Joint modelling and estimation of global and local cross-sectional dependence in large panels

Quint Wiersma, Julia Schaumburg

Vrije Universiteit Amsterdam, Tinbergen Institute

1 Contribution and main findings

- Unified approach to identifying and estimating spatio-temporal dependence structures in large panels.
- High-dimensional factor model in state-space form meets spatial error model with heterogeneous spillover intensity parameters.
- New, efficient algorithm combines model selection via Lasso and/or group Lasso with filtering and estimation.
- Monte-Carlo simulations: good performance in terms of determining presence of latent factors, sparsity pattern of loading matrices and magnitudes of coefficients.
- Application to monthly US interest rate data on 15 maturities (to be scaled up) over almost 40 years.
- Empirical evidence for time-varying number of common factors as well as increasing local dependence between neighboring maturities.

2 Dynamic factor model with spatial errors

2.1 Baseline model

- Dynamic factor model with spatial errors and exogeneous regressors:

$$\begin{aligned} y_t &= X_t\beta + \Lambda f_t + \xi_t\,, & \\ f_{t+1} &= \phi f_t + \eta_t\,, & \eta_t \sim \mathcal{N}(0,\Sigma_\eta) \\ \xi_t &= \rho W \xi_t + \varepsilon_t\,, & \varepsilon_t \sim \mathcal{N}(0,\Sigma_\varepsilon) \end{aligned}$$

  where
  - y_t is an N-dimensional time series, X_t is a $N \times K$ matrix of regressors,
  - f_t is an r-dimensional vector of latent factors,
  - Λ is an $N \times r$ loading matrix,
  - W is a $N \times N$ exogenous matrix of spatial weights,
  - $\beta = (\beta_1, ..., \beta_K)'$, $\phi = \text{diag}(\phi_1, ..., \phi_r)$, ρ are unknown coefficients,
  - Σ_η and Σ_ε are diagonal covariance matrices.
- Model can be written in standard state space form. If dimensions are not too high, filtering and estimation via Kalman filter and maximum likelihood.
- Here: N large, r unknown, Λ potentially sparse, ρ present or not.
  $\Rightarrow$ Regularized log-likelihood function (see Section 3).

2.2 Extensions

- Heterogeneous spatial dependence: extend model by replacing the spatial error equation by

$$\xi_t = PW\xi_t + \varepsilon_t\,,$$

  where $P = \text{diag}(\rho) = \text{diag}(\rho_1, \ldots, \rho_N)$.
- Heterogeneous slope coefficients: Model for y_t becomes

$$y_t = Bx_t + \Lambda f_t + \xi_t\,,$$

  where x_t is the $KN \times 1$ vector of regressors, i.e. $x_t = \text{vec}(X_t')$, and B is the $N \times KN$ block diagonal matrix $\text{diag}(\beta_1', ..., \beta_N')$, where β_i is a K-dimensional slope coefficient vector.

3 Estimation algorithm

- Objective function is ℓ_1 penalized state space log likelihood function

$$\begin{aligned} \mathcal{L}(\theta) = &-\frac{NT}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\log|F_t| + v_t' F_t^{-1} v_t\right) \\ &- \gamma_\rho \sum_{i=1}^{N}|\rho_i| - \gamma_\beta \sum_{i=1}^{N}\sum_{k=1}^{N}|\beta_{ik} - \gamma_{\Lambda,1}\sum_{i=1}^{N}\sum_{j=1}^{r}|\Lambda_{ij}| - \gamma_{\Lambda,2}\sum_{j=1}^{r}\sqrt{N}\|\Lambda_j\|_2\,, \end{aligned}$$

  where γ_ρ, γ_β, $\gamma_{\Lambda,1}$, and $\gamma_{\Lambda,2}$ control the amount of penalization.

- Standard gradient-based optimization methods are expected to be cumbersome and unstable due to high dimensions and non-smooth objective function.
  $\Rightarrow$ Instead: new tailored ECM algorithm, exploiting part of the linear structure of the model, combining coordinate descent techniques and a proximal gradient algorithm to ensure spatial stability.

