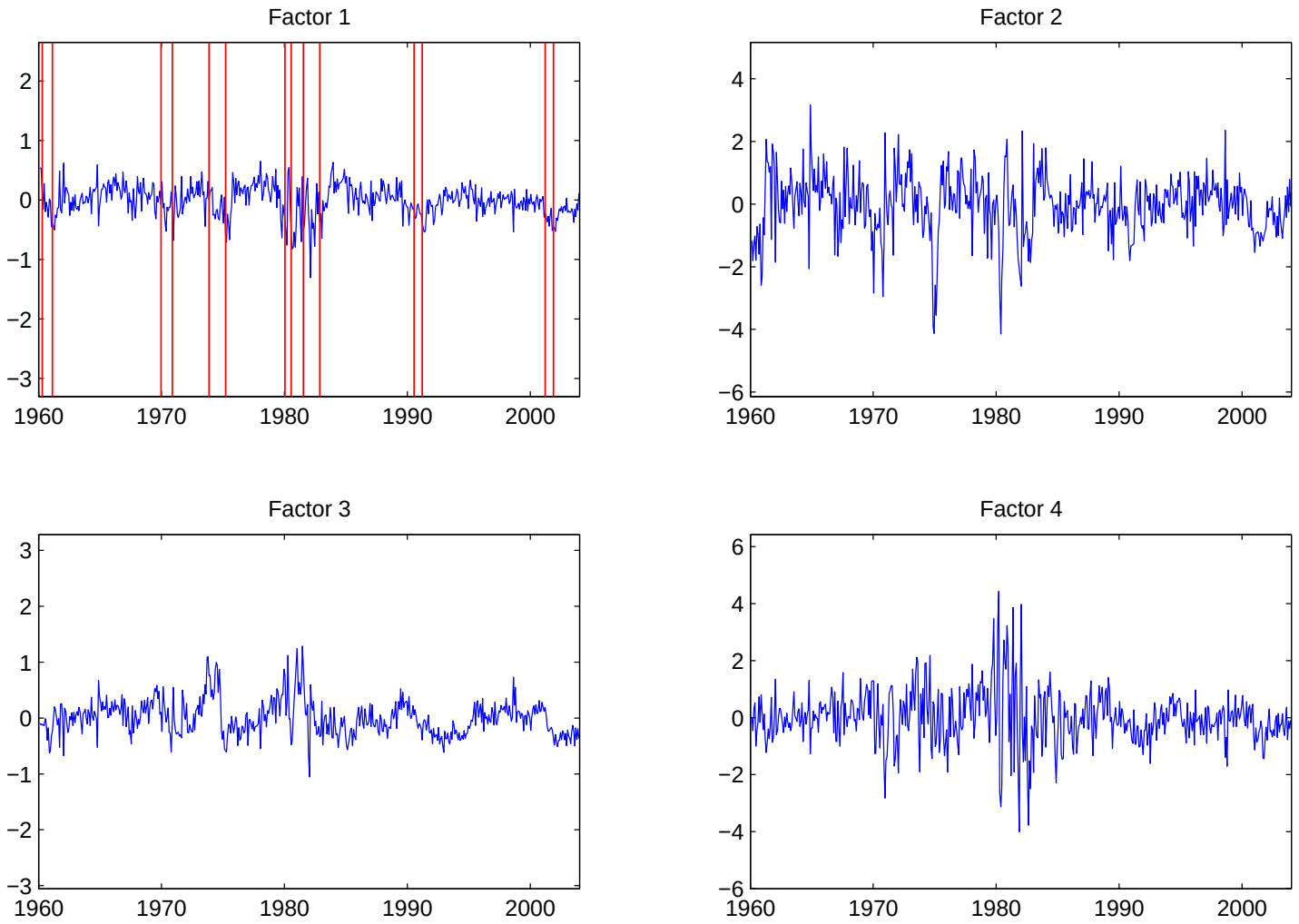
Estimated seven factors (without AR errors)



