Advances in Factor Modelling

Massimiliano Marcellino
(European University Institute)

18 March 2010, Scuola Normale Superiore

Why factor models?

- Factor models decompose the behaviour of an economic variable into a component driven by unobservable factors common to all the variables, and a variable specific (idiosyncratic component)
- Idea of few common forces driving all economic variables is appealing, e.g. RBC, DSGE literature
- Can handle large datasets -> econometric models based on large information set
- Relatively easy implementation, often good results in practice, in particular for forecasting

Main basic results

- Small factor models can be handled by the Kalman filter
- Large factor models more problematic, in particular with persistent factors and cross-correlated errors (otherwise results in statistical literature)
- The past 10 years have seen major developments: consistent estimation methods for large N, procedures for selection of number of factors, etc. see e.g. review in Bai and Ng (2008)
- Several possible applications in finance (term structure of interest rate, models for exchange rates, etc) and in economics (reduced form models of the economy, forecasting, shock propagation, etc.)

What is new?

- Estimation of large parametric factor models
- Structural factor models
- Factor augmented Error Correction Models
- Mixed frequency data and ragged edges
- Time-varying models
- Factor GMM estimation
- Block structure

What will we see?

- Summary of basic results (Bai, Ng (2008))
- Estimation of large parametric factor models (Kapetanios, Marcellino (2009))
- Time-varying models, Structural FAVAR (Banerjee, Marcellino, Masten (2008), Eickmeier, Lemke, Marcellino (2009))

State space formulation

$$
\begin{aligned}
X_t &= \Lambda f_t + \xi_t, \quad t = 1, \ldots, T \\
f_t &= A f_{t-1} + u_t
\end{aligned}
$$

where:

- X_t is the $N \times 1$ vector of stationary variables
- f_t is $r \times 1$ vector of unobservable factors; Λ, $N \times r$ matrix of loadings
- ξ_t is the $N \times 1$ vector of idiosyncratic shocks
- ξ_t and u_t are multivariate, mutually uncorrelated, standard orthogonal white noise sequences;
- $|\lambda_{\max}(A)| < 1$, $|\lambda_{\min}(A)| > 0$.

If N is small, r small, limited dynamics -> use the Kalman Filter (e.g. Stock and Watson 1989)

The FHLR approach

- The FHLR factor model is

$$X_t = B(L)u_t + \xi_t,$$

where:

- X_t is the $N \times 1$ vector of stationary variables
- u_t is the $q \times 1$ vector of i.i.d. orthonormal common shocks
- $B(L) = I + B_1 L + B_2 L^2 + ... + B_p L^p$ $(B(L) = \Lambda(I - AL)^{-1}$ in previous example)
- ξ_t is the $N \times 1$ vector of idiosyncratic shocks, can be mildly correlated

The FHLR approach - DPCA

- Estimation procedure (q known):
  - Estimate the spectral density matrix of x_t as

$$\begin{aligned} \Sigma^T(\theta_h) &= \sum_{k=-M}^{M} \Gamma_k^T \omega_k e^{-ik\theta_h}, \\ \theta_h &= 2\pi h/(2M+1), \quad h = 0, ..., 2M, \end{aligned}$$

- Calculate the first q eigenvectors of $\Sigma^T(\theta_h)$, $p_j^T(\theta_h)$, $j = 1, ..., q$, for $h = 0, ..., 2M$.

The FHLR approach - DPCA

- Define $p_j^T(L)$ as

$$
\begin{aligned}
p_j^T(L) &= \sum_{k=-M}^{M} p_{j,k}^T L^k, \\
p_{j,k}^T &= \frac{1}{2M+1} \sum_{h=0}^{2M} p_j^T(\theta_h) e^{ik\theta_h}, \;\; k = -M, ..., M.
\end{aligned}
$$

- $p_j^T(L)x_t$, $j = 1, .., q$, are the first q dynamic principal components of x_t (**DPCA**).
- Regress x_t on present, past, and future $p_j^T(L)x_t$. Fitted value is the estimated common component of x_t, $\widehat{\chi}_t$.

The FHLR approach - Choice of q

- Informal methods:
  - Estimate recursively the spectral density matrix of a subset of x_t, increasing the number of variables at each step; calculate the dynamic eigenvalues for a grid of frequencies, λ_θ^x; choose q so that when the number of variables increases the average over frequencies of the first q dynamic eigenvalues diverges, while the average of the $q+1^{th}$ does not.
  - For the whole x_t there should be a big gap between the variance of x_t explained by the first q dynamic principal components and that explained by the $q+1^{th}$ component.
- Formal methods:
  - Information criteria: Hallin Liska (2007, JASA); Amengual and Watson (JBES)

The SW approach - PCA

- The Stock and Watson (1998, 2002a,b) factor model is

$$X_t = \Lambda F_t + \xi_t,$$

  where:

  - F_t, $r \times 1$, common factors, can be correlated over time
  - Λ, $N \times r$, loadings
  - ξ_t, $N \times 1$, idiosyncratic disturbances, can be mildly cross-correlated

- If p in FHLR is finite, $F_t = (u_{1t}, ..., u_{1t-p}, ..., u_{qt}, ..., u_{qt-p})$, and $r = p * q$
- Under mild regularity conditions, the (space spanned by the) factors can be consistently estimated by the first r static principal components of X (**PCA**).
- Choice of r: fraction of explained variance, information criteria, testing (Kapetanios 2009, JBES)

Properties of PCA

- Need both N and T to grow large, and not too much correlation and heteroskedasticity in idiosyncratic errors. Consider

$$x_{it} = \lambda_i f_t + e_{it}. \tag{1}$$

Then

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N} x_{it} &= \overline{x}_t = \left(\frac{1}{N}\sum_{i=1}^{N}\lambda_i\right) f_t + \frac{1}{N}\sum_{i=1}^{N} e_{it} \\ \lim_{N\to\infty} \overline{x}_t &= \overline{\lambda} f_t \end{aligned}$$

and $\overline{x}_t$ is consistent for (the space spanned) by f_t. Can get factor loadings by OLS regression of x_{it} on $\overline{x}_t$, and

$$\lim_{T\to\infty} \widehat{\lambda}_i = \frac{\lambda_i}{\overline{\lambda}}$$

So if both N and T diverge $\widehat{\lambda}_i \overline{x}_t \to \lambda_i f_t$.

Properties of PCA

- PCA are weighted rather than simple averages, where weights depend on λ_i and $var(e_{it})$.
- Under general conditions, PCA and estimated loadings have asymptotic Normal distributions
- If N grows faster than T (such that $T^{1/2}/N$ goes to zero), the estimated factors can be treated as true factors when used in second-step regressions (e.g. for forecasting, factor augmented VARs, etc.). Namely, there are no generated regressor problems.
- The asymptotic distribution of factor based forecasts is also Normal, under general conditions, and its variance depends on the variance of the loadings and on that of the factors, so you need both N and T large to get a precise forecast.

Properties of PCA

- If the factor structure is weak (first factor explains little percentage of overall variance), PCA is no longer consistent (Onatski (2006)).
- If there is an interest in forecasting a specific variable with a large set of regressors, the latter can be pre-selected based on their correlation with the target (Boivin and Ng (2006)). Note that pre-selection also strengthens the factor structure, so less problems of weak factors.
- If the relationship between the target variable and the regressors is non-linear, it could be linearized by a kind of Taylor expansion. In practice, factors could be extracted also from cross-products of regressors, and/or cross-product of factors can be used as regressors (Bai and Ng (2008)).

Properties of PCA

- Statistical procedures to see whether factors correspond to specific macro/finance variables (Bai and Ng (2006)).
- Similar results when (some) factors are I(1). Methods to construct factor based panel unit root and cointegration tests (Bai and Ng (2004, 06))
- PCA can be used for IV/GMM estimation (Bai Ng (2006), Kapetanios Marcellino (2006)).

Parametric estimation - quasi MLE

- Kalman filter produces (quasi-) ML estimators of the factors, but considered not feasible for large N. No longer true: Doz, Giannone, Reichlin (2007).
- Model has the form

$$X_t = \Lambda F_t + \xi_t, \tag{2}$$

$$\Psi(L) F_t = B\eta_t, \tag{3}$$

  where q-dimensional vector η_t contains the orthogonal dynamic shocks driving the r factors F_t, and the matrix B is $(r \times q)$-dimensional.
- For given r and q, estimation proceeds in the following steps:

Parametric estimation - quasi MLE

1. Estimate $\widehat{F}_t$ by PCA and $\widehat{\Lambda}$ by regressing X_t on $\widehat{F}_t$. The covariance of $\widehat{\xi}_t = X_t - \widehat{\Lambda}\widehat{F}_t$, denoted as $\widehat{\Sigma}_{\xi}$, is also estimated.
2. Estimate a VAR(p) on the factors $\widehat{F}_t$, yielding $\widehat{\Psi}(L)$ and the residual covariance of $\widehat{\varsigma}_t = \widehat{\Psi}(L)\widehat{F}_t$, denoted as $\widehat{\Sigma}_{\varsigma}$.
3. To estimate B, given the number of dynamic shocks q, apply an eigenvalue decomposition of $\widehat{\Sigma}_{\varsigma}$. Let M be the $(r \times q)$ matrix of the eigenvectors corresponding to the q largest eigenvalues, and let the $(q \times q)$-dimensional matrix P contain the largest eigenvalues on the main diagonal and zero otherwise. Then, $\widehat{B} = M \times P^{-1/2}$.
4. The Kalman filter or smoother then yield new estimates of the factors, and the procedure can be iterated.

