Regularized estimation of heterogeneous panel data models with dynamic factors and local cross-sectional dependence

Siem Jan Koopman[1,2,3], Julia Schaumburg[1,2], Quint Wiersma[1,2]

[1]Vrije Universiteit Amsterdam [2]Tinbergen Institute [3]CREATES

Motivation: Modeling the yield curve

Factor models

- Three factors: level, slope, curvature

  e.g. Nelson, Siegel (1987); Litterman, Scheinkman (1991); Diebold, Li (2006); Koopman, Mallee, van der Wel (2010); Christensen, Diebold, Rudebusch (2011)

- Three factors plus macro-finance variables/factors

  e.g. Ang, Piazzesi (2004); Diebold, Rudebusch, Aruoba (2006); Moench (2008); Ludvigson, Ng (2009); Joslin, Priebsch, Singleton (2014); Coroneo, Giannone, Modugno (2016); Byrne, Cao, Korobilis (2017)

- More than three factors (with or without macro-finance components)

  e.g. Cochrane, Piazzesi (2005); Adrian, Crumb, Mönch (2013); van Dijk, Koopman, van der Wel, Wright (2014); Bauer, Hamilton (2018)

Local spillovers

- Overlap of neighboring maturities: Crumb, Gospodinov (2022a,b)
- Preferred habitat investors: Vayanos, Vila (2021)

Methodological contribution

Agnostic empirical modeling approach for panel data, combining global (factor) and local (spatial) dependence, and allowing for:

- heterogeneous spillover intensity parameters;
- heterogeneous slope coefficients and variances;
- built-in model selection (regressors, factor loadings, spatial dependence, number of factors).

Method: High-dimensional sparse dynamic factor model with regressors and spatial errors; EM-type optimization of penalized state space likelihood function.

Empirical contribution

- In-sample modeling and out-of-sample prediction of monthly treasury bond excess returns.
- Best models feature several latent dynamic factors plus heterogeneous local dependence; evidence for maturity-specific, time-varying impact of macroeconomic regressor variables.
- Good out-of-sample performance, significant improvements over expectation hypothesis and factors-only model for short maturities.

Some closely related literature

- Panel data models with common factors and (heterogeneous) spatial dependence:
  Bai, Li (2021); Aquaro, Bailey, Pesaran (2020).
- High-dimensional sparse factor models:
  Kaufmann, Schumacher (2017, 2019); Frühwirth-Schnatter, Lopes (2018).

Outline

- Introduction
- **Methodology**
- Monte Carlo simulation
- Empirical application: Predicting the yield curve
- Conclusion

HSE-DNF-X Model

Heterogeneous spatial error dynamic factor model with regressors:

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

- Data and latent factors:

$$
\underset{(N\times 1)}{y_t}, \quad \underset{(K\times 1)}{X_t}, \quad \underset{(r\times 1)}{f_t}.
$$

- Parameters:

$$
\underset{(N\times K)}{\beta}, \quad \underset{(N\times r)}{\Lambda}, \quad \underset{(r\times r)}{\phi}, \quad \underset{(r\times r)}{\Sigma_\eta}, \quad \underset{(N\times N)}{\Sigma_\varepsilon}, \quad \underset{(N\times 1)}{\rho}.
$$

- $P = \mathrm{diag}(\rho)$ and W is a known $(N \times N)$ matrix of spatial weights with $w_{ii} = 0$ and maximum eigenvalue 1.

Estimation and model selection

- If N is large, the number of coefficients is large. Furthermore,
  - the true number of factors r is unknown;
  - individual elements or entire columns of β might equal zero;
  - some spillover intensities ρ_i might be equal to to each other.
- For estimation and model selection, we combine "plain" Lasso for the elements of β with group Lasso for the columns of β and Λ together with fused Lasso for the elements of ρ.
- Penalty parameter choice via information criteria or time series cross-validation.
- Optional: improve estimates using adaptive Lasso (Zou, 2006).