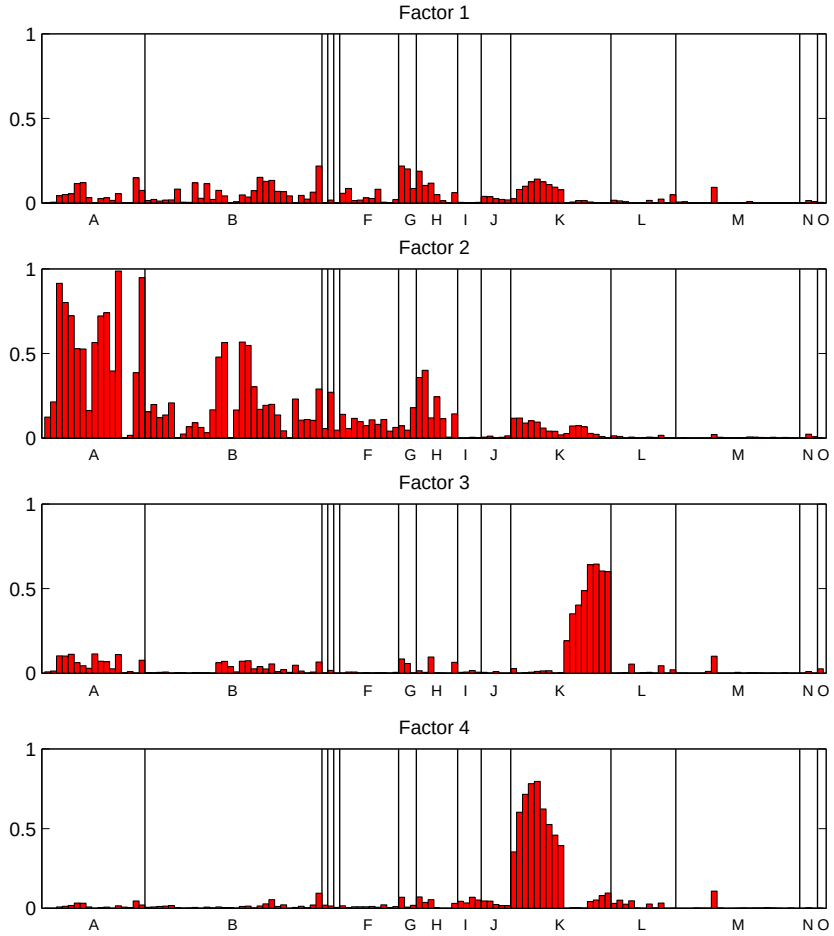
R^2 for seven factors



Estimated four factors (with AR errors)



