Efficient Estimation and Forecasting in Dynamic Factor Models with Structural Instability

Dimitris Korobilis[a], and Christian Schumacher[b]

[a]University of Glasgow
[b]Deutsche Bundesbank

Eighth ECB Workshop on Forecasting Techniques
14 June 2014

What we do

(At least) Two lessons learned from the "Great Recession":

1. Multiple causes of shocks exist
2. Transmission of shocks is nonlinear

For that reason, we focus on both issues of monitoring large datasets (using factor models) + structural instabilities

We develop fast Bayesian estimation algorithms for dynamic factor models (DFM)...

- with **time-varying coefficients** and **stochastic volatilities**
- that can be applied to **large (macroeconomic) datasets**

How we want to do it

We are going to work with a *nonlinear state-space model*.
$\rightarrow$ Nothing exciting about this, given all the computational advances in the last 25 years or so.

However, our implementation has to satisfy certain conditions:

1. **Flexibility:** We want a flexible specification of nonlinearities
2. **Numerical stability:** Simplify model identification, and provide an algorithm that works well always
3. **Replicability:** No "hidden" tunning parameters/priors that affect our results massively
4. **Computational tractability:** Provide a simple code, which has minimal possibility for programming errors, and can be used by non-econometricians

Online (EKF/UKF, UPF) and offline/batch (MCMC) estimation can become computationally demanding, and numerically unstable
$\rightarrow$ Identification of time-varying parameters + factors is hard

Other methods such as variational Bayes and basis function expansions, rely on calculating approximations to the exact posterior, which might not always work for all datasets/specifications

Many methods require vicious programming, thus, increasing the possibility of things going wrong even in simple models (e.g. Del Negro and Primiceri, 2013).

Academics can fine-tune complex models, but can economists replicate results?

Literature

Recent literature on structural instabilities and factor models:

- Banerjee, Marcellino and Masten (2008)
- Del Negro and Otrok (2008)
- Stock and Watson (2009)
- Breitung and Eickmeier (2011)
- Eickmeier, Lemke and Marcellino (2011)
- Bates, Plagborg-Moller, Stock and Watson (2013)
- Cheng, Liao and Schorfheide (2013)

Methodology

Large variable $(n \times 1)$ vector x_t follows the factor model with k factors

$$
\begin{aligned}
x_t &= \lambda_t f_t + \varepsilon_t, \varepsilon_t \sim N(0, V_t), V_t \text{ is diagonal} \qquad (1) \\
f_t &= B_t f_{t-1} + \eta_t, \eta_t \sim N(0, Q_t) \qquad (2)
\end{aligned}
$$

Following standard practice (Cooley, 1971) we define:

$$
\begin{aligned}
\lambda_t &= \lambda_{t-1} + u_t, u_t \sim N(0, R_t) \qquad (3) \\
\beta_t &= \beta_{t-1} + v_t, \beta_t = \text{ vec}\left(B_t'\right), v_t \sim N(0, W_t) \qquad (4)
\end{aligned}
$$

We have a time-invariant FM, when $R_t = W_t = 0$ and $V_t = V$ and $Q_t = Q$.

Other possibilities for structural instabilities exist (e.g. few breaks); however, we use TVP following Granger (2008)

1st step: forgetting factors

$$\begin{aligned} \lambda_t &= \lambda_{t-1} + u_t, u_t \sim N(0, R_t) & (5) \\ \beta_t &= \beta_{t-1} + v_t, \beta_t = \text{vec}\left(B_t'\right), v_t \sim N(0, W_t) & (6) \end{aligned}$$

R_t and W_t large, e.g. 130 variables & 10 factors, R_t is $1,300 \times 1,300$
$\rightarrow$ Do we ever have enough information in data (or prior) to estimate such a covariance matrix?
$\rightarrow$ Likelihood-based estimation requires to evaluate some function of $(\lambda_t - \lambda_{t-1})'(\lambda_t - \lambda_{t-1})$, but λ_t is latent!