Parametric estimation - SSS

- Model in Kapetanios and Marcellino (2009):

$$\begin{aligned} x_{nt} &= Cf_t + Du_t, \quad t = 1, \ldots, T \\ f_t &= Af_{t-1} + Bu_{t-1} \end{aligned} \tag{4}$$

(4) can be written as

$$X_t^f = \mathcal{O}\mathcal{K}X_t^p + \mathcal{E}E_t^f \tag{5}$$

where $X_t^f = (x'_{nt}, x'_{nt+1}, x'_{nt+2}, \ldots)'$, $X_t^p = (x'_{nt-1}, x'_{nt-2}, \ldots)'$, $E_t^f = (u'_t, u'_{t+1}, \ldots)'$, $\mathcal{O} = [C', A'C', (A^2)'C', \ldots]'$, $\mathcal{K} = [\bar{B}, (A - \bar{B}C)\bar{B}, (A - \bar{B}C)^2\bar{B}, \ldots]$, $\bar{B} = BD^{-1}$.

- Note that (i) $X_t^f = \mathcal{O}f_t + \mathcal{E}E_t^f$ and (ii) $f_t = \mathcal{K}X_t^p$. Best linear predictor of future X is $\mathcal{O}\mathcal{K}X_t^p$. The state is $\mathcal{K}X_t^p$ at time t. We want an estimator for $\mathcal{K}$.

Parametric estimation - SSS

- The model $X_t^f = \mathcal{O}\mathcal{K}X_t^p + \mathcal{E}E_t^f$ involves infinite dimensional vectors. In practice, use $X_{s,t}^f = (x_{nt}', x_{nt+1}', x_{nt+2}', \ldots, x_{nt+s-1}')'$ and $X_{p,t}^p = (x_{nt-1}', x_{nt-2}', \ldots, x_{nt-p}')'$. Then, regress $X_{s,t}^f$ on $X_{p,t}^p$, and apply a singular value decomposition to $\hat{\Gamma}^f\hat{\mathcal{F}}\hat{\Gamma}^p$, where $\mathcal{F} = \mathcal{O}\mathcal{K}$ and $\hat{\Gamma}^f$, and $\hat{\Gamma}^p$ are the sample covariances of $X_{s,t}^f$ and $X_{p,t}^p$ respectively. These weights are used to determine the importance of certain directions in $\hat{\mathcal{F}}$. Then, the estimate of $\mathcal{K}$ is given by

$$\hat{\mathcal{K}} = \hat{S}_m^{1/2}\hat{V}_m\hat{\Gamma}^{p^{-1}}$$

  where $\hat{U}\hat{S}\hat{V}$ represents the singular value decomposition of $\hat{\Gamma}^f\hat{\mathcal{F}}\hat{\Gamma}^p$, $\hat{S}$ contains the singular values of $\hat{\Gamma}^f\hat{\mathcal{F}}\hat{\Gamma}^p$ in decreasing order, $\hat{S}_m$ denotes the matrix containing the first m columns of $\hat{S}$ and $\hat{V}_m$ denotes the heading $m \times m$ submatrix of $\hat{V}$.
- The SSS factor estimates are $\hat{\mathcal{K}}X_t^p$.

Parametric estimation - SSS

- p must increase at a rate greater than $ln(T)^{\alpha}$, for some $\alpha > 1$, but Np at a rate lower than $T^{1/3}$. N is fixed for the moment. A range of α between 1.05 and 1.5 provides a satisfactory performance.
- s is required to satisfy $sN > m$. As N is large this restriction is not binding, $s = 1$ is enough.
- Having estimated all the parameters of the model, we can compute smoothed estimates of the factors. Yet, the starting point is the SSS estimate of the factors, and the smoother does not improve.

Parametric estimation - SSS, T asymptotics

- If we define $\hat{f}_t = \hat{\mathcal{K}} X_t^p$, then $\hat{f}_t$ converges to (the space spanned by) f_t. The speed of convergence is between $T^{1/2}$ and $T^{1/3}$ because p grows. Note that consistency is possible because f_t depends on v_{t-1}. If f_t depends on v_t, $\hat{f}_t$ converges to Af_{t-1}.
- The asymptotic distribution of $\sqrt{T^*}(vec(\hat{f}) - vec(H_m f))$ with $f = (f_1, \ldots, f_T)'$.is $N(0, V_f)$.
- Once estimates of the factors are available, estimates of the other parameters (including the factor loadings) can be obtained by OLS. Bauer (1998) proves $\sqrt{T}$ consistency and asymptotic normality.

Parametric estimation - SSS, T and N asymptotics

- If Np is $o(T^{1/3})$, p is $O(T^{1/r})$, $r > 3$, then when N and T diverge $\hat{f}_t = \hat{\mathcal{K}} X_t^p$ converges to (the space spanned by) f_t. The speed of convergence is $(T/Np)^{1/2}$. The intuition is that the estimator of $\mathcal{F} = \mathcal{O}\mathcal{K}$ in $X_{s,t}^f = \mathcal{F} X_{p,t}^p + \mathcal{E} E_t^f$ remains consistent if $Np = o(T^{1/3})$.
- With a proper standardization, $\hat{f}_t$ remains asymptotically normal
- We provide information criteria for consistent estimation of number of factors, similar to Bai and Ng (2002(but with different penalty function.

Factor estimation methods - MC Comparison

- Comparison of PCA, DPCA, MLE and SSS. We use the DGP

$$\begin{aligned} x_t &= Cf_t + \epsilon_t, \quad t = 1, \ldots, T \\ A(L)f_t &= B(L)u_t \end{aligned} \tag{6}$$

  where $A(L) = I - A_1(L) - \ldots - A_p(L)$, $B(L) = I + B_1(L) + \ldots + B_q(L)$, with $(N, T) =$ (50,50), (50,100), (100,50), (100,100), (50,500), (100,500) and (200,50). MLE for (50,50) only, due to computational burden.
- Experiments differ for the number of factors (one or several), the A and B matrices, the choice of s ($s = m$ or $s = 1$), the factor loadings (static or dynamic), the choice of the number of factors (true number or misspecified), the properties of the idiosyncratic errors (uncorrelated or serially correlated), and the way the C matrix is generated (standard normal or uniform with non-zero mean). Five groups of experiments, each is replicated 500 times.

Factor estimation methods - MC Comparison

- First set of experiments: a single VARMA factor with different specifications:

  1. $a_1 = 0.2, b_1 = 0.4$;
  2. $a_1 = 0.7, b_1 = 0.2$;
  3. $a_1 = 0.3, a_2 = 0.1, b_1 = 0.15, b_2 = 0.15$;
  4. $a_1 = 0.5, a_2 = 0.3, b_1 = 0.2, b_2 = 0.2$;
  5. $a_1 = 0.2, b_1 = -0.4$;
  6. $a_1 = 0.7, b_1 = -0.2$;
  7. $a_1 = 0.3, a_2 = 0.1, b_1 = -0.15, b_2 = -0.15$;
  8. $a_1 = 0.5, a_2 = 0.3, b_1 = -0.2, b_2 = -0.2$.
  9. As 1 but $C = C_0 + C_1 L$.
  10. As 1 but one factor assumed instead of $p + q$

Factor estimation methods - MC Comparison

- Second group of experiments: as in 1-10 but with each idiosyncratic error being an AR(1) process with coefficient 0.2 (exp. 11-20). Experiments with cross correlation yield similar ranking of methods.
- Third group of experiments: 3 dimensional VAR(1) for the factors with diagonal matrix with elements equal to 0.5 (exp. 21).
- Fourth group of experiments: as 1-21 but the C matrix is U(0,1) rather than N(0,1).
- Fifth group of experiments: as 1-21 but using $s = 1$ instead of $s = m$.

Factor estimation methods - MC Comparison

- We compute the correlation between true and estimated common component. We have also computed the spectral coherency for selected frequencies. We also report the rejection probabilities of an LM(4) test for no correlation in the idiosyncratic component. The values are averages over all series and over all replications.
- Detailed results are in paper: for exp. 1-21, groups 1-3, see Tables 1-7; for exp. 1-21, group 4, see Table 8 for (N=50, T=50); for exp. 1-21, group 5, see Tables 9-11.

Factor estimation methods - MC Comparison, N=T=50

- Single ARMA factor (exp. 1-8): looking at correlations, SSS clearly outperforms PCA and DPCA. Gains wrt PCA rather limited, 5-10%, but systematic. Larger gains wrt DPCA, about 20%. Little evidence of correlation of idiosyncratic component , but rejection probabilities of LM(4) test systematically larger for DPCA.
- Serially correlated idiosyncratic errors (exp. 11-18): no major changes. Low rejection rate of LM(4) test due to low power for $T = 50$.
- Dynamic effect of factor (exp. 9 and 19): serious deterioration of SSS, a drop of about 25% in the correlation values. DPCA improves but it is still beaten by PCA.. Choice of s matters: for $s = 1$ SSS becomes comparable with PCA (Table 9).

Factor estimation methods - MC Comparison, N=T=50

- Misspecified number of factors (exp. 10 and 20): no major changes, actually slight increase in correlation. Due to reduced estimation uncertainty.
- Three autoregressive factors: (exp. 21): gap PCA-DPCA shrinks, higher correlation values than for one single factor. SSS deteriorates substantially, but improves and becomes comparable to PCA when $s = 1$ (Table 11).
- Full MLE gives very similar and only very slightly better results than PCA, and is dominated clearly by SSS.

Factor estimation methods - MC Comparison, other results

- *Larger temporal dimension (N=50, T=100,500).* Correlation between true and estimated common component increases monotonically for all the three methods, ranking of methods across experiments not affected. Performance of LM tests for serial correlation gets closer and closer to the theoretical one. (Tab 2,3)
- *Larger cross-sectional dimension (N=100, 200, T=50).* SSS is not affected (important, $N > T$), PCA and DPCA improve systematically, but SSS still yields the highest correlation in all cases, except exp. 9, 19, 21. (Tab 4,7).
- *Larger temporal and cross-sectional dimension (N=100,T=100 or N=100,T=500).* The performance of all methods improves, more so for PCA and DPCA that benefit more for the larger value of N. SSS is in general the best in terms of correlation(Tab 5,6).
- *Uniform loading matrix.* No major changes (Tab 8)
- *Choice of s.* PCA and SSS perform very similarly (Tab 9-11).