R^2 for four factors



Box-Ljung Q(5) statistics

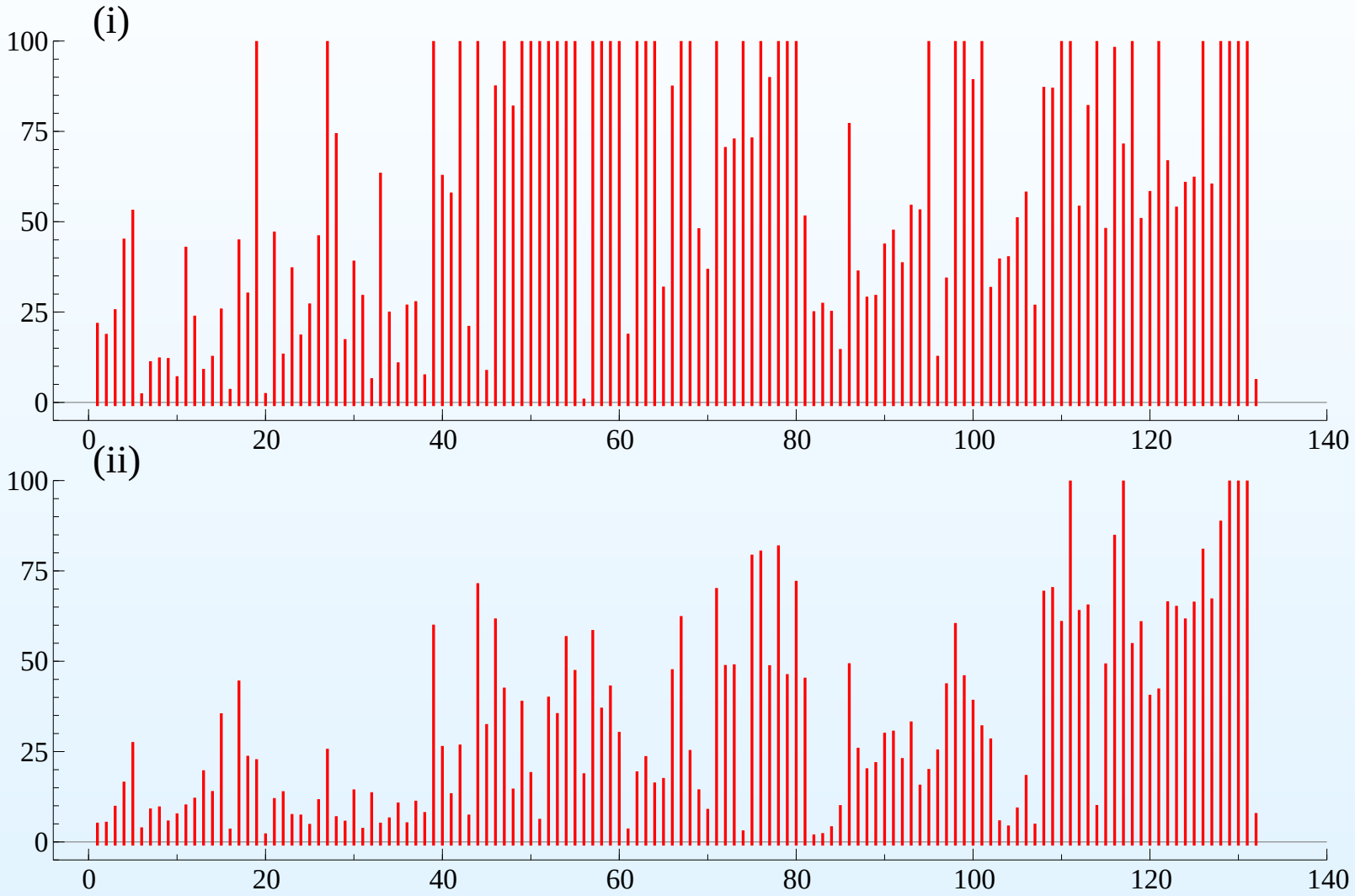


Illustration 3: Smooth modeling of interest rate yields

Empirical models for interest rates and yield curve:

- Nelson and Siegel (1987)
- Diebold and Li (2006)
- Diebold, Rudebusch and Arioba (2006)
- Koopman, Mallee and van der Wel (2008, forthcoming)

We can view Nelson-Siegel model as a restricted DFM.

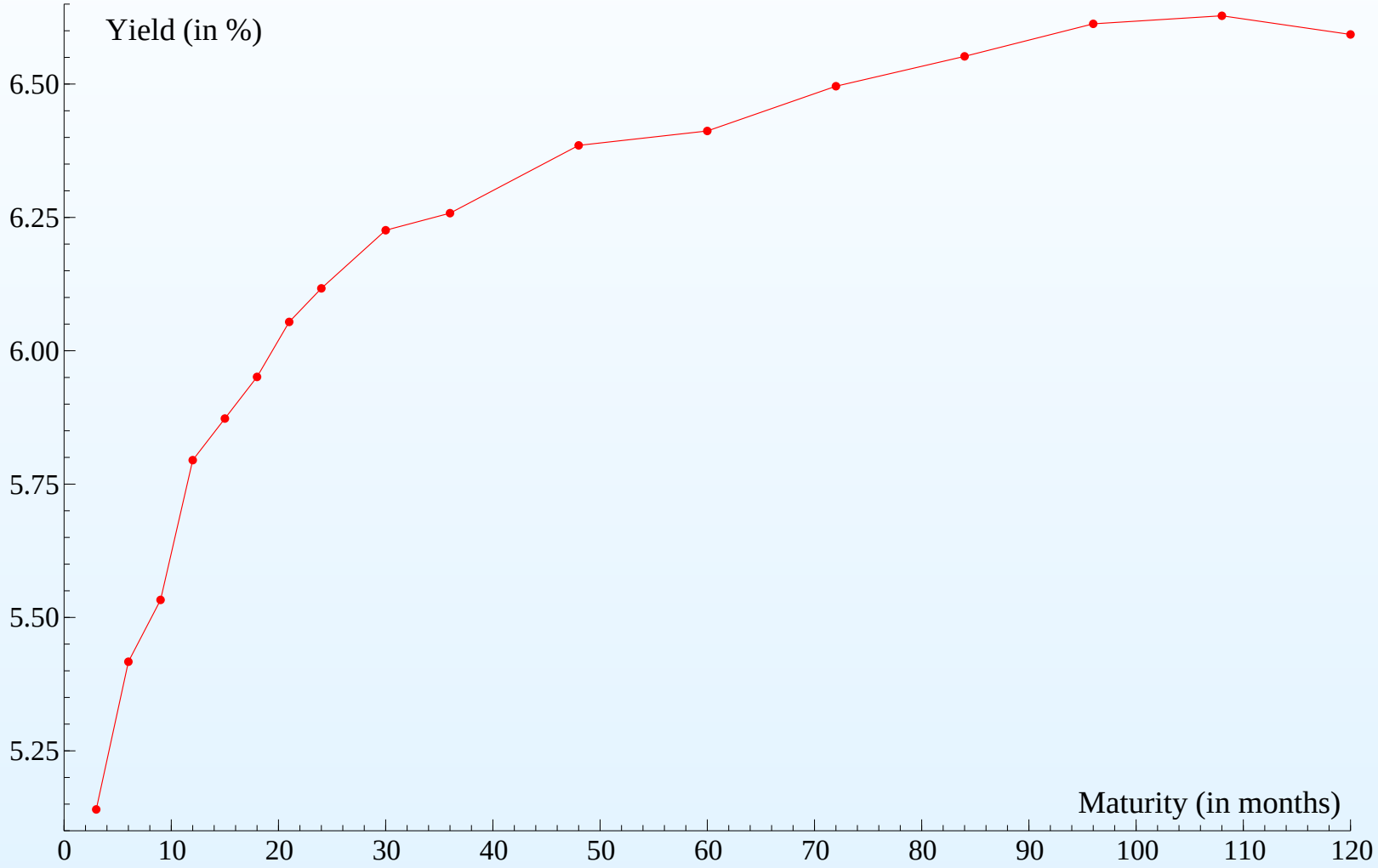
In finance, theoretical models for yield curve:

