Yield Curve Modeling and Forecasting using Semiparametric Factor Dynamics

Wolfgang Karl Härdle
Piotr Majer

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics and Economics
Humboldt–Universität zu Berlin
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de

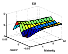
Yield Curve

Figure 1: Yield curves of Greece, Portugal, Spain, Italy; 199901-201011

Objectives

- Modeling the interest rate term structure using Dynamic Factor Model
  - Estimating and predicting factors and factor loadings
  - Understanding the dynamics of factor loadings
  - Impact of macroeconomy variables
  - Panel data analysis

- Forecasting the term structure
  - Forecasting evaluation against (dynamic) Nelson-Siegel Model



Statistical Challenges

- Time-varying high-dimensional object
- Dimension reduction: extraction of common factors
- Factors estimated non-parametrically
- Capturing dynamics by parametric time series models for factor loadings

Smooth in space and parametric in time

Economic Implications

- Understanding macroeconomic activity
- Conducting monetary policy
- Pricing and hedging interest rate products
- Portfolio allocation

Outline

1. Motivation ✓
2. Data
3. Dynamic Semiparametric Factor Model (DSFM)
4. Nelson-Siegel Model
5. Yield Curve Modeling and Forecasting
6. Conclusion

Data

- Zero-coupon bonds returns (observed monthly) - $Y_t \in \mathbb{R}^{11}$:
  - Italy (IT), Greece (GR), Portugal (PT), Spain (ES)
- Period covered: 199901-201203
- Dependence on Maturity - $X_{t,1} \in \mathbb{R}$
- Further explanatory variables - $X_{t,2} \in \mathbb{R}$
  - inflation rate (INF), manufacturing capacity utilization (CU)
  - unemployment rate (EMP), industrial production (IP), ΔGDP

Summary statistics

	Mean	Median	SD	Skewness	Kurtosis
1-year	2.8736	2.7843	1.1719	0.0523	2.1771
3-year	3.5394	3.4493	0.9836	0.4591	3.0203
5-year	3.9361	3.8228	0.8649	0.6655	3.5370
10-year	4.6039	4.4694	0.6801	0.5987	3.4534

Table 1: Statistical summary of the Italian 1, 3, 5, 10-year zero-coupon bond yields. Sample period 199901-201203; SD denotes Standard Deviation.

Notation

$$\underbrace{(X_{1,1}, Y_{1,1}), \ldots, (X_{J_1,1}, Y_{J_1,1})}_{t=1}, \ldots, \underbrace{(X_{1,T}, Y_{1,T}), \ldots, (X_{J_T,T}, Y_{J_T,T})}_{t=T},$$

where:

$X_{j,t} \in \mathbb{R}^d,\ Y_{j,t} \in \mathbb{R}$

T - the number of observed time periods (days)

J_t - the number of the observations in (day) t

$\mathrm{E}(Y_t|X_t) = F_t(X_t)$.

Quantify $F_t(X_t)$

Dynamic Semiparametric Factor Model

- Park et al. (2009), Fengler et al. (2007)

$$\mathsf{E}(Y_t|X_t) = \sum_{l=0}^{L} Z_{t,l}\, m_l(X_t) = Z_t^\top m(X_t) = Z_t^\top A^* \Psi(X_t)$$

$$\begin{aligned} Z_t &= (\mathbf{1}, Z_{t,1}, \ldots, Z_{t,L})^\top \text{low dim (stationary) time series} \\ m &= (m_0, m_1, \ldots, m_L)^\top, \text{tuple of functions} \\ \Psi(X_t) &= \{\psi_1(X_t), \ldots, \psi_K(X_t)\}^\top, \psi_k(x) \text{ space basis} \\ A^* &: (L+1) \times K \text{ coefficient matrix} \end{aligned}$$

B-Splines

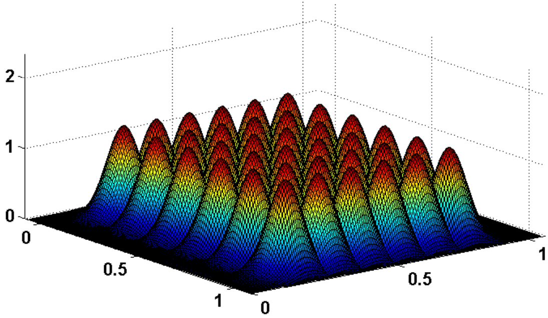


Figure 2: B-splines basis functions; order of B-splines: quadratic; number of knots: $6 \times 6 = 36$ ▸ B-Splines