Factor estimation methods - MC Comparison, summary

- DPCA shows consistently lower correlation between true and estimated common components than SSS and PCA. It shows, in general, more evidence of serial correlation of idiosyncratic components, although not to any significant extent.
- SSS beats PCA, but gains are rather small, in the range 5-10%, and require a careful choice of s.
- SSS beats MLE, which is only sligthly better than PCA.
- All methods perform very well in recovering the common components.

Factor forecasts under structural breaks

Based on work with Banerjee and Masten

Motivation

- Factor models useful for forecasting macro variables, large literature showing this. Alternative methods available, several comparative studies, e.g. Eickmeier Ziegler (2007).
- Fewer results for the large N small T case, even though factor models should be efficient also in this case, and even with $N>T$. Interesting applications include New EU member countries and the Euro area.
- Additional problem for the New EU member countries and the Euro area is parameter instability

What we will see

- Forecasting performance of diffusion index-based methods in short samples with structural change, based on:

  - Detailed simulation study: DGPs with different types of structural change, relative forecasting performance of factor models and traditional time series methods.

  - In the paper also empirical applications for the Euro area and Slovenia: relatively short samples of data and structural changes are likely.

Main findings

- Coherence b/w the empirical and simulation results.
- Relatively good performance of factor-based forecasts in short samples with structural change.

Monte Carlo Experiments

- **Purpose**: Understand the sensitivity of the performance of factor- and non-factor methods to:

  - *T* and *N*
  - various features likely to characterize the data in practice: degree of persistence of the factors and the presence of structural change
  - The data are generated by a dynamic factor model that allows for autoregressive factors, auto- and cross-correlation in idiosyncratic errors and time-varying parameters

DGP

- $x_{it} = \lambda_{it}^{'} f_t + e_{it}$
- $f_t = A_t f_{t-1} + u_t, \qquad A_t = \alpha_t I_r$
- $\alpha_t = d\left(\alpha_{t-1} + 1/T\, \eta_t\right) + \left(1-d\right)\alpha_1 I\left(T_B\right) + (1-d)\alpha_0, \;\; \alpha_0 = \overline{\alpha}$
- $I\left(T_B\right) = \begin{cases} 0, \forall t \leq T_B \\ 1, \forall t > T_B \end{cases} \quad d = \begin{cases} 0, \text{breaking} & \alpha \\ 1, \text{time varying} & \alpha \end{cases}$
- $\lambda_{it} = \lambda_{it-1} + (c/T)\zeta_{it}$
- $(1-aL)e_{it} = \left(1+b^2\right)v_{it} + bv_{i+1,t} + bv_{i-1,t}$
- $y_t = \iota' f_{t-1} + \varepsilon_t$

- f_t and λ_t are $r \times 1$, r=1, …, 5
- e_{it}, v_{it}, and ε_t are i.i.d. $N(0,1)$, while ζ_{it} and u_t are i.i.d. $N(0,I_r)$. u_t is independent of e_{it}, v_{it}, ε_t and ζ_{it}
- Factor persistence α:

o stable and fixed (d = *0* and $T_B = T$, $\alpha = \{0.3, 0.5, 0.7\}$)

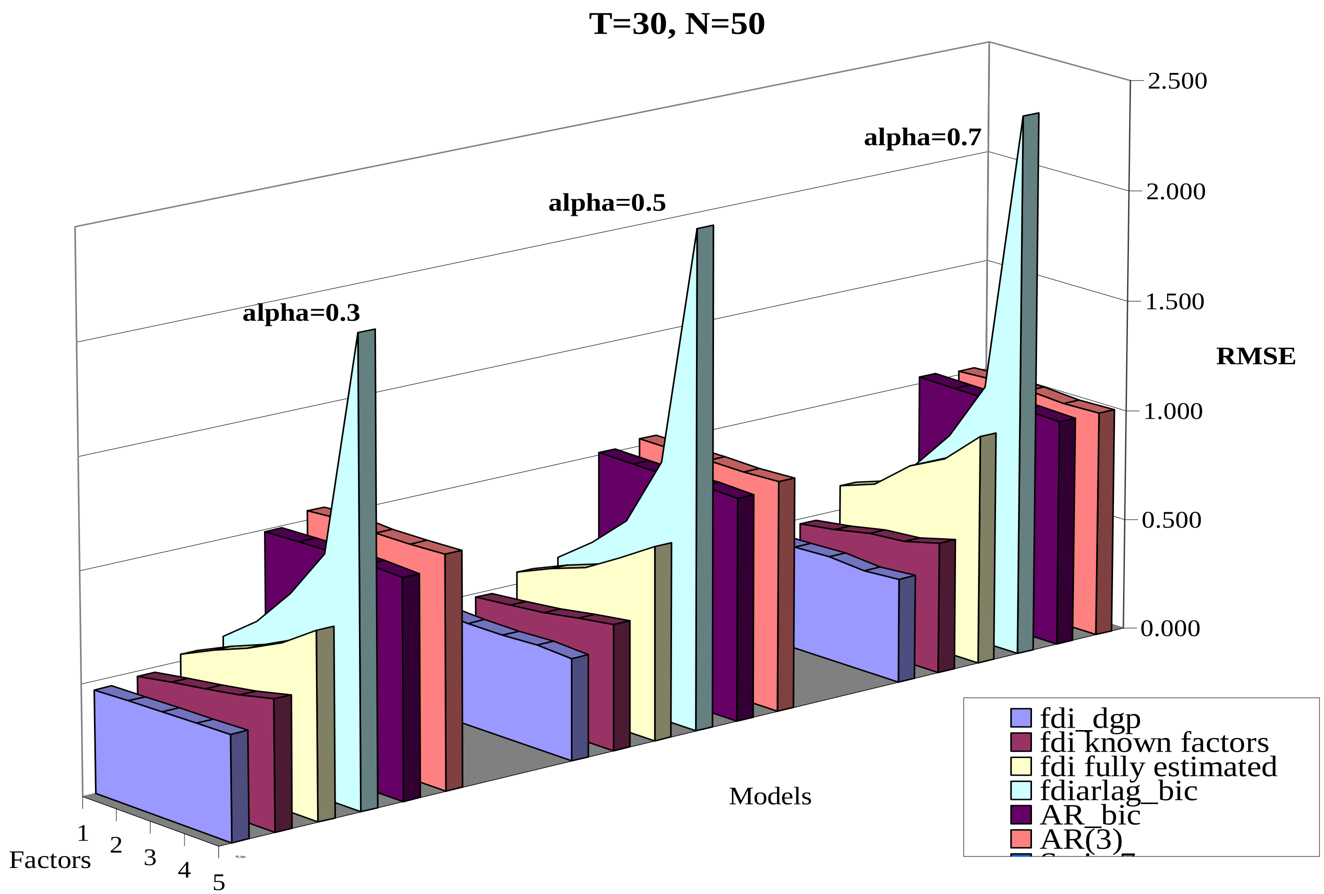
o continuously time-varying persistence (d = *1*, $\overline{\alpha} = \{0.3, 0.5, 0.7\}$)

o discrete break in persistence of factors (d = *0* and $T_B = T/2$, $\alpha_1 = 0.4$ when $\alpha_0 = 0.3$ and $\alpha_1 = -0.4$ when $\alpha_0 = 0.7$ -> persistence from 0.3 to 0.7 and viceversa)

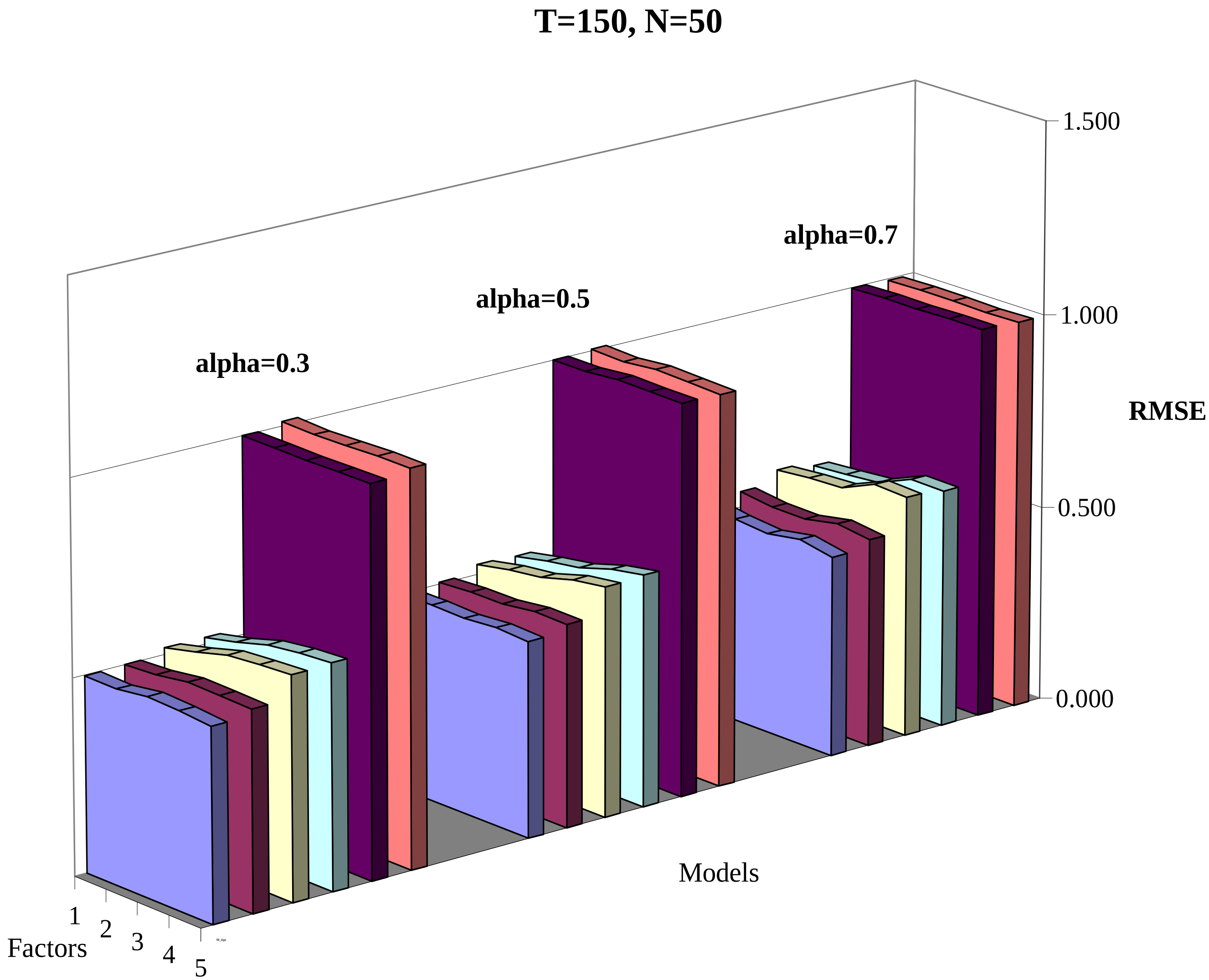
- Case of double variance of factors
- Time-varying factor loadings ($c = 5$)
- Cross-correlated idiosyncratic components ($a = 0.5,\ b = 1$)
- *T* and *N* combinations:
  - T = 30, N = 50 relevant for new EU members on quarterly frequency
  - T = 50, N = 50
  - T = 50, N = 100
  - T = 150, N = 50