- Vasicek (1977)
- Cox, Ingersoll and Ross (1985)
- Duffie and Kahn (1996)

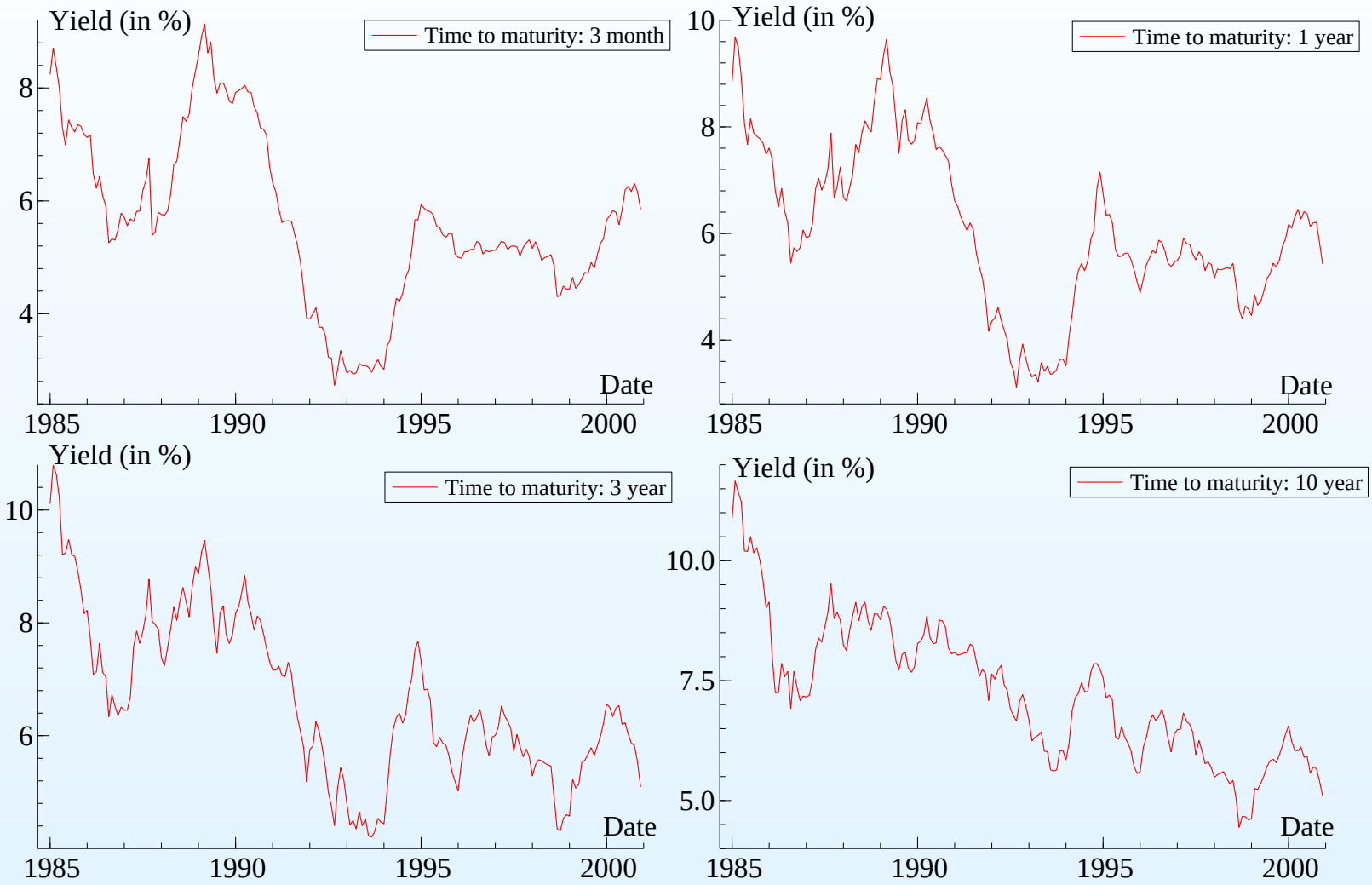
Here models can also be regarded as restricted DFM.

Restrictions concentrate on the amount of smoothness in the factor loadings, see aslo Bowsher and Meeks (2008).
Here we develop "tests" for the appropriate amount of smoothness.