Estimation

- Penalized log likelihood:

$$
\begin{aligned}
L(\theta) = & -\frac{NT}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\log|F_t| + v_t' F_t^{-1} v_t\right) \\
& - \sum_{i=2}^{N}\gamma_\rho|\rho_i - \rho_{i-1}| - \sum_{i=1}^{N}\sum_{k=1}^{K_{max}}\gamma_{\beta,ik}^{l}|\beta_{ik}| - \sum_{k=1}^{K_{max}}\sqrt{N}\gamma_{\beta,k}^{gl}\|\beta_{\bullet k}\|_2 \\
& - \sum_{j=1}^{r_{max}}\sqrt{N}\gamma_{\Lambda,j}\|\Lambda_{\bullet j}\|_2, \qquad (1)
\end{aligned}
$$

  where v_t is the prediction error, F_t is its variance, $\Lambda_{\bullet j}$ denotes the j-th column of Λ and $\beta_{\bullet k}$ the k-th column of β.
- Penalty parameters γ_ρ, $\gamma_{\beta,ik}^{l}$, $\gamma_{\beta,k}^{gl}$, and $\gamma_{\Lambda,j}$ control model sparsity.
- Direct numerical optimization of (1) can be cumbersome; use Expectation-Conditional Maximization (ECM; Meng, Rubin 1993) algorithm instead.

ECM algorithm

Input: Penalty parameters $\gamma^{l}_{\Lambda,ij}$, $\gamma^{gl}_{\Lambda,j}$, $\gamma^{l}_{\beta,ik}$, $\gamma^{gl}_{\beta,k}$, and γ_ρ and initial $\theta^{(0)}$.

Iterate until convergence: For iteration $k+1$.

- E-Step:
  - Given $\theta^{(k)}$, run Kalman filter and smoother to obtain estimates of $\mathbb{E}[f_t|y_1, ..., y_T]$ and corresponding variance and autocovariance matrices.
  - Plug these objects into conditional expectation of penalized complete data log likelihood $Q\left(\theta^{(k+1)}|\theta^{(k)}\right)$.
- CM-Step:
  - Estimate $\phi^{(k+1)}$ and $\Sigma_\eta^{(k+1)}$ using least squares.
  - Sequentially optimize $Q\left(\theta^{(k+1)}|\theta^{(k)}\right)$ with respect to Λ, β, Σ_ε, and ρ, using proximal gradient descent, see Boyd et al (2011).

Adaptive ECM Algorithm

- For the adaptive version of our ECM algorithm we follow the approach of Lu and Su (2016) to obtain penalization weights for β and Λ.
  1. Obtain OLS estimates of the slopes β by regressing the K_{max} regressors on the N time series, denoted by $\widetilde{\beta}$.
  2. Obtain PCA estimates of Λ and factors f_t from the demeaned data $\widetilde{y}_t = \widetilde{\beta} X_t$, denoted by $\widetilde{\Lambda}$ and $\widetilde{f}_t$.
  3. Compute r_{max} eigenvalues of $\widehat{\Sigma}_{\widehat{f}} = T^{-1} \sum_{t=1}^{T} \widehat{f}_t \widehat{f}_t'$ arranged in descending order, denoted by $\tau_1, \ldots, \tau_{r_{max}}$, where $\widehat{f}_t = (NT)^{-1} (\sum_{t=1}^{T} \widetilde{f}_t \widetilde{y}_t') \widetilde{y}_t$.
- The adaptive penalty parameters are then given by

$$\gamma_{\beta,ik}^{l} = \frac{\gamma_{\beta}^{l}}{|\widetilde{\beta}_{ik}|} \quad \gamma_{\beta,k}^{gl} = \frac{\gamma_{\beta}^{gl}}{\|\widetilde{\beta}_{\bullet k}\|_2} \quad \gamma_{\Lambda,j} = \frac{\gamma_{\Lambda}}{\tau_j}$$

Outline

- Introduction
- Methodology
- **Monte Carlo simulation**
- Empirical application: Predicting the yield curve
- Conclusion

Monte Carlo simulation

Goal: Investigate the finite sample performance of our ECM algorithm in terms of estimation and model selection accuracy.

- Simulate from HSE-DNF-X model (1000 replications),

$$y_t = \beta X_t + \Lambda f_t + \xi_t\,, \quad f_{t+1} = \phi f_t + \eta_t\,, \quad \xi_t = PW\xi_t + \varepsilon_t.$$

- $X_{t,ik},\ \epsilon_{t,i},\ \eta_{t,j} \sim N(0,1)$.
- Regressors: $K = 1$, $K_{max} = 2$, slope coefficients $\beta_{ik} \sim U(1/2, 1)$ for $k = 1$ and $i \leq N \div 2$, otherwise $\beta_{ik} = 0$.
- Factors: $r = 2$, $r_{max} = 3$, loadings $\lambda_{ij} \sim N(0, 1)$ for $j = 1, 2$, $\lambda_{ij} = 0$ for $j = 3$; $\phi = \text{diag}(0.5, 0.7, 0.9)$.
- Spatial spillover intensity: $\rho_i = 0.2$ for $i = 1, ..., N/2$, $\rho_i = 0.6$ for $i = N/2 + 1, ..., N$.
- Sample sizes: $N \in \{10, 25, 50\}$, $T \in \{250, 500, 1000\}$.