Regardless of the filtering methods used (linear/nonlinear), we can easily estimate instead:

$$\begin{aligned} R_t &= \left(\mu_1^{-1} - 1\right) P^{\lambda}_{t-1|t-1}, & (7) \\ W_t &= \left(\mu_2^{-1} - 1\right) P^{\beta}_{t-1|t-1}, & (8) \end{aligned}$$

- In the formulation above, $0 < \mu_1, \mu_2 \leq 1$ are *forgetting factors*
- They allow estimation in an exponentially weighted window of data
- Observations j period in the past have weight μ_i^j
- $\mu_i = 1$ gives recursive OLS
- $\mu_i < 1$ means that we are using an effective window of $\frac{1}{1-\mu_i}$
- Very straightforward interpretation as a "prior on time-variation": lower values imply more time-variation
- Massively lower sensitivity to its values than the respective tuning constant of the IW prior (k_Q in Primiceri, 2005)
- We can easily search for optimal μ_i in a grid (Koop and Korobilis, 2013), or use MCMC to estimate it (Windle and Carvalho, 2013)

2nd step: Decay factors

For V_t and Q_t we want to use a simple recursive estimator to avoid simulation methods.
Following Koop and Korobilis (2014) we define EWMA filters

$$\widehat{V}_t = \delta_1 \widehat{V}_{t-1} + (1 - \delta_1)\, diag\left(\widehat{\varepsilon}_t \widehat{\varepsilon}_t'\right), \quad (9)$$

$$\widehat{Q}_t = \delta_2 \widehat{Q}_{t-1} + (1 - \delta_2)\, \widehat{\eta}_t \widehat{\eta}_t', \quad (10)$$

given initial values $V_0 = diag\left(\underline{V}\right)$ and $Q_0 = \underline{Q}$

This procedure provides fast point estimates of the covariances, for inference using the mean (e.g. MSFE forecasts, or point estimate of the factors).

$\rightarrow$ EWMA has similar properties to IGARCH(1,1)

Equivalent Bayesian procedure (Wishart Matrix Discounting, WMD):

$$
\begin{aligned}
V_t &\sim iW\left(S_t, n_t\right), \quad (11)\\
Q_t &\sim iW\left(\Psi_t, v_t\right), \quad (12)
\end{aligned}
$$

given priors $V_0 \sim iG\left(\underline{S}, \underline{v}\right)$ and $Q_0 \sim iW\left(\underline{\Psi}, \underline{n}\right)$, where

- $n_t = \delta_1 n_{t-1} + 1$
- $S_t = \left(1 - n_t^{-1}\right) S_{t-1} + n_t^{-1} \left[S_{t-1}^{1/2} \widetilde{V}_{t-1}^{-1/2} \left(\widehat{\varepsilon}_t \widehat{\varepsilon}_t'\right) \widetilde{V}_{t-1}^{-1/2} S_{t-1}^{1/2}\right]$,
- $v_t = \delta_2 v_{t-1} + 1$
- $\Psi_t = \left(1 - v_t^{-1}\right) \Psi_{t-1} + v_t^{-1} \left[\Psi_{t-1}^{1/2} \widetilde{Q}_{t-1}^{-1/2} \left(\widehat{\eta}_t \widehat{\eta}_t'\right) \widetilde{Q}_{t-1}^{-1/2} \Psi_{t-1}^{1/2}\right]$.

In both cases, δ_1, δ_2 are *decay factors* with similar properties as the forgetting factors defined previously.

Quintana and West (1987); West and Harrison (1997)

Our feasible approximations

We build on ideas of Koop and Korobilis (2013) that once covariance matrices are known, does not need repeated sampling (Monte Carlo) for estimation; just one run of the KF or KFS is enough.

In order to achieve this we define:

1. Don't estimate $(\lambda_t, f_t, \beta_t)$ jointly as one state using nonlinear SS methods; rather break the problem into three separate, conditionally linear SS models
2. When we require the time t sq. errors (e.g. $(\widehat{\varepsilon}_t \widehat{\varepsilon}_t')$), then use instead

$$\widehat{\varepsilon}_t = x_t - \widehat{\lambda}_{t|t-1} f_t \tag{13}$$

   (similarly for $\widehat{\eta}_t$)

Approximation 1: Estimate (conditionally) linear models

We follow the main idea of Doz et al. (2011):

1. Obtain the Principal component (assuming that any other factor estimate of the TVP-DFM is expensive to obtain)
2. Estimate λ_t and β_t using Kalman filter/smoother conditional on PC
3. Estimate factors f_t conditional on all coefficients and volatilities using Kalman filter/smoother

- Single iteration provides reasonable estimates of vols, tvp & f_t
- "Exact methods" would be subject to large estimation error, anyway
- As in Doz et al. (2011) we may iterate for increased precision

$\rightarrow$ In simulations the log-lik will always converge (but might not always be monotonically increasing)

Approximation 2: Feasible estimation of residuals

Our second approximation implies to use:

$$\widehat{\varepsilon}_t = x_t - \widehat{\lambda}_{t|t-1} f_t^{PC} \tag{14}$$

(similarly for $\widehat{\eta}_t$).