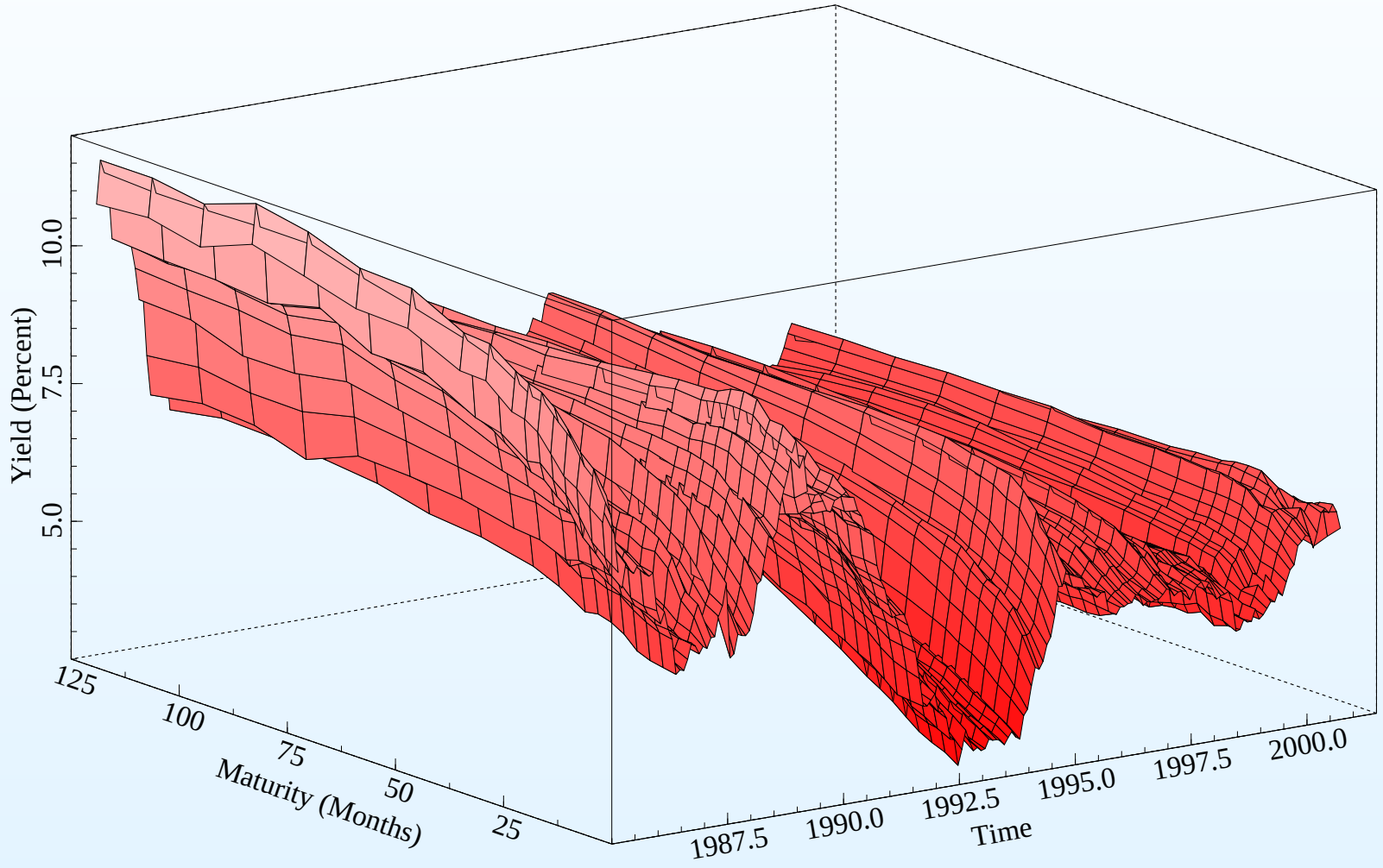
U.S. Yield Curve May 1997



Time Series of Four Maturities



Term Structure of Interest Rates over Time



Dynamic Factor Model for Yields (DFMY)

- Time series panel of N monthly yield observations $y_t = (y_t(\tau_1), \ldots, y_t(\tau_N))'$ with $y_t(\tau_i)$ the yield at time t with maturity τ_i
- The general dynamic factor model is given by:

$$\begin{aligned} y_t &= \mu_y + \Lambda f_t + \varepsilon_t, & \varepsilon_t &\sim N(0, H), \\ f_t &= Z\alpha_t \\ \alpha_{t+1} &= \mu_\alpha + T\alpha_t + R\,\eta_t, & \eta_t &\sim N(0, Q), \end{aligned}$$

  for $t = 1, \ldots, T$. It is a bit more state space.
- Here, vector f_t represents an r-dimensional stochastic process that is generated by the p-dimensional state vector α_t and disturbance η_t
- We take H diagonal and have for the $p \times 1$ initial state $\alpha_1 \sim N(a_1, P_1)$, and assume $N >> r$ and $r \leq p$
- Identification restrictions are imposed on Λ and on μ_y, μ_α