Some simulation results

	latent factors		regressors			local dependence		
	RMSE	#factors	RMSE	sparsity	#regressors	RMSE	#groups	cos. similarity
				$N = 10$				
$T = 250$	0.222	2.095	0.032	0.985	1.003	0.091	2.466	0.979
$T = 500$	0.203	2.019	0.026	0.998	1.000	0.086	2.460	0.982
$T = 1000$	0.187	2.002	0.024	0.999	1.000	0.084	2.489	0.983
				$N = 25$				
$T = 250$	0.102	2.432	0.015	0.987	1.001	0.037	2.924	0.997
$T = 500$	0.099	2.396	0.013	0.998	1.000	0.034	2.568	0.997
$T = 1000$	0.098	2.318	0.012	0.999	1.000	0.033	2.367	0.997
				$N = 50$				
$T = 250$	0.118	2.674	0.010	0.985	1.011	0.024	3.809	0.999
$T = 500$	0.119	2.593	0.009	0.998	1.000	0.022	3.369	0.998
$T = 1000$	0.115	2.377	0.008	0.999	1.000	0.020	3.018	0.999

- As T increases, coefficients are estimated more accurately.
- Model selection (sparsity patterns, number of regressors and factors, group structure in local dependence) is fairly accurate for realistic T, N.

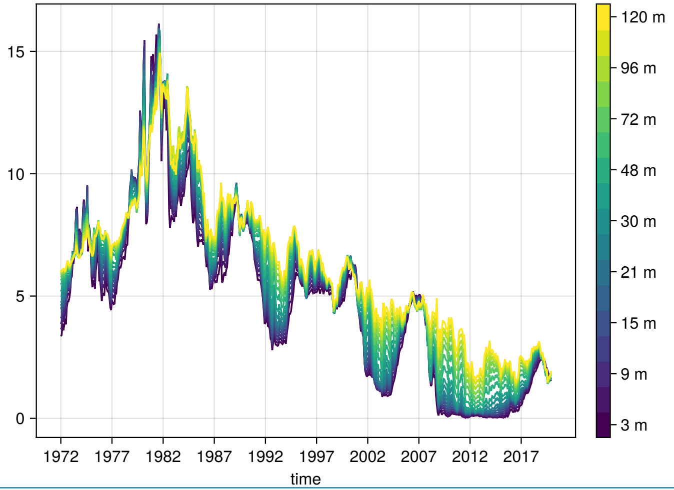
Outline

- Introduction
- Methodology
- Monte Carlo simulation
- **Empirical application: Predicting the yield curve**
- Conclusion

Data

- Monthly constant maturity zero-coupon Treasury yield curve from the data set of Liu and Wu (2020, JFE).
- 576 monthly observations, covering January 1972 – December 2019.
- Excess returns for 17 maturities, using 1 month as risk-free rate:
  - Short: 3, 6, 9, 12 months
  - Medium: 15, 18, 21, 24, 30, 36, 48 months
  - Long: 60, 72, 84, 96, 108, 120 months
- Regressors: IP growth, Real interest rate (Coroneo, Giannone, Modugno 2016; Joslin, Priebsch, Singleton 2014).
- Full model allows for heterogeneous but potentially sparse slopes, up to five factors, heterogeneous, grouped or constant local dependence; W with one direct neighbor structure.
- Analysis: Full-sample, rolling windows (20 years; 324 windows).

Yield data


15
10
5
0
1972
1977
1982
1987
1992
1997
2002
2007
2012
2017
time
120 m
96 m
72 m
48 m
30 m
21 m
15 m
9 m
3 m