- We use an estimate $\widehat{\lambda}_{t|t-1}$ instead of $\lambda_t = \widehat{\lambda}_{t|t}$
- That way, we can run the Kalman filter using one iteration:

$\rightarrow$ V_t can be updated based on $\widehat{\varepsilon}_t$, then update $\lambda_{t|t}$ given V_t

- This allows easy identification of vols and tvps:

$\rightarrow$ We estimate V_t using information which is available at t; similarly for λ_t

$\rightarrow$ MCMC would provide sample from $p(V_t|-)$ given a "guess" of λ_t and vice-versa (convergence not guaranteed)

Empirical evidence

We implement three exercises:

1. Monte Carlo simulations
2. Forecasting German GDP using a large panel of variables
3. Forecasting Eurozone sovereign bond-yields

I won't go through details with 1. and 2. due to the limited time.
Only comment is:

- If the true model is TVP-DFM, then our estimator gives better SFF0 statistics than PC and Doz et al. (2011)
- For German data, we obtain better MSFE than forecasting with PC, and time-varying volatility contributes to forecast accuracy more than time-varying loadings or VAR lag polynomial parameters.

I WILL FOCUS ON THE EURO-AREA BOND-YIELDS.

Eurozone sovereign bond yields

- We have 10y bond rates for 10 countries:
  Austria, Belgium, Finland, France, Greece, Ireland, Italy, Netherlands, Portugal, Spain
- Data are expressed in spreads from the 10y German bund
- Sample is 1999m1 - 2012m12
- Data have an obvious factor structure (comovements during Global, and Euro-Area crises)
- Data have obvious structural breaks, and changing volatility
- Most importantly: Data are explosive $\rightarrow$ 1st PC collapses to sample mean (loadings degenerate to 1) $\rightarrow$ Application of TVP-DFM is nonsense!

The data

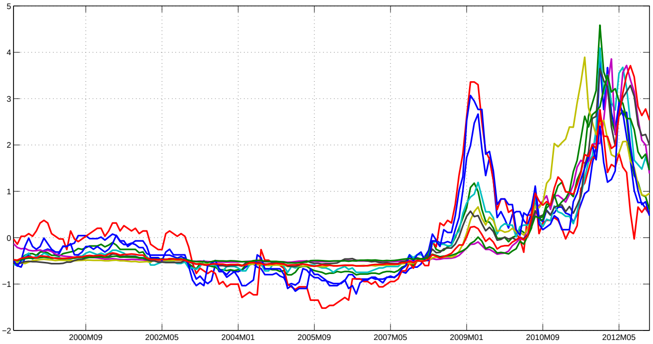
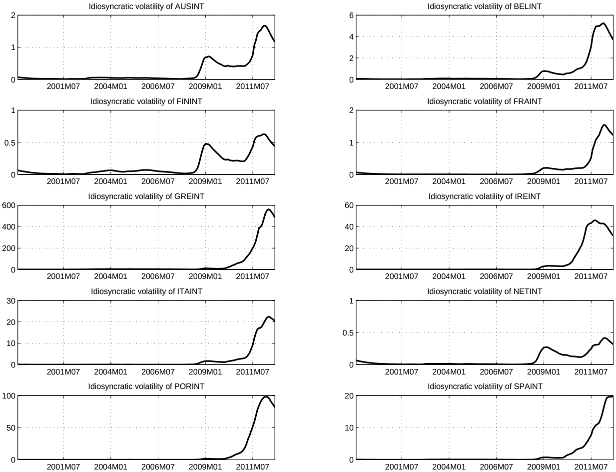
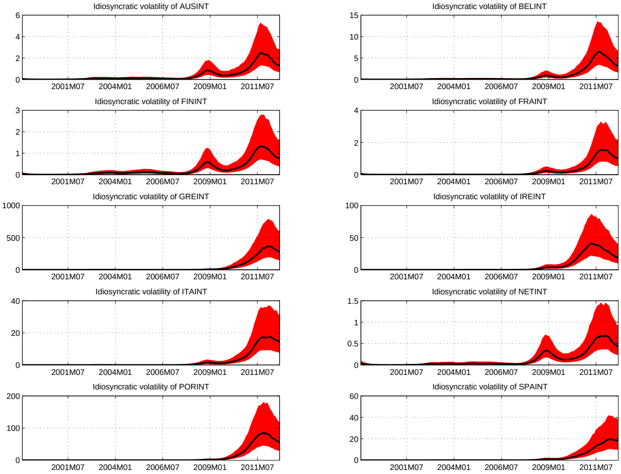


Figure: Returns on 10-year bond for 10 Eurozone countries, 1999m1-2012m12. The data are expressed as spreads from the 10-year yield of the German bund, and then standardized (mean zero, variance one) in order to be of comparable scale.