- Models estimated on artificial data:
  - AR(1)-benchmark, AR(3) and AR with BIC selection
  - Factor models: (1) *fdi_dgp* (not estimated), (2) known factors, estimated coefficients, (3) fully estimated, but knowing the model's structure, and (4) lags of factors and *y* (*fdiarlag_bic*).

Figure 3: Time-varying lambda - *h* = 1



MSE relative to AR(1)


T=150, N=50
alpha=0.3
alpha=0.5
alpha=0.7
1.500
1.000
0.500
0.000
RMSE
Models
Factors
1
2
3
4
5

Summary of Monte Carlo results

(a) Continuous changes in factor persistence do not seem to matter, even in short samples.

(b) Discrete changes do matter but the impact on relative performance of factor methods leads either to improvement or deterioration, depending upon the value of the (starting) persistence parameter, the direction of the change and the magnitude of the T and N dimensions.

(c) Time varying factor loadings are important except when T and N are large (in line with SW 2002) – empirically relevant (Figure 3)

(d) Ranking of the impact of the different kinds of stability is (c) to (b) to (a)

(e) Factor models outperform AR models in the majority of cases, even in short samples subject to changes

(f) Relative performance of factor models deteriorates fast with number of factors in the DGP (especially when selection is BIC based)

(g) As expected, the variance of the idiosyncratic component of the target variable is important. (Figure 4)

(h) Similar results for $h=4$

Which models could handle instability in short samples?

- ARTV (tend to overperform STAR, NN and MS)

- Stock and Watson’s P/T model with Stochastic Volatility

- Koopman and Marcellino, generalizations of ARTV

- *Factor models with TV-parameters*

- BVAR (De Mol, Giannone and Reichlin (2006)) and BRRR (Carriero, Kapetanios and Marcellino (2007))

Classical time-varying FAVAR models
Estimation, forecasting and structural analysis

Based on work with Eickmeier and Lemke

Scope

- Develop a classical, Kalman filter-based approach to estimate FAVARs with time-varying parameters

- Allow for time variation in the loadings, factor dynamics and conditional variance-covariance matrix

- Application to large US dataset, 1972-2007
  - forecasting
  - monetary transmission

Related literature

- Constant parameter FAVAR (Bernanke et al. 2005) exploiting large set of variables and has proved useful for
  - forecasting (e.g. Stock and Watson 2002a, 2002b, 2006, Eickmeier and Ziegler 2008)
  - structural analyses (e.g. Boivin et al. 2009, Kose et al. 2003, 2008, Beck et al. forthcoming, Eickmeier and Hofmann 2009)
- VARs allowing for smooth time variation in parameters
  - e.g. Cogley and Sargent 2005, Sims and Zha 2006
- Time-varying parameter (TV-) Bayesian FAVARs
  - monetary policy (MP) applications (Baumeister et al. 2009, Korobilis 2009)
  - internat. business cycle and inflation comovements (Del Negro and Otrok 2008, Liu and Mumtaz 2009, Mumtaz and Surico 2008)

Summary of results

- Method
  - Fast and numerically stable
  - Results not prior-dependent
  - Flexible accounting for various sources of time variation

- Estimation results
  - Minor changes in factor dynamics
  - Discernible variation in volatility
  - Marked changes in factor loadings for some variables

Summary of results (cont.)

- Forecasting
  - Overall good performance, also after 1995
  - TV-FAVAR forecasts more accurate than constant-parameter counterparts
  - Forecast superiority especially for inflation and financial variables

- Structural analysis
  - Overall plausible IRFs to MP shocks
  - Smaller size of shocks over time, in particular after 1985
  - Weaker price and output responses over time to MP shocks
  - Stronger reaction of consumption and investment during recessions

Outline

1. Model and estimation approach

2. Data and evidence on time-variation in the parameters

3. Forecasting

4. Structural analysis

5. Conclusion

1. Model and estimation approach

Loadings

- Observables as functions of factors and idiosyncratic components
$$x_{i,t} = \Lambda'_{i,t} F_t + e_{i,t}, \quad i = 1, \ldots, N$$
- Assume regularity conditions as in Stock/Watson (1998, 2002)
- F = latent factors (later one observable, FFR)
- Independent random walks for time-varying loadings: smoothness conditions

Factor dynamics

- Dynamics of factors, triangular P_t

$$P_t F_t = \mathcal{K}_{1,t} F_{t-1} + \ldots + \mathcal{K}_{p,t} F_{t-p} + u_t,$$

$$E(u_t u_t') = S_t$$

- Time variation in contemporaneous relations and autoregressive dynamics

- Again, independent random walks for parameters

Shock volatility

- Time-varying volatility of FAVAR innovations.

- Additional latent volatility factor? → Nonlinear state space model.

- Here: specify time-varying vola as exponential-affine function of past factors.

$$S_{gg,t} = \exp(c_g + b_g' F_{t-1}), \quad g = 1, \ldots, G$$

- Estimate *c* and *b* parameters by ML

- Alternatives: tv vola as function of exogenous variables, other functional forms

Estimation of factors

- Estimate G factors by principal components (PC) as justified by Stock and Watson (1998, 2002, 2008)

- Treat them as known (Stock and Watson 2002, 2008, Bai and Ng 2002)

Estimation of tv parameters

- Time-varying loadings
  - estimate (by ML) random-walk innovation variances of time-varying parameters, equation by equation
  - Kalman smoother to back out parameter paths

- Time-varying VAR parameters
  - also estimate VAR equations individually (exploit conditional independence!)
  - get ML estimates of random-walk variances of time-varying K and P parameters and of tv vola parameters
  - Kalman smoother to back out parameter paths, impose non-explosiveness condition for VAR
  - VAR order for F_t set to 2

Comparison with previous (Bayesian) TV-FAVARs

- Our classical approach does not rely on simulation-based inference → fast and computationally less burdensome

- Outcome not dependent on prior distribution.

- Relatively flexible due to two-step approach
  - allow for time-variation in both loadings and autoregressive VAR parameters (no identification problem)
  - shock volatility modelled as a function of the factors (in contrast to stochastic volatility - with additional latent factors)
  - model straightforward to extend to allow for time-varying volatility and serial correlation in idiosycratic components