Excess returns

- Levels close to nonstationary; use excess returns instead (Bianchi, Büchner, Tamoni 2020).
- $p_t^{(n)}$: Log price of zero coupon bond with maturity n at time t
- Log yield:

$$y_t^{(n)} = -\frac{1}{n} p_t^{(n)} \tag{2}$$

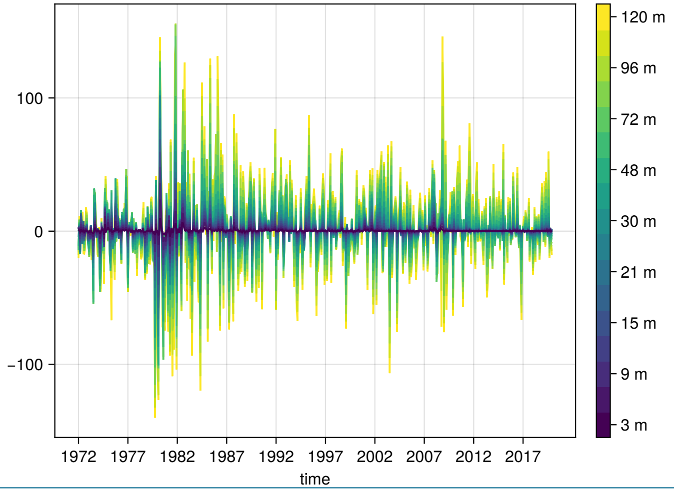
- Log holding period return (buy n-year bond at time t and sell it as $n-1$-year bond at $t+1$):

$$r_{t+1}^{(n)} = p_{t+1}^{(n-1)} - p_t^{(n)} \tag{3}$$

- Excess log return (our dependent variable):

$$rx_{t+1}^{(n)} = r_{t+1}^{(n)} - y_t^{(1)} \tag{4}$$

Excess return data



In-sample results

- Data-driven model selection via BIC.
- 5 factors selected for full sample.
- Macro variables play a minor role: real interest rate is deselected, some negative coefficients for IP growth.
- Local dependence:
  - Short:

3m	6m	9m	12m
0.9947	0.9947	0.9947	0.8461

  - Medium:

15m	18m	21m	24m	30m	36m	48m
0.7166	0.8600	0.9817	0.9000	0.9000	0.8699	0.8699

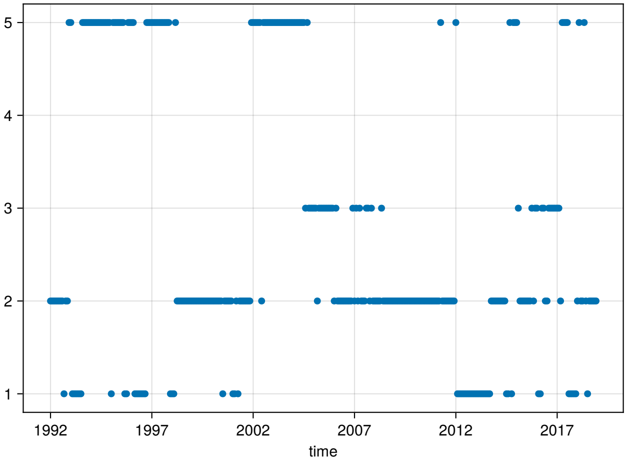
  - Long:

60m	72m	84m	96m	108m	120m
0.8545	0.7559	0.7827	0.7827	0.6184	0.6184

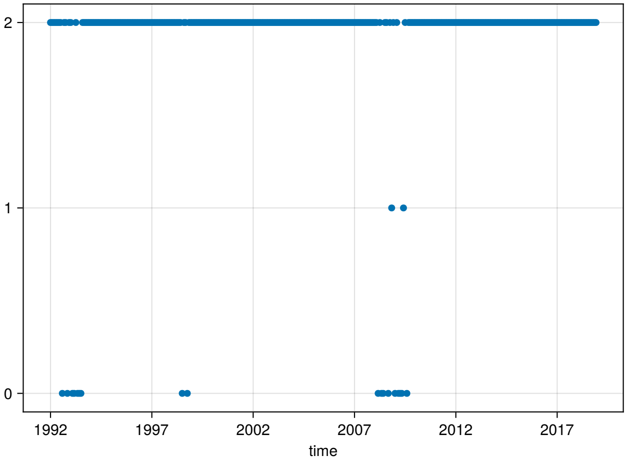
Rolling windows and out-of-sample prediction

- Re-estimation and penalty parameter choice in every window using 3y validation sample and mean absolute forecast error as criterion.
- Iterative out-of-sample forecasts for $h = 1, 2, .., 12$ months.
- VAR(3) forecasting model for IP growth and real interest rate.
- Benchmarks: Our model with only dynamic factors, expectation hypothesis (EH; unpredictable excess returns).
- Forecast evaluation using relative MAFEs and test for multi-horizon predictive ability.
- Evaluation period: January 1992 – December 2019.