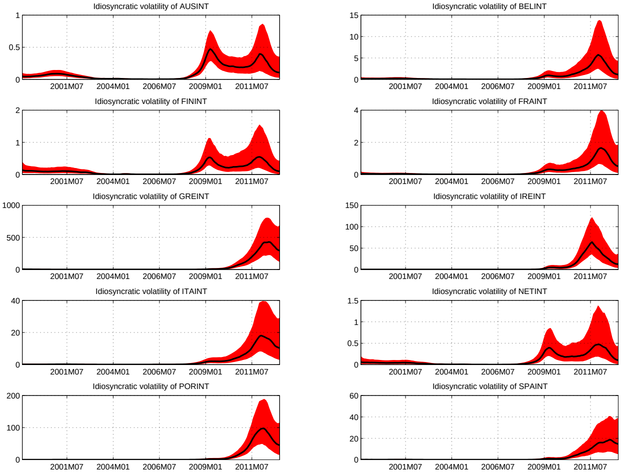
Idiosyncratic volatilities, EWMA



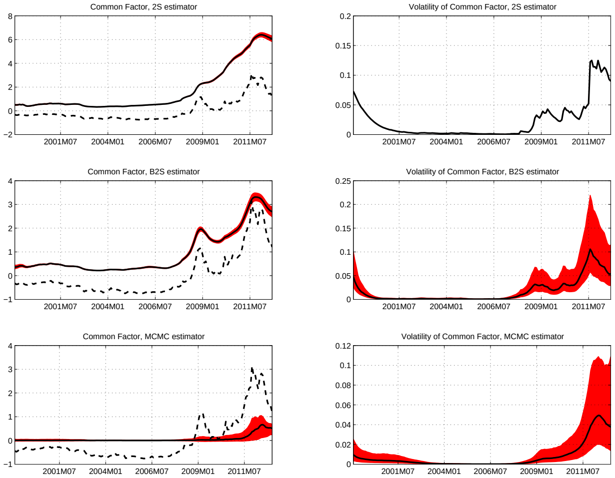
Idiosyncratic volatilities, WMD



Idiosyncratic volatilities, MCMC + SV



Factor estimates, and factor vols



Forecasting bond yields

We implement forecasting, using 1 factor model to remove uncertainty regarding number of factors

Many studies use one (EA) or two (core/periphery) factors, or three (level, slope, curvature) if a DNS model applies to the whole yield curve

For the Bayesian, such uncertainty can be dealt with using MLs, PLs, BMA, shrinkage priors, EB priors, hierarchical priors etc.

We forecast using the whole TVP-DFM system, i.e.

$$x_{t+1} = \lambda_{t+1} f_{t+1} + \varepsilon_{t+1} \quad (15)$$

$$f_{t+1} = B_{t+1} f_t + \eta_{t+1} \quad (16)$$

but we don't simulate $\lambda_{t+h} = \lambda_t$ and $B_{t+h} = B_t \; \forall \; h$ (i.e. we only forward f_{t+h}, but not tvps and vols)

Table 1. Initial conditions and priors used in different models

TVP-DFM			
Parameter	2S-EWMA	2S-WMD	MCMC
λ_t		$\lambda_0 \sim N(0, 4 \times I)^{a}$	
β_t		$\beta_0 \sim N(0, 1 \times I)^{a}$	
f_t		$f_0 \sim N(0, 4 \times I)^{a}$	
R_t		FF with $\mu_1 = 0.99^{a}$	
W_t		FF with $\mu_2 = 0.99^{a}$	
V_t	$V_{0,i} \equiv 0.1^{b}$	$V_{0,i} \sim IW(I, 1+2)^{c,b}$	$log(V_{0,i}) \sim N(\log(0.5), 0)^{b}$
Q_t	$Q_0 \equiv 0.1$	$Q_0 \sim IW(I, k+2)^{c}$	$log(Q_0) \sim N(\log(0.5), 0)$

DFM	
Parameter	MCMC
λ	$\lambda \sim N(0, 4 \times I)$
β	$\beta \sim N(0, 1 \times I)$
f_t	$f_0 \sim N(0, 4 \times I)$
V	$V_i \sim IG(0.01, 0.01)^{b}$
Q	$Q \sim IG(0.01, 0.01)$

[a]These coefficients are common in all three specifications of the TVP-DFM

[b]$V_{0,i}$ (similarly, V_i) denotes the i-th diagonal element of V_t (similarly, V), $i = 1, ..., 10$.