2. Data and evidence on time-variation in the parameters

US data

- Sample period 1972Q1-2007Q2

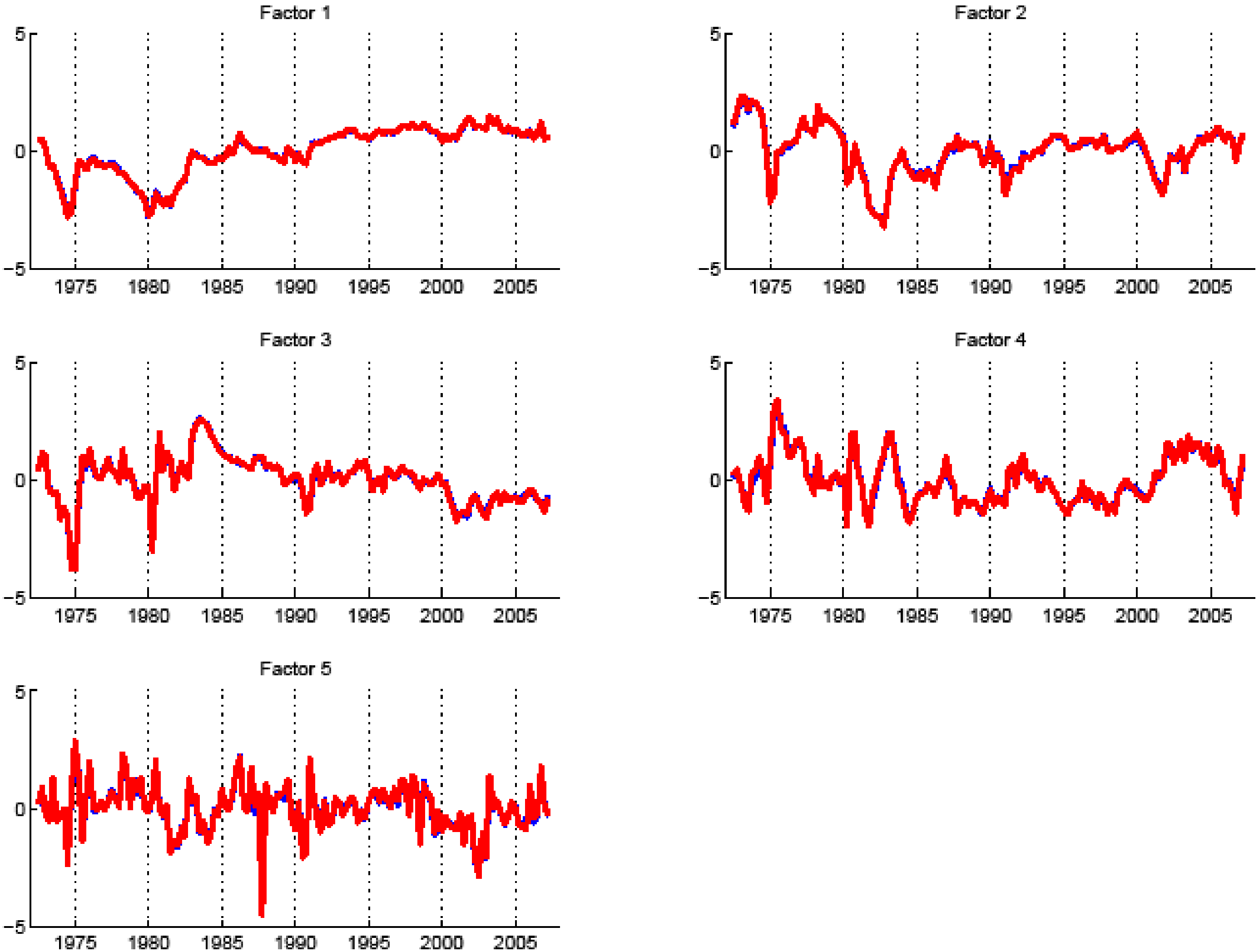
- Original (balanced) dataset with 808 quarterly series

- Remove series with a commonality < 0.6 (based on 5 factors)
  $\longrightarrow$ left with $N = 338$ series
  - 119 real economic activity measures (e.g. GDP and components, industrial production, labor market variables, expectations)
  - 136 price measures (e.g. deflators of GDP and components, CPI, PPI, wages, commodity prices)
  - 83 monetary and financial indicators (e.g. interest rates, stock prices, money and credit aggregates, exchange rates)

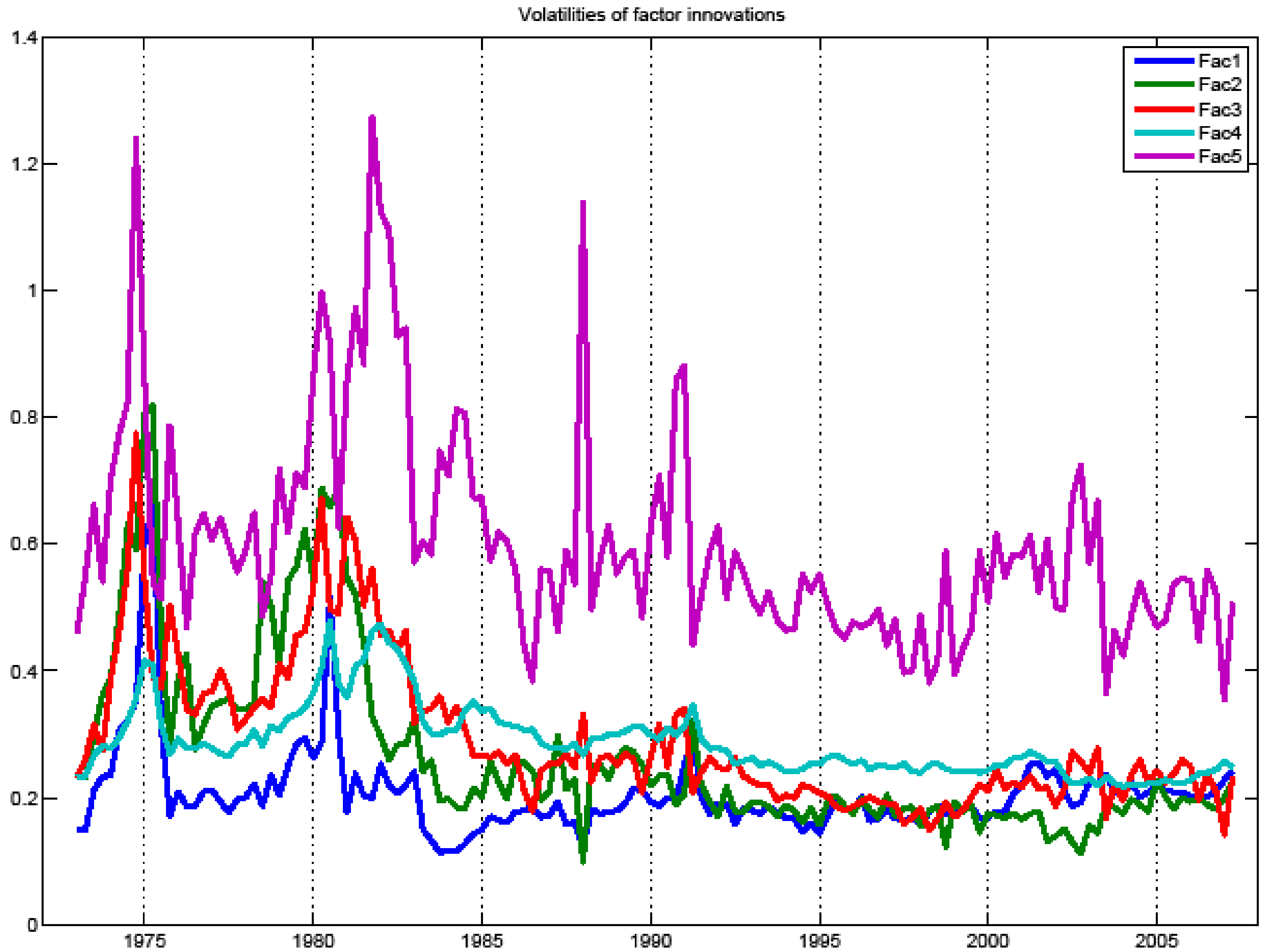
- Data are seasonally adjusted, stationary, outlier adjusted, demeaned and standardized

Estimated factors

- First 5 PCs (blue) and factors re-estimated from a cross-section regression on time-varying loadings (red)