Estimation

Recall

$$Y_{t,j} = \sum_{l=0}^{L} Z_{t,l} m_l(X_{t,j}) + \varepsilon_{t,j} = Z_t^\top A\psi(X_{t,j}) + \varepsilon_{t,j}$$

$\psi(x) = \{\psi_1(x), \ldots, \psi_K(x)\}^\top$ tensor B-spline basis

$$(\widehat{Z}_t, \widehat{A}) = \arg\min_{Z_t, A} \sum_{t=1}^{T} \sum_{j=1}^{J} \left\{ Y_{t,j} - Z_t^\top A\psi(X_{t,j}) \right\}^2$$

- Selection of L by explained variance

Implementation

- Explained variance

$$EV(L) = 1 - \frac{\sum_{t=1}^{T} \sum_{j=1}^{J} \left\{ Y_{t,j} - \sum_{l=0}^{L} Z_{t,l} m_l(X_{t,j}) \right\}^2}{\sum_{t=1}^{T} \sum_{j=1}^{J} \left\{ Y_{t,j} - \bar{Y} \right\}^2}$$

Statistical Inference

- Asymptotically negligible difference between $\widehat{Z}_t$ and Z_t
- Time series modeling of $\widehat{Z}_t$

Panel DSFM

$$Y^i_{t,j} = \sum_{l=0}^{L} Z^i_{t,l} \widetilde{m}_l(X_{t,j}) + \varepsilon^i_{t,j}, \quad 1 \leq j \leq J,\ 1 \leq t \leq T,$$

$1 \leq i \leq I$ - country index

- Domestics specific effects and time evolution captured by Z^i_t

$\widetilde{m}_l$ - common factors

$$(\widehat{Z}^1_t, \dots, \widehat{Z}^I_t, \widehat{A}) = \arg \min_{Z^1_t, \dots, Z^I_t, A} \sum_{i=1}^{I} \sum_{t=1}^{T} \sum_{j=1}^{J} \left\{ Y^i_{t,j} - Z^{i\top}_t A \psi(X_{t,j}) \right\}^2 \tag{1}$$

Dynamic Nelson-Siegel Model

$$
\begin{aligned}
Y_{t,j} =& L_t + S_t \left\{ \frac{1-\exp(-\lambda X_{t,j})}{\lambda X_{t,j}} \right\} + C_t \left\{ \frac{1-\exp(-\lambda X_{t,j})}{\lambda X_{t,j}} - \exp(-\lambda X_{t,j}) \right\} + \varepsilon_{t,j} \\
=& Z_t^\top m(X_{t,j}) + \varepsilon_{t,j},
\end{aligned}
$$

- $Z_t = (L_t, S_t, C_t)^\top$ loadings: L_t - level, S_t - slope, C_t - curvature
- $m(\cdot) = (\mathbf{1}, \frac{1-\exp(-\lambda(\cdot))}{\lambda(\cdot)}, \frac{1-\exp(-\lambda(\cdot))}{\lambda(\cdot)} - \exp(-\lambda(\cdot)))$ common factors

Dynamic Nelson-Siegel Model

- Z_t estimated day-by-day by OLS (for fixed λ)
- VAR(1) model for $Z_t \in \mathbb{R}^3$

$$Z_t = \mu + \mathcal{A} Z_{t-1} + \eta_t, \qquad (2)$$

- Ability to reproduce (historical) yield curve properties
  - mean reversion
  - basket of different shapes
- Parsimonious structure

Yield Curve Modeling

- DSFM approaches
  - yield curves modeled domestically (DSFM)
  - panel term structure model (PDSFM)
- Explanatory variables $X_{t,j}$
  - time to maturity
  - inflation Rate

Explained Variance

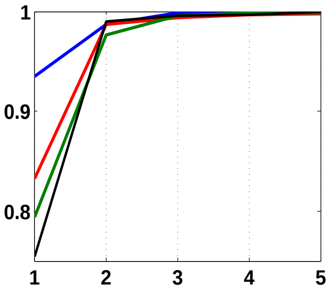


Figure 3: Explained variance for the first 5 factors for domestic DSFM: PT, GR, ES and IT (left panel) and PDSFM (right panel).