Algorithm 1: Coordinate Descent Proximal ECM Algorithm

Input: Tuning parameters $\gamma_{\Lambda,1} = \gamma^*_{\Lambda,1}$, $\gamma_{\Lambda,2} = \gamma^*_{\Lambda,2}$, $\gamma_\beta = \gamma^*_\beta$, and $\gamma_\rho = \gamma^*_\rho$ and initial parameter vector $\theta^{(0)}$.

Iterate until convergence: For iteration k:

E-step

- Run the Kalman filter and smoother recursions to obtain estimates for state $\hat{f}_t^{(k)}$, state variance $V_t^{(k)}$, and state covariance $V_{t,t-1}^{(k)}$ (based on the estimates of iteration $k-1$).
- With these estimates, calculate components of expected conditional EM log likelihood function.

CM-step

- Estimate $\phi^{(k)}$ and $\Sigma_\eta^{(k)}$ via least squares.
- Estimate $\Lambda^{(k)}$, $\beta^{(k)}$, $\Sigma_\varepsilon^{(k)}$, and $\rho^{(k)}$ as follows:
  - Given the current estimates $\beta^{(k-1)}$, $\Sigma_\varepsilon^{(k-1)}$ and $\rho^{(k-1)}$, estimate $\Lambda^{(k)}$ as follows:
    Iterate until convergence: For iteration n.
    - Update $\Lambda^{(k,n)}$ using a blockwise proximal gradient descent algorithm for sparse group Lasso.
  - Similarly, given the current estimates $\Lambda^{(k)}$, $\Sigma_\varepsilon^{(k-1)}$ and $\rho^{(k-1)}$, estimate $\beta^{(k)}$ as follows:
    Iterate until convergence: For iteration n.
    - Update $\beta^{(k,n)}$ using a multipe response coordinate descent algorithm for Lasso.
  - Next, given the current estimates $\Lambda^{(k)}$, $\beta^{(k)}$ and $\rho^{(k-1)}$ update $\Sigma_\varepsilon^{(k)}$.
  - Finally, given the current estimates $\Lambda^{(k)}$, $\beta^{(k)}$ and $\Sigma_\varepsilon^{(k)}$ estimate $\rho^{(k)}$ using the proximal gradient descent algorithm.

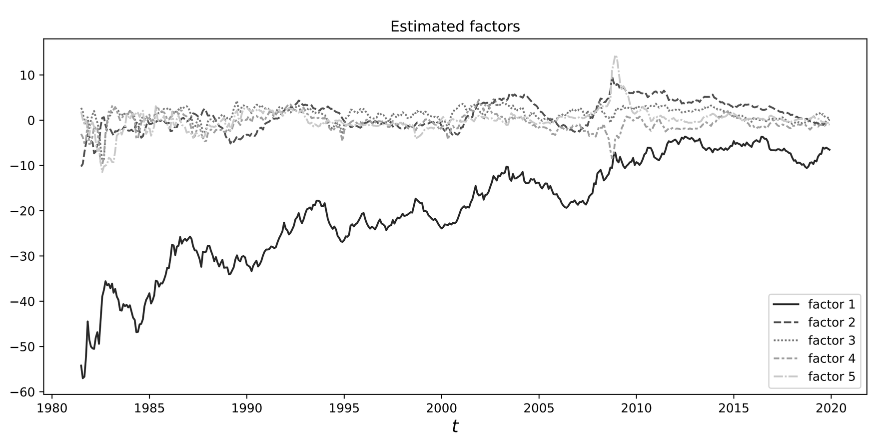
4 Monte Carlo simulations

		$r=2$			$r=4$		
		Frob. norm (Λ)	correct zeros (Λ)	RMSE (ρ)	Frob. norm (Λ)	correct zeros (Λ)	RMSE (ρ)
$N=20$	$T=200$	0.0155	0.9849	0.0155	0.0362	0.9127	0.0212
	$T=500$	0.0108	0.9986	0.0102	0.0307	0.9553	0.0149
	$T=1000$	0.0091	0.9999	0.0073	0.0297	0.9597	0.0119
$N=50$	$T=200$	0.0103	0.9862	0.0092	0.0242	0.9246	0.0111
	$T=500$	0.0080	0.9991	0.0057	0.0145	0.9823	0.0069
	$T=1000$	0.0072	1.0000	0.0041	0.0127	0.9920	0.0055