DFMY – Stationary Case

- Take a VAR(k) model for the $r \times 1$ vector f_t:

$$f_{t+1} = \sum_{j=0}^{k-1} \Gamma_{t-j} f_{t-j} + \zeta_t, \qquad \zeta_t \sim NID(0, Q_\zeta)$$

- Stationarity imposed by restriction that $|\Gamma(z)| = 0$ solution has all roots outside the unit circle
- We can write this f_t in DFMY. For example, for $k = 1$ we have

$$\begin{aligned} \alpha_t &= f_t, \\ Z &= I_r, \\ R &= I_r, \\ T &= \Gamma_0, \\ Q &= Q_\zeta \end{aligned}$$

DFMY – Nonstationary Case

- Now the f_t are generated by a cointegrated vector autoregressive process:

$$\Delta f_{t+1} = \beta\gamma' f_t + \sum_{j=0}^{k-1} \Gamma_j \Delta f_{t-j} + \zeta_t, \qquad \zeta_t \sim N(0, Q_\zeta),$$

- The $r \times s$ matrices β and γ have full column rank; in the nonstationary case $s < r$ and all f_t nonstationary
- We propose an alternative but observationally equivalent specification for f_t via factor rotation:

$$\bar{f}_t = \begin{pmatrix} \bar{f}_t^N \\ \bar{f}_t^S \end{pmatrix} = \begin{bmatrix} \beta_\perp & \gamma \end{bmatrix}' f_t,$$

also need to construct new loading matrix $\bar{\Lambda} = \begin{bmatrix} \bar{\Lambda}^N & \bar{\Lambda}^S \end{bmatrix}$

The Smooth DFMY (SDFMY)

- For cross-sectional observation i we can write the DFM as:

$$y_t(\tau_i) = \mu_{y,i} + \sum_{j=1}^{r} \lambda_{ij} f_{jt} + \varepsilon_{it}, \qquad t = 1, \ldots T, \qquad i = 1, \ldots, N,$$

  where λ_{ij} is the loading of factor j on maturity i

- We propose to let the loading parameter be an unknown function $g_j(\cdot)$ for each factor j
- In model, $y_t = \Lambda f_t + \varepsilon_t$, the jth column of Λ is a smooth functional
- Assume functions $g_1(\cdot), \ldots, g_r(\cdot)$ to be smooth functions of time to maturity:

$$\lambda_{ij} = g_j(\tau_i), \qquad \Lambda = [g_1(\cdot), \ldots, g_r(\cdot)],$$

- We consider cubic spline for $g_j(\cdot)$ but any functional form will do
- Here $g_j(\tau)$, more general specification is $g_j(c_j + X_j \beta_j)$

SDFMY with splines: selection of knots

- We consider cubic spline for $g_j(\cdot)$ with K_j knots (smoothness)
- But how many knots K_j to select in the splines?
  - Small number of knots: Loadings lie on same polynomial for considerable number of maturities
  - Large number of knots: Get closer to unrestricted DFM
- Propose using a Wald test procedure to determine the knots
- This is standard as we are testing linear restrictions
- Amounts to an iterative general-to-specific approach:
  1. Start with all knots (unrestricted Λ in DFM)
  2. Calculate test statistic for all knots
  3. Remove knot with smallest non-significant statistic
  4. Continue with 2 and 3 until all knots are significant
- Other (more advanced) selection procedures can also be adopted.

Special DFMY I

- The DFMY is given by:

$$\begin{aligned} y_t &= \mu_y + \Lambda f_t + \varepsilon_t, & \varepsilon_t &\sim N(0, H), \\ f_t &= U\alpha_t \\ \alpha_{t+1} &= \mu_\alpha + T\alpha_t + R\,\eta_t, & \eta_t &\sim N(0, Q), \end{aligned}$$

- It nests all term structure models that we consider
- Functional Signal plus Noise – Bowsher and Meeks (2008)
  - They model yield curve as a spline function:

$$y_t = \mu_y + W f_t + \varepsilon_t, \qquad \varepsilon_t \sim NID(0, H),$$

    with spline weight matrix W as loading matrix and with f_t containing the (time-varying) knot values
  - They consider a CVAR(k) specification for factors, $\mu_\alpha = 0$
  - Model is based on the time-varying spline model for electricity load forecasting, see Harvey and Koopman (1997)