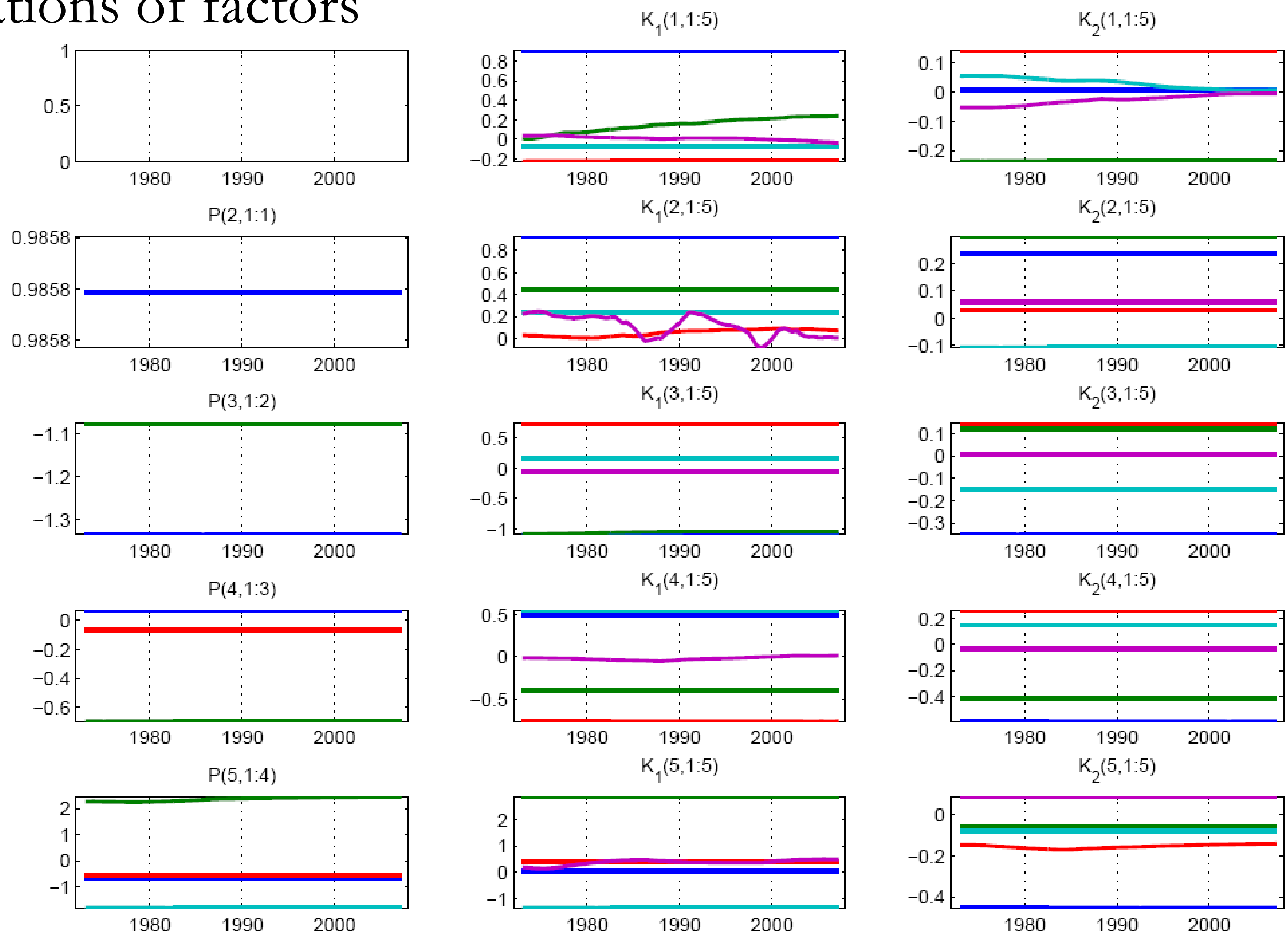


Time-variation of factor innovation volatility



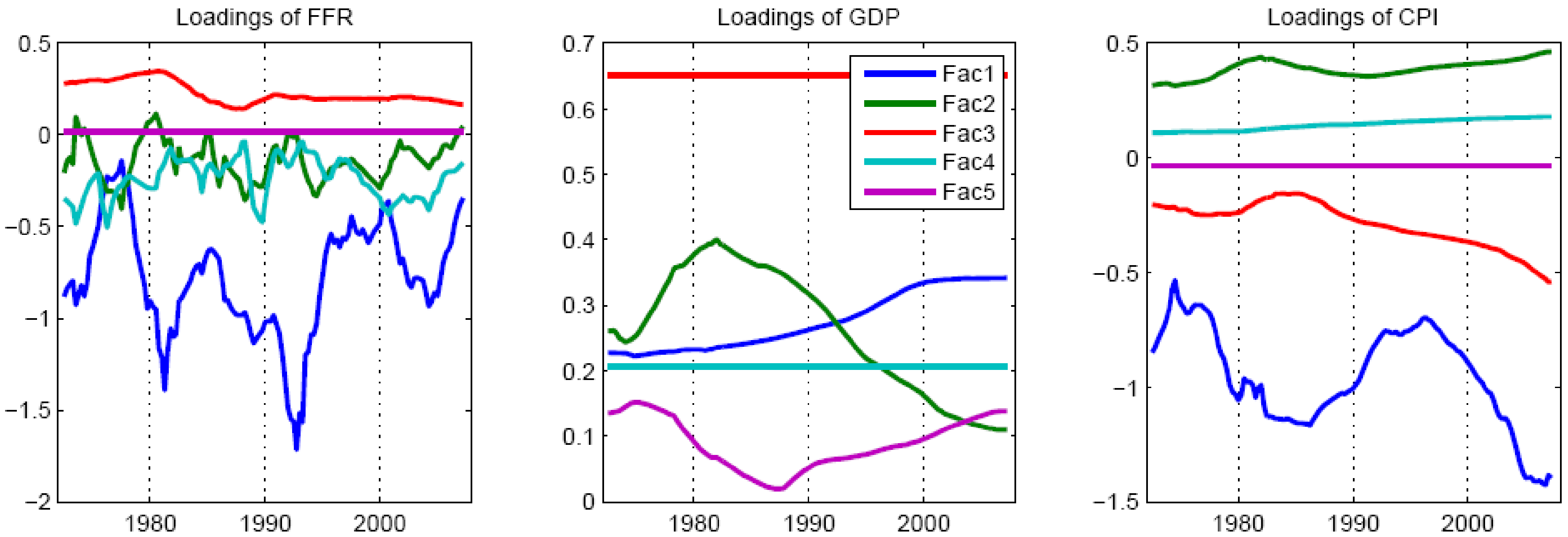
Time-variation of VAR parameters

- Dynamic (*K* parameters) and contemporaneous (*P* parameters) relations of factors



Time-variation of factor loadings

- Loadings of FFR, GDP growth, CPI inflation



Summary: evidence on parameter variation

- PC estimator appropriate (will extend cross-sectional regression to iterative procedure and exploit full system later)

- Substantial time variation in shock variances and loadings

- Limited time variation in dynamic and contemporaneous relations of factors

3. Forecasting

Forecasting with the TV-FAVAR

- Goal: predict macro and financial variables

- In-sample (for now) forecasts for
  - entire sample
  - recessions only
  - post-1995 sample (e.g. D'Agostino et al. 2006)

- Forecast procedure
  - add target variable one by one to first 5 PCs
  - estimate a TV-VAR on the 6 variables/factors
  - carry out iterative forecasts for horizons 1 to 4

Forecast results

- (Relative) RMSEs of forecasts of activity

	RMSE(AR)			Const. FAVAR/AR			TV-FAVAR/const. FAVAR		
h	all periods	recessions	after 1995	all periods	recessions	after 1995	all periods	recessions	after 1995
ΔGDP									
1	0.0077	0.0113	0.0049	0.798	0.620	1.020	1.000	1.274	0.964
2	0.0078	0.0121	0.0048	0.847	0.665	0.972	1.004	1.228	1.039
3	0.0080	0.0131	0.0050	0.917	0.810	1.058	0.986	1.070	0.939
4	0.0080	0.0133	0.0050	0.957	0.847	1.094	0.969	1.027	0.918
ΔConsumption									
1	0.0062	0.0106	0.0038	0.815	0.659	0.957	0.993	1.088	0.899
2	0.0063	0.0110	0.0038	0.916	0.834	1.152	0.960	1.037	0.852
3	0.0063	0.0114	0.0039	0.948	0.852	1.110	0.962	1.009	0.866
4	0.0064	0.0115	0.0040	0.945	0.849	1.124	0.953	1.005	0.829
ΔNon-residential investment									
1	0.0191	0.0299	0.0122	0.853	0.743	0.977	1.006	1.158	0.952
2	0.0213	0.0347	0.0138	0.841	0.691	1.014	0.982	1.022	0.926
3	0.0225	0.0381	0.0142	0.879	0.768	1.090	0.947	0.990	0.872
4	0.0228	0.0392	0.0149	0.902	0.793	1.116	0.928	0.974	0.849
ΔResidential investment									
1	0.0375	0.0613	0.0170	0.747	0.606	1.053	0.951	1.103	0.707
2	0.0438	0.0694	0.0214	0.772	0.779	1.049	0.957	1.024	0.728
3	0.0452	0.0748	0.0229	0.873	0.829	1.128	0.940	1.002	0.719
4	0.0458	0.0768	0.0237	0.873	0.805	1.113	0.952	1.023	0.748

Forecast results

- (Relative) RMSEs of forecasts of inflation measures

	RMSE(AR)			Const. FAVAR/AR			TV-FAVAR/const. FAVAR		
h	all periods	recessions	after 1995	all periods	recessions	after 1995	all periods	recessions	after 1995
ΔGDP deflator									
1	0.0026	0.0032	0.0018	0.829	0.550	1.042	0.921	1.021	0.807
2	0.0032	0.0045	0.0018	0.820	0.606	1.155	0.955	1.023	0.815
3	0.0036	0.0054	0.0019	0.812	0.667	1.176	0.919	0.853	0.782
4	0.0038	0.0062	0.0021	0.814	0.712	1.323	0.892	0.781	0.710
ΔCPI									
1	0.0048	0.0067	0.0046	0.787	0.649	0.879	0.954	1.201	0.732
2	0.0054	0.0074	0.0045	0.803	0.760	0.906	1.006	1.170	0.797
3	0.0054	0.0079	0.0040	0.843	0.747	1.043	0.974	1.065	0.808
4	0.0062	0.0098	0.0046	0.824	0.727	1.018	0.941	0.961	0.782
ΔCPI ex food and energy									
1	0.0036	0.0074	0.0015	0.723	0.590	1.307	1.165	1.495	0.461
2	0.0043	0.0088	0.0017	0.773	0.687	1.169	1.047	1.191	0.558
3	0.0047	0.0089	0.0020	0.755	0.705	1.037	1.027	1.100	0.579
4	0.0053	0.0104	0.0024	0.727	0.647	0.979	1.011	1.091	0.539
ΔPPI									
1	0.0088	0.0131	0.0078	0.825	0.695	0.988	0.893	0.949	0.912
2	0.0096	0.0138	0.0084	0.846	0.834	0.937	0.907	0.917	0.918
3	0.0095	0.0147	0.0082	0.861	0.842	0.951	0.927	0.937	0.922
4	0.0102	0.0165	0.0088	0.860	0.816	0.975	0.919	0.907	0.883

Forecast results

- (Relative) RMSEs of forecasts of mon./fin. variables

	RMSE(AR)			Const. FAVAR/AR			TV-FAVAR/const. FAVAR		
h	all periods	recessions	after 1995	all periods	recessions	after 1995	all periods	recessions	after 1995
ΔM2									
1	0.0064	0.0052	0.0054	0.819	1.052	0.882	0.789	0.864	0.703
2	0.0074	0.0055	0.0059	0.902	1.126	0.992	0.844	0.937	0.805
3	0.0076	0.0061	0.0055	0.914	1.057	1.081	0.912	0.934	0.903
4	0.0079	0.0061	0.0057	0.891	1.049	1.060	0.944	0.916	0.875
ΔC&I loans									
1	0.0127	0.0168	0.0109	0.894	0.865	0.910	0.835	0.816	0.716
2	0.0159	0.0227	0.0135	0.878	0.852	0.855	0.813	0.802	0.732
3	0.0169	0.0230	0.0159	0.859	0.886	0.825	0.824	0.813	0.808
4	0.0176	0.0212	0.0173	0.864	0.971	0.824	0.882	0.928	0.836
ΔReal estate loans									
1	0.0092	0.0074	0.0126	0.907	0.849	0.892	0.768	0.813	0.750
2	0.0113	0.0098	0.0144	0.893	0.738	0.891	0.750	0.951	0.717
3	0.0116	0.0102	0.0134	0.919	0.771	0.958	0.824	0.985	0.823
4	0.0122	0.0106	0.0137	0.901	0.756	0.932	0.849	1.001	0.845
ΔS&P 500									
1	0.0644	0.0862	0.0574	0.952	0.902	1.007	0.961	0.948	0.939
2	0.0653	0.0896	0.0584	0.977	0.982	1.035	0.966	0.935	0.901
3	0.0653	0.0904	0.0594	0.980	0.979	1.028	0.967	0.965	0.896
4	0.0655	0.0906	0.0597	0.985	0.985	1.013	0.965	0.981	0.913
ΔHouse price									
1	0.0137	0.0115	0.0133	0.820	0.690	0.899	0.917	0.923	0.861
2	0.0137	0.0116	0.0132	0.885	0.774	0.923	0.983	1.050	0.961
3	0.0144	0.0116	0.0136	0.903	0.757	0.962	0.972	1.038	0.955
4	0.0144	0.0117	0.0137	0.906	0.794	0.964	0.981	1.033	0.960

Summary: Forecasting with the TV-FAVAR

- Const. FAVAR in general beats AR (gains from large information set), even some gains for h=1 for stock prices and exchange rates (not shown).