In-Sample Fit

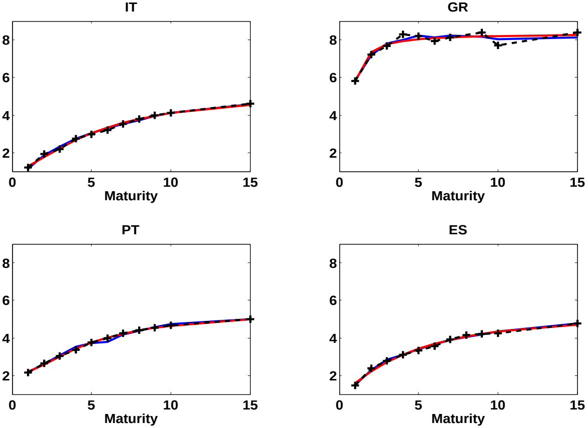


Figure 4: The term structure of interest rates (dotted black) observed on 20100331, DSFM (blue) and the Nelson-Siegel fitted data.

DSFM Factor $\widehat{m}_1$

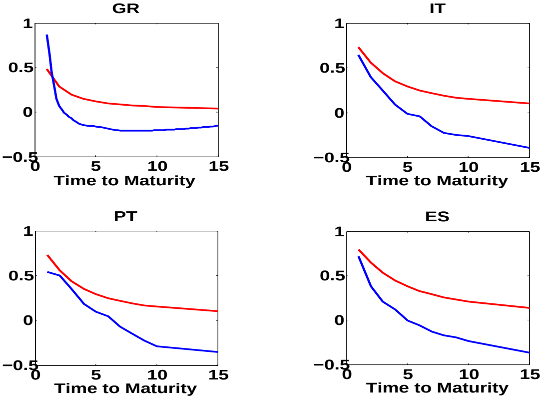


Figure 5: Estimated $\widehat{m}_1$ depending on time to maturity [Years] using domestic DSFM and Nelson-Sigel slope factor with $\lambda_{GR} = 0.049$, $\lambda_{IT} = 0.127$, $\lambda_{PT} = 0.109$ and $\lambda_{ES} = 0.174$, respectively.

DSFM Factor $\widehat{m}_2$

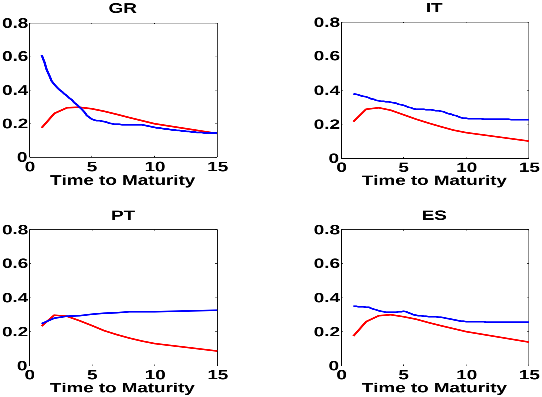


Figure 6: Estimated $\widehat{m}_2$ depending on time to maturity [Years] using domestic DSFM and Nelson-Sigel slope factor with $\lambda_{GR} = 0.049$, $\lambda_{IT} = 0.127$, $\lambda_{PT} = 0.109$ and $\lambda_{ES} = 0.174$, respectively.

Estimated Factor Loadings, $\widehat{Z}_t$

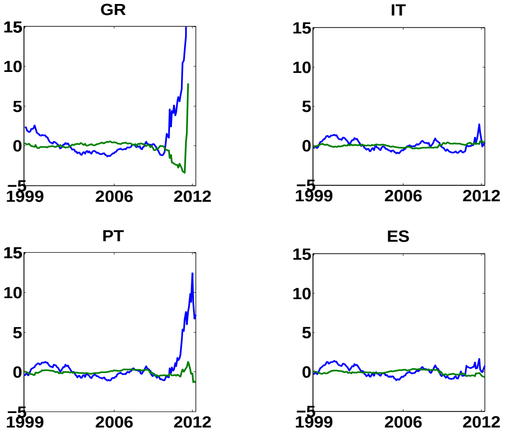


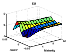
Figure 7: Estimated factor loadings $\widehat{Z}_t$ of the yield curve over whole sample using domestic DSFM; blue line corresponds to $\widehat{Z}_{t,1}$, green - $\widehat{Z}_{t,2}$.