Special DFMY II

- Nelson and Siegel (1987) – Diebold 2006 papers: DL & DRA
  - The yield curve is expressed as a linear combination of smooth factors

$$f_1 + \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right)\cdot f_2 + \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)\cdot f_3$$

    which gives

$$\begin{aligned} y_t &= \Lambda_{ns} f_t + \varepsilon_t, \qquad \varepsilon_t \sim NID(0,H), \\ \Lambda_{ns} &= \left[1 \quad \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) \quad \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)\right]. \end{aligned}$$

  - Interpretation as level, slope and curvature for the factors
  - DRA has restricted VAR(1) for the state, $\mu_y = 0$ and $\mu_\alpha \neq 0$
  - So here we have strong smoothing restrictions on Λ

Special DFMY III

- Arbitrage-Free NS – Christensen, Diebold and Rudebusch (2007)
  - The NS model is not arbitrage free
  - CDR employ “reverse engineering” and obtain an Arbitrage-Free NS model
  - Dynamics of latent factors now coming from solution of SDE, plus ‘correction’ term for μ_y
  - Restrictions on Λ, T and μ_y
- Affine Term Structure Models – Duffie and Kan (1996)
  - Term structure can be explained by dynamics of unobserved short rate
  - Short rate depends on unobserved factors
  - Focus on Gaussian case
  - Restrictions on Λ, T and μ_y

Logikelihood and AIC values

We consider the unrestricted DFMY. We show the value of the loglikelihood at the MLE of $\widehat{\psi}$: $\ell(\widehat{\psi})$ and corresponding *AIC*

Model	$\ell(\widehat{\psi})$	AIC	Model	$\ell(\widehat{\psi})$	AIC
VAR(1)	3894.5	-7595	CVAR(1)	3899.0	-7606
VAR(2)	3918.5	-7625	CVAR(2)	3923.7	-7637
VAR(3)	3922.6	-7615	CVAR(3)	3927.7	-7627
VAR(4)	3932.2	-7616	CVAR(4)	3937.3	-7628

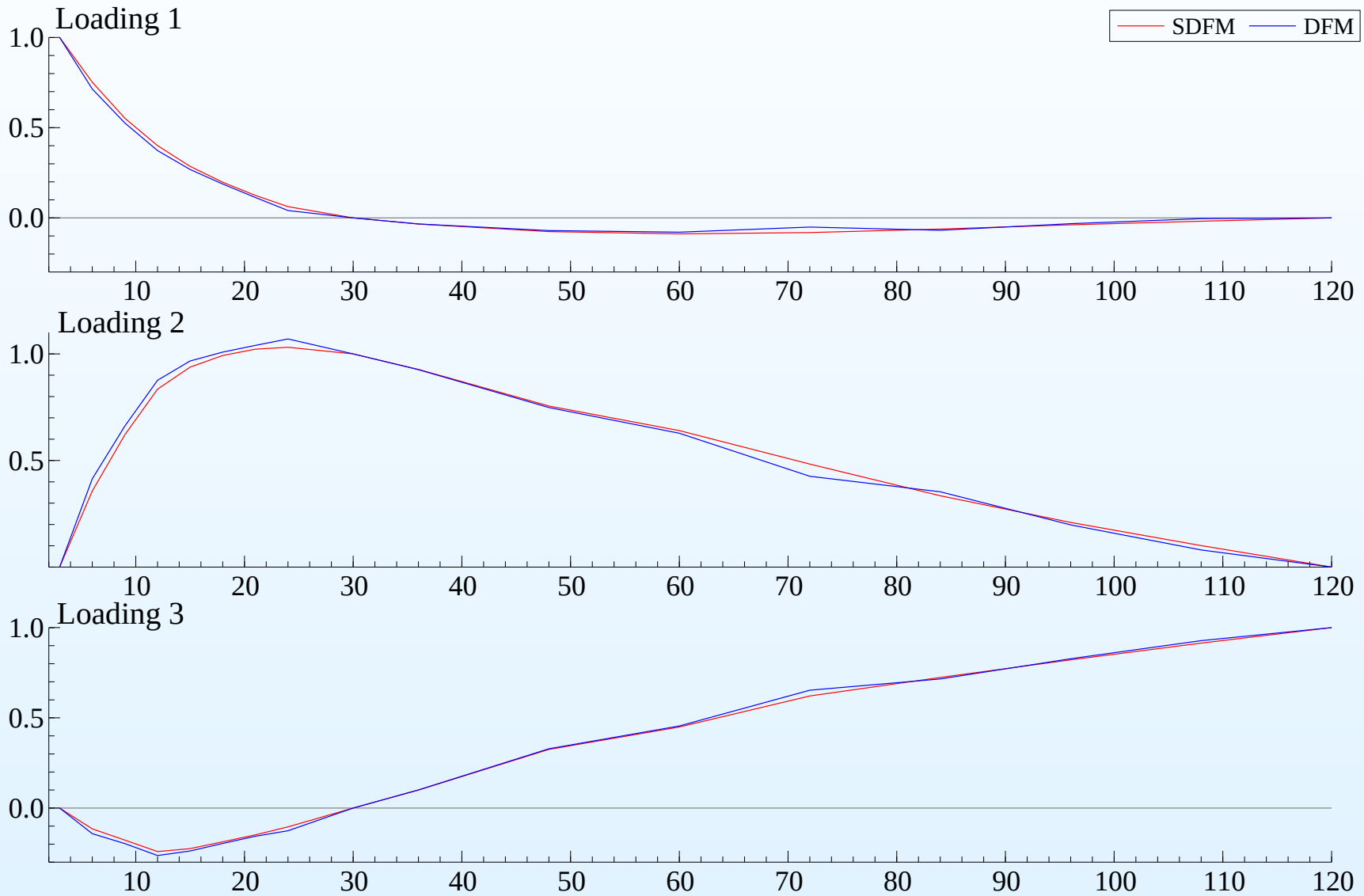
Note: Similar results hold for the NS and FSN model