- For most variables, gains from FAVAR over AR larger during recessions. Pattern due to marked increase in RMSE of AR model.

- Over entire period, TV-FAVAR better than const. FAVAR in most cases. Largest gains (10-25%) for inflation measures, C&I and real estate loans, FFR, M2.

Summary: Forecasting with the TV-FAVAR (cont.)

- TV-FAVAR and const. FAVAR comparable during recessions. Large information set matters more than time variation in parameters.

- Relative performance of const. FAVAR deteriorates in 1995-2007, RMSE of AR drastically reduced for most variables …

- … but TV-FAVAR often even better, and better than const. FAVAR for virtually any variable and horizon. Results for inflation particularly good.

4. Structural analysis

Reasons for changing monetary policy transmission

- Long-lasting changes
  - modifications in conduct of monetary policy (e.g. Boivin and Giannoni 2002) → may have weakened MP transmission
  - liberalization and innovation on financial markets → may have weakened or strengthened MP transmission
  - globalization (e.g. Boivin and Giannoni 2008) → may have weakened MP transmission

- (Temporary) changes over the business cycle
  - due to e.g. menu costs and a convex supply curve monetary policy may affect activity more strongly in recessions than in expanions (Peersman and Smets 2002)

Existing empirical evidence on changes in monetary policy transmission

- Despite of many studies, still no consensus on how MP transmission has changed over time

- Consensus that volatility of MP shocks has declined since the early 1980s (e.g. Primiceri 2005, Canova and Gambetti 2009)

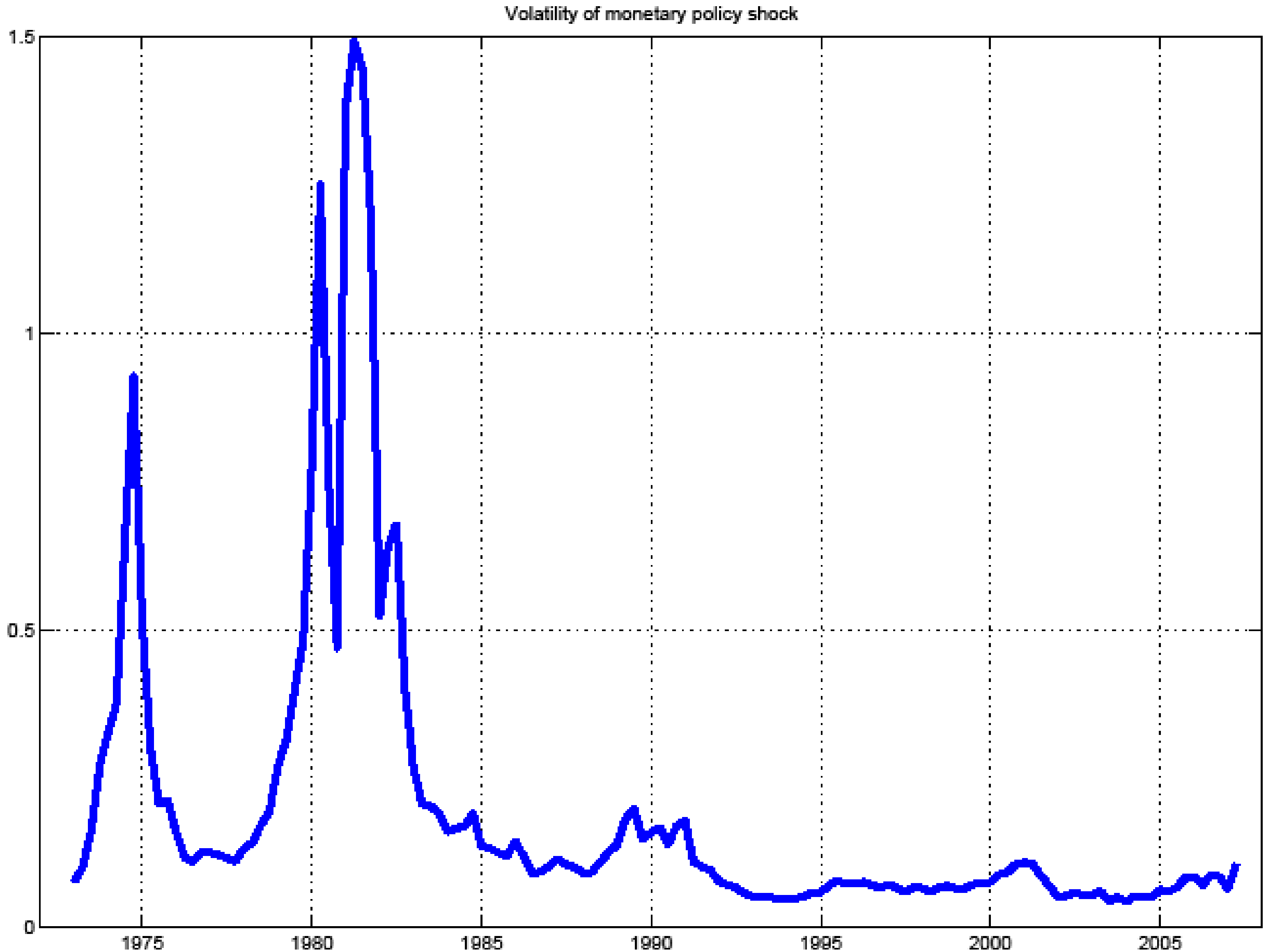
- MP found to have greater effects during recessions than during expansions in the EA (Peersman and Smets 2002) – evidence for US missing

Monetary policy shock identification

- Following Bernanke et al. (2005): add the FFR as observable in the factor VAR (after having estimated factors from slow-moving variables only and having removed FFR from space spanned by latent factors)

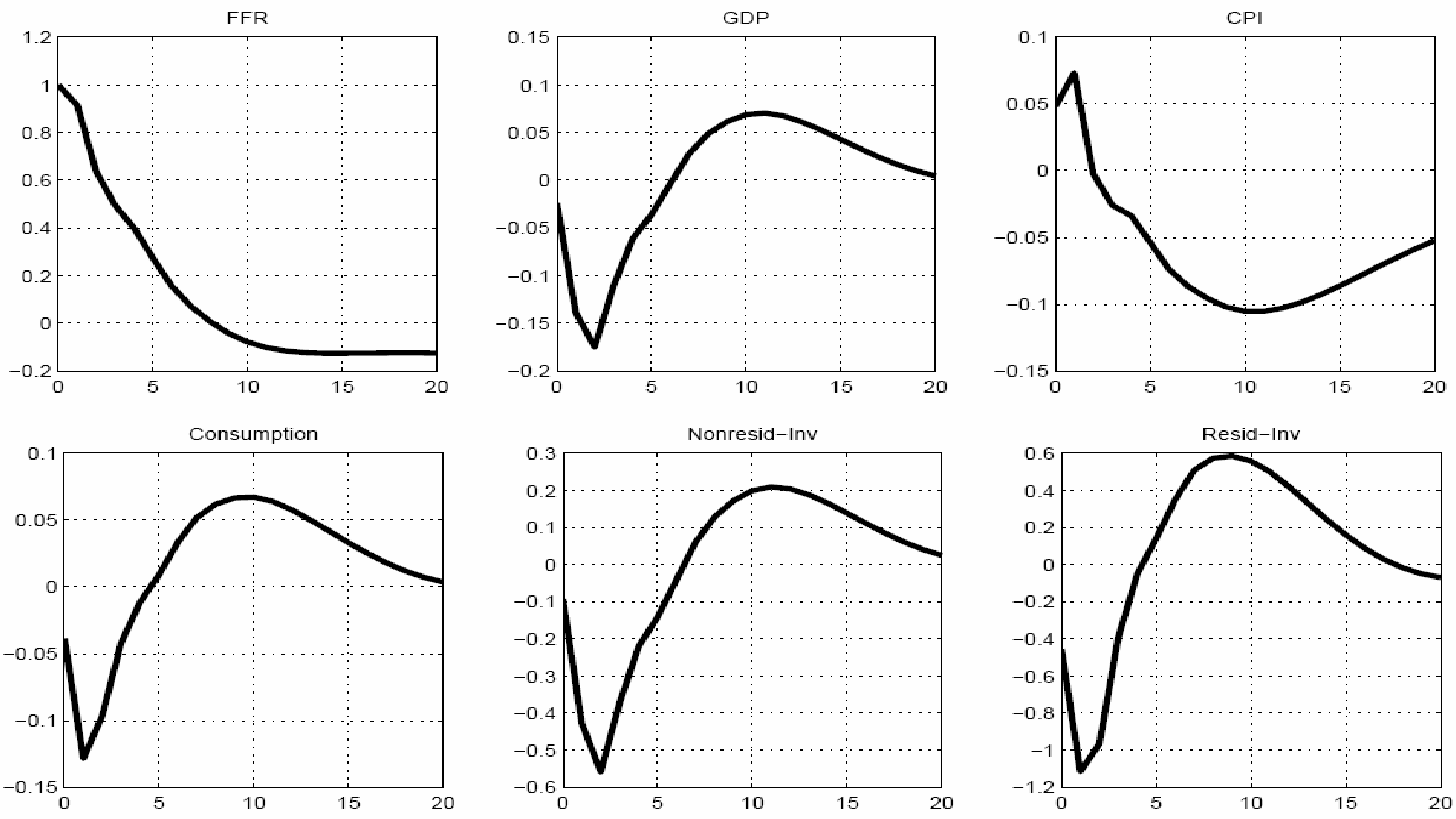
- Structural representation used for estimation implies recursive/triangular structure of VAR innovations with FFR ordered last