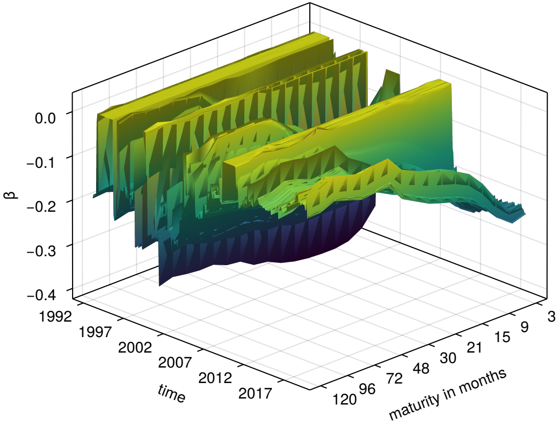
Rolling windows: number of factors



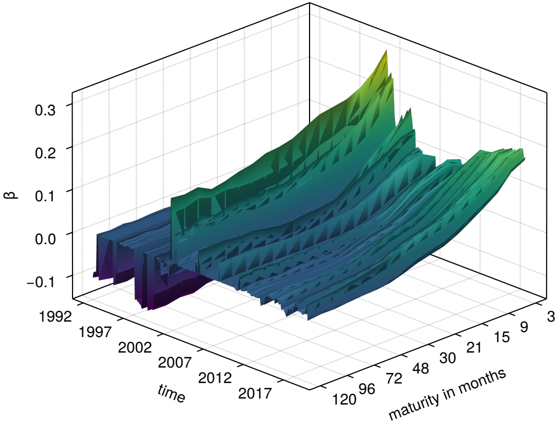
Rolling windows: number of regressors



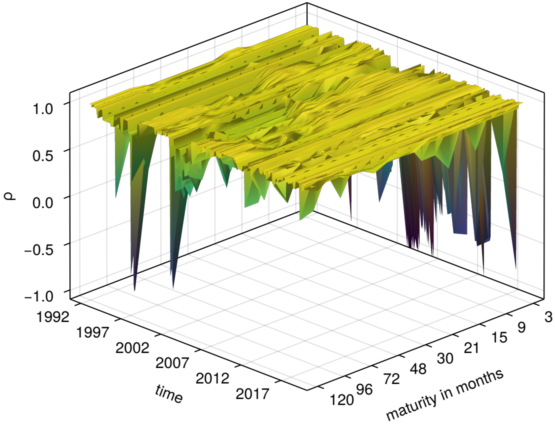
Rolling windows: slopes IP growth



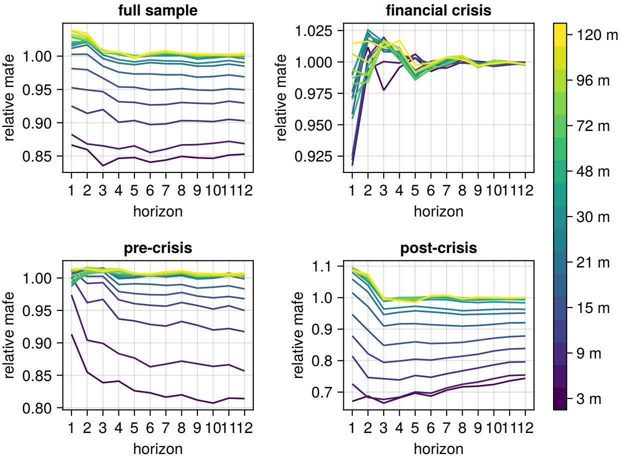
Rolling windows: slopes real interest rate



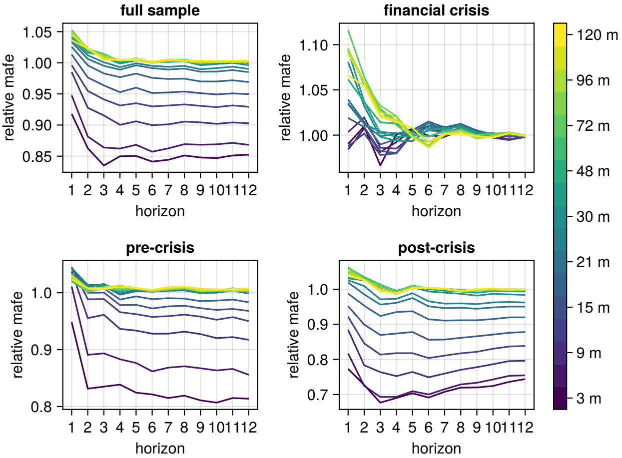
Rolling windows: local dependence



Forecast comparison: HSE-DFM-X vs. EH



Forecast comparison: HSE-DFM-X vs. factors-only



Multi-horizon forecast comparison: EH

- Test for multi-horizon superior predictive ability (Quaedvlieg, 2019).
- Test statistics (based on MAFE):