$\widehat{Z}_t$ Properties

- Unit root hypothesis is not rejected at significance 5%
- Highly persistent processes
- Stationarity of $\Delta\widehat{Z}_t \equiv \widehat{Z}_t - \widehat{Z}_{t-1}$ not rejected
- $VAR(p)$ models for $\Delta\widehat{Z}_t \in \mathbb{R}^2$

$$\Delta\widehat{Z}_t = c + \alpha_1 \Delta\widehat{Z}_{t-1} + \ldots + \alpha_p \widehat{Z}_{t-p} + \varepsilon_t$$

- Model selection based on HQ and SC information criteria



Including Further Explanatory Variables

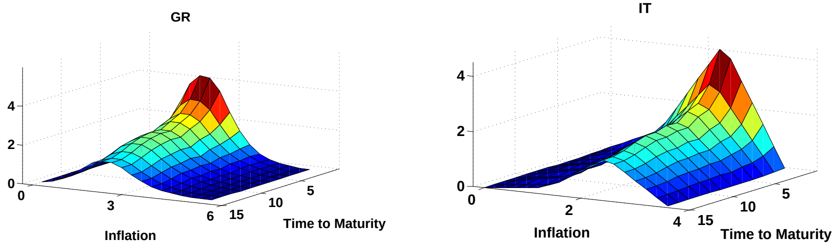


Figure 8: Estimated first factor depending on time to maturity and inflation rate for Greece and Italy

Including Further Explanatory Variables

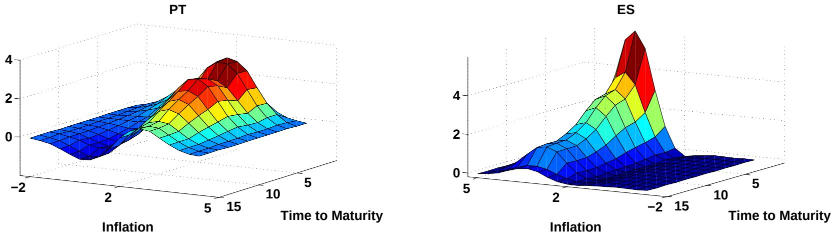


Figure 9: Estimated first factor depending on time to maturity and inflation rate for Portugal and Spain

Including Further Explanatory Variables

- Impact of the Inflation Rate
  - stronger on short rates
  - similar across countries
  - peak at the 2% rate - central bank target

Estimated PDSFM Factors

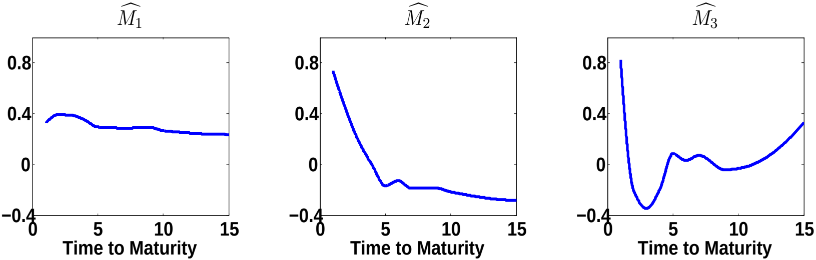


Figure 10: Estimated factors $\widehat{M}$ depending on time to maturity (years), using PDSFM with 3 factors

RMSE

	GR	IT	PT	ES
NS	0.5600	0.0685	0.2009	0.0636
DSFM	0.2886	0.0872	0.4195	0.0695
DSFM(INF)	0.6813	0.1550	0.5520	0.2110
PDSFM	0.6810	0.1474	0.4516	0.1434

Table 2: RMSE derived by Nelson-Siegel model (NS), domestic DSFM, DSFM with inflation rate and PDSFM (3 factors) in dependence on time to maturity.

Factors and macroeconomic fundamentals

- Contemporaneous correlation between factor loadings and macroeconomic variables:

$$\Delta \widehat{Z}_t = C + \beta_1 INF_t + \beta_2 CU_t + \beta_3 EMP_t + \beta_4 IP_t + \beta_5 \Delta GDP_t + \varepsilon_t \quad (3)$$

- $\widehat{Z}_{t,1}$ driven by INF, ΔGDP and IP at significance level 5%
- $\widehat{Z}_{t,2}$ can not be explained by (3)

Forecasting Setup

- Period: 200701-201203
- Rolling windows shifted over 1 month grids
- Forecasting horizon: 12 months
- Forecasting approaches
  - domestic DSFM ($L = 2$)
  - dynamic NS