Volatility of monetary policy shocks



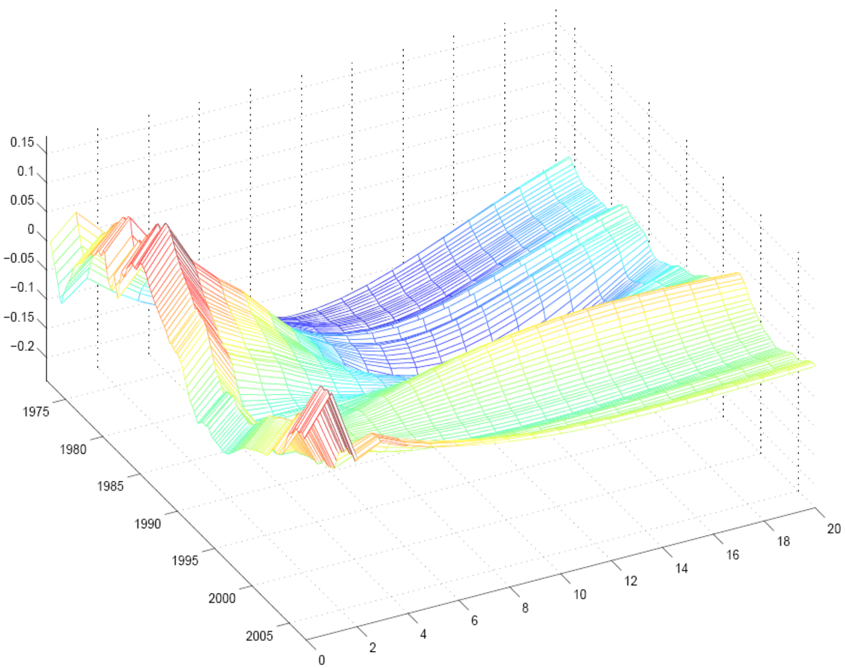
Impulse responses – constant parameters

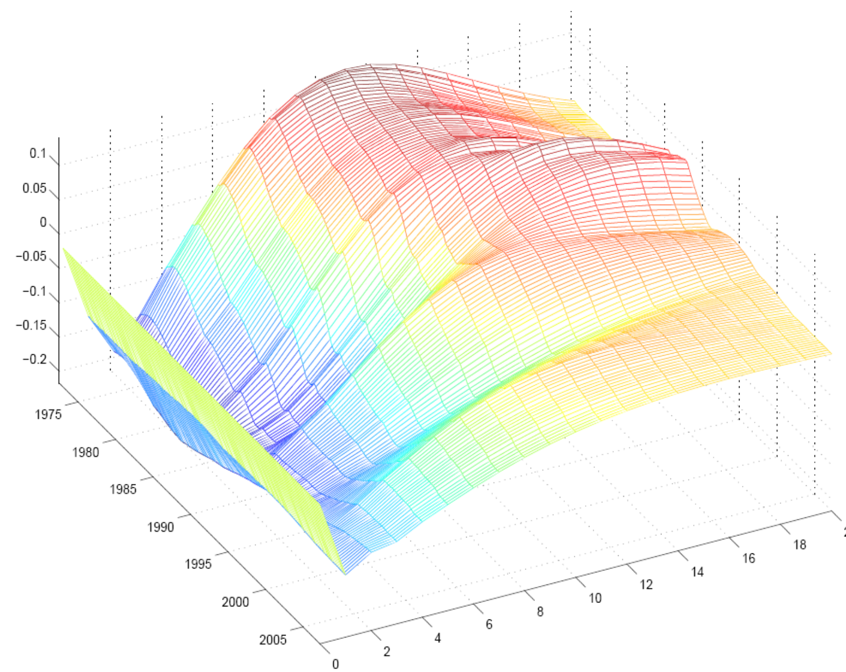
• IRFs of FFR, GDP growth, CPI inflation (top), consumption, non-residential and residential investment growth (bottom)



Impulse responses – time-varying parameters

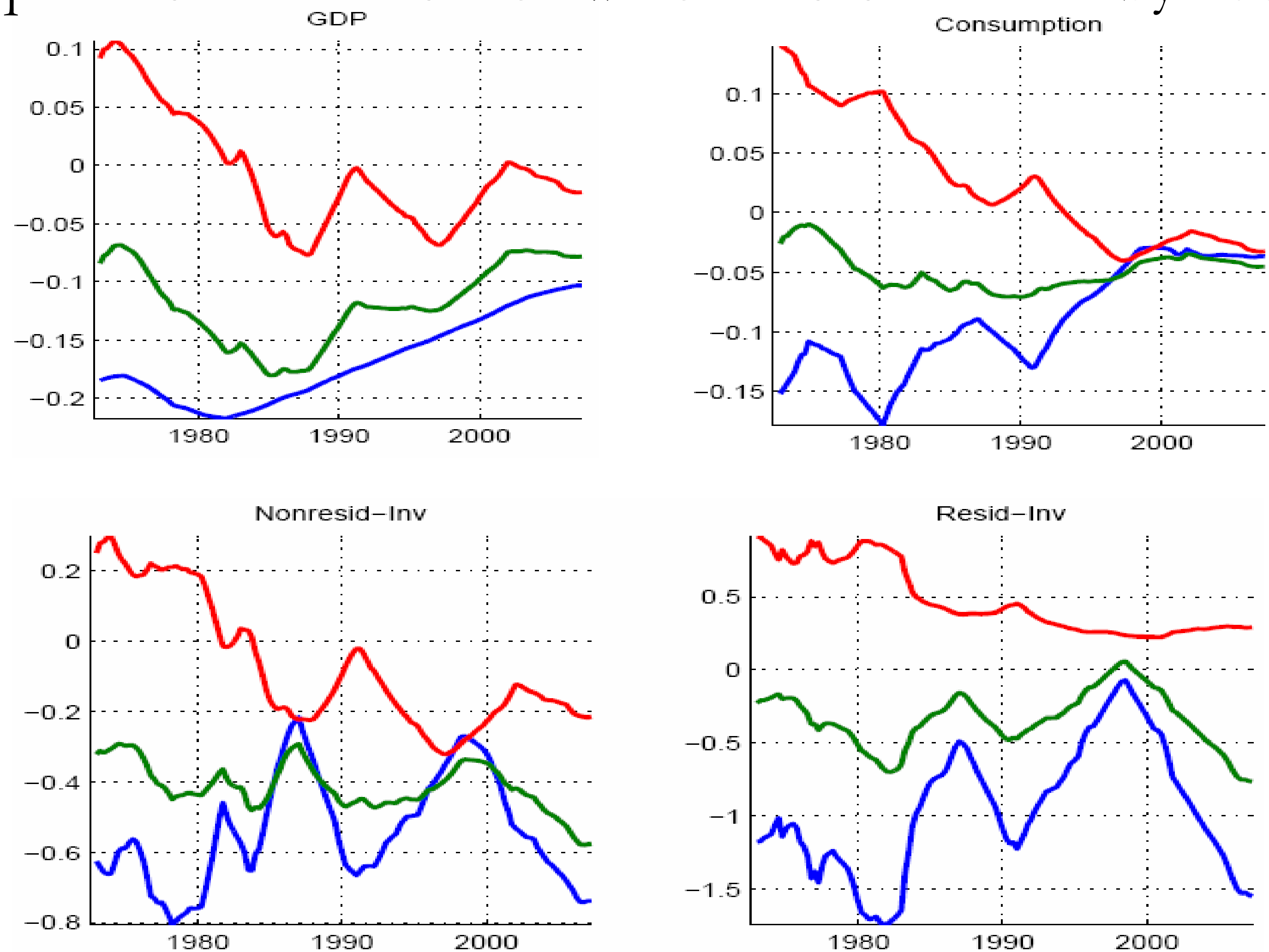
- IRFs of CPI inflation (left) and GDP growth (right) to a MP shock which raises the FFR by 100 bp.





Impulse responses – time-varying parameters

- IRFs of GDP and cons. growth (top), non-resid. and resid. inv. growth (bottom), for fixed horizons of 2 (blue), 4 (green) and 8 (red) quarters to a MP shock which raises the FFR by 100 bp.



Summary: Structural analysis

- Smaller MP shock variance after the early 1980s. Small increases during recessions

- Effects on inflation and – since the mid-1980s for shorter horizons – also on GDP growth weakened over time

- Stronger effects of MP shocks during recessions than during expansions on consumption and investment growth, but not on GDP growth

5. Conclusion

Conclusion

- Classical approach to estimate a FAVAR with smoothly time-varying parameters.

- Main findings from our applications
  - Substantial time variation in loadings and shock volatility, less time variation in dynamic and contemporaneous factor relations
  - TV-FAVAR forecasts generally better than const. FAVAR forecasts, especially for more recent years and for inflation and financial variables
  - Weaker effects of MP on inflation and (since the mid-1980s and for shorter horizons) on GDP growth; stronger effects of MP shocks on consumption and investment growth in recessions than in expansions

- Model and estimation approach transferable to other applications with large datasets, where accounting for possibly time-varying structure appears relevant