	3 m	6 m	9 m	12 m	15 m	18 m
full sample	7.98*	6.87*	6.13*	5.65*	4.94*	3.44*
pre-crisis	5.34*	4.08*	3.01*	2.35*	1.79*	0.75
financial crisis	1.04	1.75*	0.78	0.71	-0.03	-0.16
post-crisis	10.35*	9.07*	9.78*	9.64*	9.33*	7.97*
	21 m	24 m	30 m	36 m	48 m	60 m
full sample	1.33*	0.53	-0.83†	-1.24†	-1.52†	-1.76†
pre-crisis	-0.46	-0.69	-0.69	-0.73†	-0.88†	-1.08†
financial crisis	0.05	-0.02	-0.38	-0.22	0.48	0.56
post-crisis	4.13*	2.29*	-0.38	-1.33†	-2.24†	-2.32†
	72 m	84 m	96 m	108 m	120 m	
full sample	-1.99†	-1.98†	-2.26†	-2.57†	-3.15†	
pre-crisis	-1.27†	-1.29†	-1.66†	-1.93†	-2.02†	
financial crisis	0.23	0.22	0.00	-0.30	-0.84	
post-crisis	-2.21†	-2.09†	-2.03†	-2.06†	-2.55†	

*HSE-DFM-X outperforms EH ($\alpha = 10\%$), †EH outperforms HSE-DFM-X ($\alpha = 10\%$)

Multi-horizon forecast comparison: factors-only

- Test for multi-horizon superior predictive ability (Quaedvlieg, 2019).
- Test statistics (based on MAFE):

	3 m	6 m	9 m	12 m	15 m	18 m
full sample	7.89^*	6.83^*	6.02^*	5.55^*	4.83^*	3.18^*
pre-crisis	5.39^*	4.27^*	3.08^*	2.32^*	1.67^*	0.51
financial crisis	0.68	0.49	-0.02	0.46	-0.39	$-1.56^\dagger$
post-crisis	9.81^*	8.94^*	9.72^*	9.50^*	9.26^*	8.96^*
	21 m	**24 m**	**30 m**	**36 m**	**48 m**	**60 m**
full sample	1.07^*	0.18	$-1.44^\dagger$	$-1.89^\dagger$	$-2.24^\dagger$	$-2.94^\dagger$
pre-crisis	$-0.82^\dagger$	$-1.21^\dagger$	$-1.29^\dagger$	$-1.20^\dagger$	$-1.25^\dagger$	$-1.43^\dagger$
financial crisis	$-1.72^\dagger$	$-1.73^\dagger$	$-2.17^\dagger$	$-2.16^\dagger$	$-1.71^\dagger$	$-2.37^\dagger$
post-crisis	6.94^*	3.97^*	0.43	-0.68	$-1.62^\dagger$	$-2.21^\dagger$
	72 m	**84 m**	**96 m**	**108 m**	**120 m**	
full sample	$-2.92^\dagger$	$-2.40^\dagger$	$-2.37^\dagger$	$-2.41^\dagger$	$-2.40^\dagger$	
pre-crisis	$-1.50^\dagger$	$-1.41^\dagger$	$-1.68^\dagger$	$-1.87^\dagger$	$-1.92^\dagger$	
financial crisis	$-2.19^\dagger$	$-1.65^\dagger$	$-1.47^\dagger$	$-1.40^\dagger$	-1.26	
post-crisis	$-2.12^\dagger$	$-1.57^\dagger$	$-1.32^\dagger$	$-1.14^\dagger$	$-1.22^\dagger$	

*HSE-DFM-X outperforms factors-only model ($\alpha = 10\%$), †factors-only model outperforms HSE-DFM-X ($\alpha = 10\%$)

Conclusion

- Agnostic method for empirical analysis of panel data with regressors as well as dependencies over cross-section and time.
- Tailored algorithm to efficiently combine estimation and model selection. Works in simulations.
- Some new insights on the properties of treasury yields (excess returns):
  - Even after controlling for up to five latent factors, local dependence appears to play an important role.
  - Impact of macro variables on bond returns is time-varying.
  - Including model features in an adaptive way improves out-of-sample forecasting performance at the short end of the yield curve.

Thank you!