Forecasting

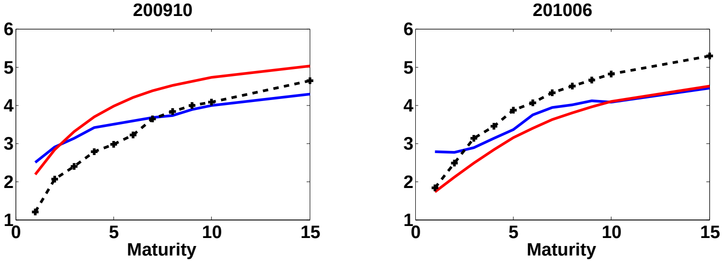


Figure 11: Term structure of interest rates (dotted black) observed on 20091030 (left) and 20100630 (right) for Italy with the DSFM (blue) and the dynamic Nelson-Siegel (red) forecasts.

RMSPE

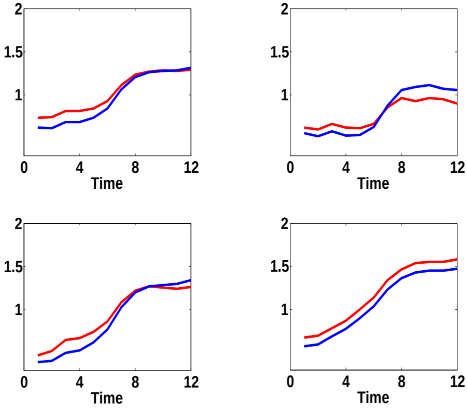


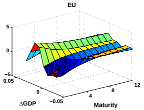
Figure 12: Root Mean Squared Prediction Error DSFM (2 factors) and NS for whole yield curve (top left), 2 year rate (top right), 7 year rate (bottom left) and 15 year rate (bottom right).

Conclusion

- Two factors are sufficient to model the term structure of interest rates domestically
- Panel term structure required 3 factors
- Estimated factor loadings are unit root, highly persistent processes
- Macroeconomic variables included
- Yield curves are modeled and forecasted successfully

Yield Curve Modeling and Forecasting using Semiparametric Factor Dynamics

Wolfgang Karl Härdle
Piotr Majer



Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics and Economics
Humboldt–Universität zu Berlin
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de

References

Diebold, F., Li, C.
Forecasting the Term Structure of Government Bond Yields
Journal of Econometrics, 130(2): 337–364, 2006

Fengler, M., Härdle, W. and Mammen, E.
A Dynamic Semiparametric Factor Model for Implied Volatility String Dynamics
Journal of Financial Econometrics, 5(2): 189–218, 2007

Hodrick, R., Prescott, E.
Postwar U.S. Business Cycles: An Empirical Investigation
Journal of Money, Credit, and Banking, 29(1): 1-16, 1997

References

Nelson, C. and Siegel A.
Parsimonious Modeling of Yield Curves
The Journal of Business, 60(4), 473–489, 1987

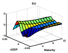
Park, B., Mammen, E., Härdle, W. and Borak, S.
Time Series Modelling with Semiparametric Factor Dynamics
J. Amer. Stat. Assoc., 104(485): 284-298, 2009

B-Splines

▸ B-Splines

Univariate **B-spline** basis $\Psi = \{\psi_1(X), \ldots, \psi_K(X)\}^\top$ is a series of $\psi_k(X)$ functions defined by $x_0 \leq x_2 \leq \ldots \leq x_{K-1}$, K knots and order p, i.e. for $p = 2$ (quadratic)

$$\psi_j(x) = \begin{cases} \frac{1}{2}(x - x_j)^2 & \text{if } x_j \leq x < x_{j+1} \\ \frac{1}{2} - (x - x_{j+1})^2 + (x - x_{j+1}) & \text{if } x_{j+1} \leq x < x_{j+2} \\ \frac{1}{2}\left\{1 - (x - x_{j+2})^2\right\} & \text{if } x_j \leq x < x_{j+1} \\ x & \text{otherwise} \end{cases}$$



B-Splines

▸ B-Splines

- Knots K and order p has to be specified in advance (EV criterion); K corresponds to bandwidth
- In higher dimensions, for $dim(X) = d > 1$

$$\Psi = \{\psi_1(X_1), \ldots, \psi_{K_1}(X_1)\} \times \ldots \times \{\psi_1(X_d), \ldots, \psi_{K_d}(X_d)\}$$

- Flexible and computationally efficient approach to capture various spatial structures