SDFMY: spline knot selection

Maturity	Panel A: Unrestricted model			Panel B: Final result		
6	2.65	4.22*	6.08*	59.08**	6.50*	5.24*
9	0.79	2.40	5.59*	-	6.58*	8.92**
12	0.23	1.35	4.29*	-	16.25**	19.62**
15	0.04	0.33	1.51	-	24.17**	26.83**
18	0.00	0.02	0.28	-	-	-
21	0.95	0.74	1.52	18.55**	-	-
24	3.51	2.37	3.99*	23.13**	-	7.35**
36	1.14	1.50	6.68**	-	-	26.88**
48	0.44	2.88	13.47**	-	30.07**	52.87**
60	1.19	4.99*	18.04**	-	26.79**	54.39**
72	2.59	5.74*	15.67**	-	22.80**	43.00**
84	2.59	4.57*	8.81**	-	-	17.85**
96	0.76	1.68	1.79	7.68**	-	5.10*
108	0.01	0.05	0.00	-	-	-

Note: 3, 30 and 120 months not shown as these knots can not be removed

SDFM Factor Loadings – CVAR



All Models - Overview

Finally, we provide an overview of all models and test the restrictions imposed on the DFM by the various models

Model	$\ell(\widehat{\psi})$	n_ψ	AIC
DFM-VAR(2)	3918.5	106	-7625
DFM-CVAR(2)	3923.7	105	-7637
SDFM-VAR(2)	3906.8	85	-7644
SDFM-CVAR(2)	3913.6	85	-7657
FSN-VAR(2)	3479.0	64	-6830
FSN-CVAR(2)	3483.7	63	-6841
NS-VAR(2)	3808.4	65	-7487
NS-CVAR(2)	3813.5	64	-7499
AFNS	3253.3	42	-6423
ATSM	3393.0	30	-6726

All Models - Ljung-Box Statistics

	CVAR(2) factors				VAR(1) factors		
Maturity	DFMY	NS	FSN	SDFMY	NS	AFNS	ATSM
3	5.8	6.2	84.3**	6.0	11.6	10.5	17.9
6	7.1	7.4	11.2	7.4	12.6	11.8	33.1**
9	19.2*	19.3*	11.6	18.7*	31.8**	39.9**	55.1**
12	22.5**	29.2**	16.7	23.1**	36.2**	53.1**	52.6**
18	12.8	13.0	13.2	12.7	22.2**	28.1**	22.2**
21	12.2	12.0	13.8	12.2	18.8*	22.4**	19.1*
24	10.2	11.2	15.3	10.6	18.9*	21.6**	22.0**
30	9.3	9.4	10.5	9.1	17.2	15.8	16.0
60	5.9	5.7	9.2	5.6	11.5	10.3	11.2
84	8.4	9.4	10.4	9.1	15.3	12.9	17.4
120	10.1	9.7	6.9	11.0	9.7	9.3	12.2

Note: To preserve space 15, 36, 48, 72, 96 and 108 months omitted here

All Models - Tests of Restrictions

Panel A: Stationary Models			
Model	LR	k	p-value
NS	220.2	41	0.000
FSN	879.0	42	0.000
SDFM	23.4	21	0.320

Panel B: Nonstationary Models			
Model	LR	k	p-value
NS	220.4	41	0.000
FSN	879.8	42	0.000
SDFM	20.2	20	0.450

Panel C: Arbitrage-Free Models			
Model	LR	k	p-value
AFNS	1330.4	64	0.000
ATSM	1051.0	76	0.000

Dynamic factor analysis: a likelihood-based approach with applications in economics and finance

Siem Jan Koopman

Department of Econometrics, VU University Amsterdam

Tinbergen Institute

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

Thank you !