[c]The IW for a univariate variable is equivalent to the Inverse Gamma (IG) distribution.

The previous Table shows initial conditions for all models estimated
We try to use similar priors and initial conditions whenever possible

In the next Table we present posterior predictive likelihoods, PLs, (not to be confused with prior predictive, i.e. marginal, likelihoods)

In order to achieve this, we use Monte Carlo Integration to obtain samples from all the posteriors
$\rightarrow$ We do that whenever this possible, e.g. in the EWMA we have point estimate.

The PLs in the Table below are relative to the PL of the constant parameter DFM estimated with MCMC.

Table 2. Relative PL's for the bond yield data

	$h=1$	$h=3$	$h=6$	$h=9$	$h=12$
			2S-EWMA		
Austria	1.05	1.00	0.89	0.77	0.74
Belgium	0.90	0.86	0.89	0.82	0.80
Finland	1.03	0.95	0.93	0.86	0.80
France	1.17	1.15	1.05	1.09	1.01
Greece	1.85	1.79	1.73	1.86	1.88
Ireland	1.45	1.09	0.95	0.85	0.77
Italy	2.90	2.81	3.03	4.62	4.20
Netherlands	0.95	0.89	0.92	0.81	0.76
Portugal	1.04	1.05	1.22	1.13	1.16
Spain	1.17	1.01	0.90	0.81	0.78
			2S-WMD		
Austria	1.17	1.08	0.98	0.88	0.82
Belgium	0.95	0.91	0.96	0.89	0.86
Finland	1.19	1.05	1.01	0.89	0.85
France	1.17	1.17	1.10	1.19	1.08
Greece	1.91	1.98	1.93	1.82	1.52
Ireland	1.53	1.15	0.98	0.91	0.83
Italy	3.41	3.08	3.18	5.07	4.64
Netherlands	1.06	0.98	1.03	0.87	0.81
Portugal	1.05	1.00	1.16	1.14	1.15
Spain	1.29	1.10	0.99	0.89	0.87
			MCMC		
Austria	1.32	1.15	0.99	0.91	0.84
Belgium	1.04	0.95	0.96	0.92	0.87
Finland	1.20	1.02	0.91	0.77	0.79
France	1.13	1.16	1.06	1.14	1.05
Greece	2.79	2.67	2.55	2.43	1.86
Ireland	1.82	1.28	1.09	0.95	0.84
Italy	2.11	1.99	2.11	3.03	2.87
Netherlands	0.90	0.88	0.95	0.86	0.82
Portugal	0.97	0.97	1.12	1.06	1.12
Spain	1.22	1.02	0.95	0.84	0.83

Discussion

We can make the following observations:

1. There is a strong case for using a DFM with structural instabilities for the Eurozone data, compared to a constant parameter DFM.
2. Our proposed estimation methods compare as well in achieving similar predictive likelihoods to MCMC.
3. We should also consider in this evaluation that the Monte Carlo versions of our two-step algorithms are several times faster than MCMC.
4. Additionally, we haven't imposed any identification restrictions to obtain our results. This is not the case for the MCMC algorithm, where identification is far more challenging.

1. Our algorithms can also be extended in several directions. For example, the results above suggest that for some countries (e.g. Belgium) a TVP-DFM might not be appropriate, and one might be better-off with a constant parameter DFM. Similar results have been found in the macro-finance literature, e.g. Sims and Zha (2006). If we allow each DFM equation to have a different forgetting factor for $\lambda_{i,t}$, that is, if we make μ_1 a diagonal matrix instead of a scalar, we can allow some countries to have constant loadings while others to have time-varying.
2. In a similar spirit, one can optimize the model in terms of all the forgetting and decay factors used.
3. As in Koop and Korobilis (2014), we can make the argument that estimation error can be balanced by the fact that we can estimate millions of TVP-DFMs and reduce model uncertainty instead of worrying about estimation accuracy in a heavily parametrized model!

Thank you!