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), $\Delta$GDP

Yield Curve Modeling and Forecasting
## Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-year</td>
<td>2.8736</td>
<td>2.7843</td>
<td>1.1719</td>
<td>0.0523</td>
<td>2.1771</td>
</tr>
<tr>
<td>3-year</td>
<td>3.5394</td>
<td>3.4493</td>
<td>0.9836</td>
<td>0.4591</td>
<td>3.0203</td>
</tr>
<tr>
<td>5-year</td>
<td>3.9361</td>
<td>3.8228</td>
<td>0.8649</td>
<td>0.6655</td>
<td>3.5370</td>
</tr>
<tr>
<td>10-year</td>
<td>4.6039</td>
<td>4.4694</td>
<td>0.6801</td>
<td>0.5987</td>
<td>3.4534</td>
</tr>
</tbody>
</table>

**Table 1:** Statistical summary of the Italian 1, 3, 5, 10-year zero-coupon bond yields. Sample period 199901-201203; SD denotes Standard Deviation.
Dynamic Semiparametric Factor Model (DSFM)

Notation

\[(X_{1,1}, Y_{1,1}), \ldots, (X_{J_1,1}, Y_{J_1,1}), \ldots, (X_{1,T}, Y_{1,T}), \ldots, (X_{J_T,T}, Y_{J_T,T}),\]

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) \(t\)
- \(E(Y_t|X_t) = F_t(X_t)\).

Quantify \(F_t(X_t)\)
Dynamic Semiparametric Factor Model

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

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

\[ Z_t = (1, Z_{t,1}, \ldots, Z_{t,L})^\top \text{ low dim (stationary) time series} \]

\[ m = (m_0, m_1, \ldots, m_L)^\top, \text{ tuple of functions} \]

\[ \Psi(X_t) = \{ \psi_1(X_t), \ldots, \psi_K(X_t) \}^\top, \psi_k(x) \text{ space basis} \]

\[ A^* : (L + 1) \times K \text{ coefficient matrix} \]

Yield Curve Modeling and Forecasting
B-Splines

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

Recall

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

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

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

- Selection of \( L \) by explained variance

Yield Curve Modeling and Forecasting
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 \( \hat{Z}_t \) and \( Z_t \)
- Time series modeling of \( \hat{Z}_t \)
Panel DSFM

\[ Y_{t,j}^i = \sum_{l=0}^{L} Z_{t,i}^l \tilde{m}_l(X_{t,j}) + \varepsilon_{t,j}^i, \quad 1 \leq j \leq J, \quad 1 \leq t \leq T, \]

1 \leq i \leq I - country index

- Domestic specific effects and time evolution captured by \( Z_{t,i}^l \)
- \( \tilde{m}_l \) - common factors

\[
(\hat{Z}_t^1, \ldots, \hat{Z}_t^l, \hat{A}) = \arg \min_{Z_t^1, \ldots, Z_t^l, A} \sum_{i=1}^{I} \sum_{t=1}^{T} \sum_{j=1}^{J} \left\{ Y_{t,j}^i - Z_t^i \top A \psi(X_{t,j}) \right\}^2
\]

(1)
Dynamic Nelson-Siegel Model

\[ Y_{t,j} = L_t + S_t \left\{ \frac{1 - \exp(-\lambda X_{t,j})}{\lambda X_{t,j}} \right\} + C_t \left\{ \frac{1 - \exp(-\lambda X_{t,j})}{\lambda X_{t,j}} - \exp(-\lambda X_{t,j}) \right\} + \varepsilon_{t,j} \]

\[ = Z_t^\top m(X_{t,j}) + \varepsilon_{t,j}, \]

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

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

$$Z_t = \mu + A Z_{t-1} + \eta_t,$$

- 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 $\hat{m}_1$

Figure 5: Estimated $\hat{m}_1$ 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 $\hat{m}_2$

Figure 6: Estimated $\hat{m}_2$ 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, $\hat{Z}_t$

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

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

\[
\Delta \hat{Z}_t = c + \alpha_1 \Delta \hat{Z}_{t-1} + \ldots + \alpha_p \hat{Z}_{t-p} + \varepsilon_t
\]

- 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 $\hat{M}$ depending on time to maturity (years), using PDSFM with 3 factors
### RMSE

<table>
<thead>
<tr>
<th></th>
<th>GR</th>
<th>IT</th>
<th>PT</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS</td>
<td>0.5600</td>
<td>0.0685</td>
<td>0.2009</td>
<td>0.0636</td>
</tr>
<tr>
<td>DSFM</td>
<td>0.2886</td>
<td>0.0872</td>
<td>0.4195</td>
<td>0.0695</td>
</tr>
<tr>
<td>DSFM(INF)</td>
<td>0.6813</td>
<td>0.1550</td>
<td>0.5520</td>
<td>0.2110</td>
</tr>
<tr>
<td>PDSFM</td>
<td>0.6810</td>
<td>0.1474</td>
<td>0.4516</td>
<td>0.1434</td>
</tr>
</tbody>
</table>

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:

\[ \Delta \hat{Z}_t = C + \beta_1 \text{INF}_t + \beta_2 \text{CU}_t + \beta_3 \text{EMP}_t + \beta_4 \text{IP}_t + \beta_5 \Delta \text{GDP}_t + \epsilon_t \] (3)

- \( \hat{Z}_{t,1} \) driven by \( \text{INF} \), \( \Delta \text{GDP} \) and \( \text{IP} \) at significance level 5%
- \( \hat{Z}_{t,2} \) 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
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.
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

Fengler, M., Härdle, W. and Mammen, E.
A Dynamic Semiparametric Factor Model for Implied Volatility String Dynamics

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

Park, B., Mammen, E., Härdle, W. and Borak, S.
Time Series Modelling with Semiparametric Factor Dynamics
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

- Knots $K$ and order $p$ have 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.