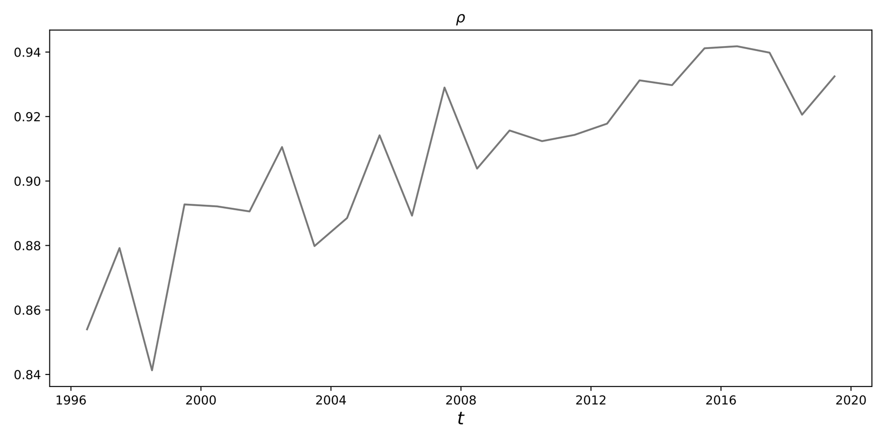
5 Application: Dynamics of the yield curve

- Background: debate on appropriate number of factors underlying yield curve dynamics (Crump, Gospodinov 2019) as well as preferred habitat model of yield curve (Vayanos, Vila 2019).
- Monthly yield data from Liu and Wu (2020) on 15 maturities capturing the short, medium and long end of the yield curve: 1, 3, 6, 9, 12, 24, 36, 48, 60, 84, 120, 150, 180, 210, and 240 months. Sample covers July 1981–December 2019.
- Full sample as well as rolling window estimation of 180 observations; model selection and choice of penalty parameters may result in time-varying number of factors as well as different sparsity patterns.
- Current results: homogeneous spatial intensity parameter, no explanatory variables.

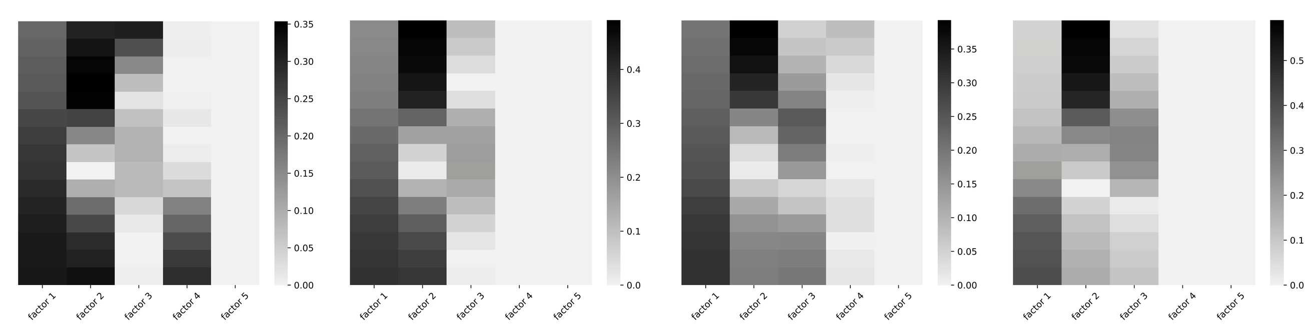
Full-sample filtered factors



Time-varying spatial intensity parameter



Loading matrix structures: 1999, 2005, 2010, 2018



6 Work in progress

- Out-of-sample forecasts, benchmarked by dynamic Nelson-Siegel model variants.
- Extend empirical model by macroeconomic and financial variables with heterogeneous slope coefficients.
- Allow for maturity- or maturity group-specific spillover intensities.

References

[1] Richard K Crump and Nikolay Gospodinov. Deconstructing the yield curve. *FRB of New York Staff Report*, (884), 2019.

[2] Yan Liu and Jing Cynthia Wu. Reconstructing the yield curve. Technical report, National Bureau of Economic Research, 2020.

[3] Dimitri Vayanos and Jean-Luc Vila. A preferred-habitat model of the term structure of interest rates. Technical report, National Bureau of Economic Research, 2009, 2019.