‘Forecasting multiple time series using quasi-randomized neural networks - Example of a Dynamic Nelson Siegel model’, JIRF 2017

Thierry Moudiki (ISFA, Université Lyon 1)

May 18th, 2017

What?

A model for forecasting multiple correlated time series, using Random Vector Functional Link (RVFL) neural networks

Plan

- RVFL models description
- The model described in the paper
- Example of a Dynamic Nelson Siegel (DNS) model for spot interest rates

Plan

Plan

- **RVFL models description**
- The model described in the paper
- Example of a Dynamic Nelson Siegel (DNS) model for spot interest rates

RVFL models description

$y \in \mathbb{R}^n$, to be explained by $X^{(j)} \in \mathbb{R}^n$, $j \in \{1, \ldots, p\}$

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_i^{(j)} + \sum_{l=1}^{L} \gamma_l g\left(\sum_{j=1}^{p} W^{(j,l)} X_i^{(j)}\right) + \epsilon_i$$

for $i \in \{1, \ldots, n\}$. With:

- g: *activation function*
- L: number of **nodes** in the **hidden layer**
- $W^{(j,l)}$ elements of the hidden layer, **simulated** from uniform distribution
- β_j and γ_l: **learned** from the observed data y and $X^{(j)}$, $j \in \{1, \ldots, p\}$
- ϵ_i: **residual differences** between output and the RVFL model

RVFL models description

- **Result**: *RVFL networks are universal approximators of a continuous function on a bounded finite dimensional set [. . .]*. From Igelnik and Pao (1995). *Stochastic choice of basis functions in adaptative function approximation and the functional-link net*.
- **Danger**: Overfitting
- **Overfitting**? Increasing the number nodes $L \rightarrow$ in-sample bias decreases $\rightarrow$ out-of-sample error increases

RVFL models description

Overfitting?

From "An Introduction to Statistical Learning" (Springer, 2013), authors: G. James, D. Witten, T. Hastie and R. Tibshirani

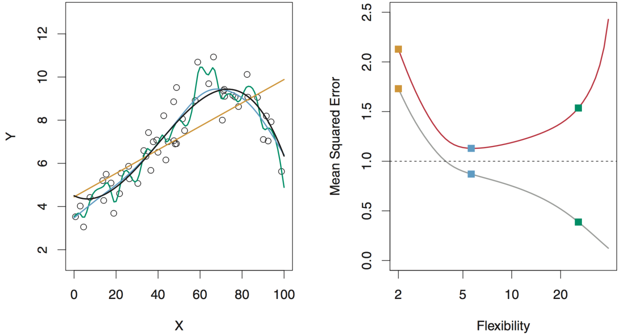


FIGURE 2.9. Left: *Data simulated from f, shown in black. Three estimates of f are shown: the linear regression line (orange curve), and two smoothing spline fits (blue and green curves).* Right: *Training MSE (grey curve), test MSE (red curve), and minimum possible test MSE over all methods (dashed line). Squares represent the training and test MSEs for the three fits shown in the left-hand panel.*

Plan

- RVFL models description
- **The model described in the paper**
- Example of a Dynamic Nelson Siegel (DNS) model for spot interest rates

The model described in the paper

- $p \in \mathbb{N}^*$ time series $(X_t^{(j)})_t, j = 1, \ldots, p$, observed at $n \in \mathbb{N}^*$ discrete dates
- **aim**: simultaneous forecasts of the p time series at time $n + h$, $h \in \mathbb{N}^*$
- **output variables** for RVFL regression:

$$Y^{(j)} = \left(X_n^{(j)}, \ldots, X_{k+1}^{(j)}\right)^T$$

- $X_n^{(j)}$: most contamporaneous value of j^{th} time series
- output variables stored in

$$\mathbf{Y} \in \mathbb{R}^{(n-k) \times p}$$

- predictors (lags) stored in:

$$\mathbf{X} \in \mathbb{R}^{(n-k) \times (k \times p)}$$

The model described in the paper

- $\mathbf{X}$ consists in p blocks of k lags, for each one of the observed p time series.
- j_0^{th} block of $\mathbf{X}$, for $j_0 \in \{1, \ldots, p\}$ contains k columns:

$$\left(X_{n-i}^{(j_0)}, \ldots, X_{k+1-i}^{(j_0)}\right)^T \tag{1}$$

with $i \in \{1, \ldots, k\}$

- possibility to add other regressors, such as dummy variables, indicators of special events
- **Multitask learning**: the predictors are shared by the p series

The model described in the paper

In the paper:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_i^{(j)} + \sum_{l=1}^{L} \gamma_l g \left(\sum_{j=1}^{p} W^{(j,l)} X_i^{(j)} \right) + \epsilon_i$$

- y_i = contamporaneous values of a given series
- $X_i^{(j)}$ = lags
- g: *activation function* $= x \mapsto max(x, 0)$
- **Hidden layer** elements $W^{(j,l)}$: **low discrepancy** sequence
- **Discrepancy** of sequence of N points in $V \in [0,1)^d$:

$$sup_{V \in [0,1)^d} \left| \frac{number\ of\ points\ in\ V}{N} - v(V) \right|$$

  where $v(V)$ = volume of V: **how well the points are dispersed within V?**.
- **Additional constraints**: $||\beta||^2 \leq u$ and $||\gamma||^2 \leq v$, to reduce the risk of overfitting

The model described in the paper

Solving for the regression parameters

- We obtain the Lagrangian (one Lagrange multiplier λ_i per additional constraint):

$$\mathcal{L}(\mathbf{X}; \beta, \gamma) = ||y - \mathbf{X}\beta - \Phi(\mathbf{X})\gamma||^2 + \lambda_1 \beta^T \beta + \lambda_2 \gamma^T \gamma$$

with $\mathbf{X}$ = lagged predictors, $\Phi(\mathbf{X})$ = transformed lagged predictors

- First derivatives according to β and γ lead to:

$$\begin{pmatrix} \mathbf{X}^T\mathbf{X} + \lambda_1 I_{k \times p} & \mathbf{X}^T\Phi(\mathbf{X}) \\ \Phi(\mathbf{X})^T\mathbf{X} & \Phi(\mathbf{X})^T\Phi(\mathbf{X}) + \lambda_2 I_L \end{pmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T y \\ \Phi(\mathbf{X})^T y \end{pmatrix}$$

The model described in the paper

- We denote:

$$A = \begin{pmatrix} \mathbf{X}^T\mathbf{X} + \lambda_1 I_{k\times p} & \mathbf{X}^T\Phi(\mathbf{X}) \\ \Phi(\mathbf{X})^T\mathbf{X} & \Phi(\mathbf{X})^T\Phi(\mathbf{X}) + \lambda_2 I_L \end{pmatrix} =: \begin{pmatrix} B & C^T \\ C & D \end{pmatrix}$$

and $S = D - CB^+C^T$. Using algorithm for **blockwise matrix inversion**:

$$A^+ = \begin{pmatrix} B^+ + B^+C^TS^+CB^+ & -B^+C^TS^+ \\ -S^+CB^+ & S^+ \end{pmatrix} =: \begin{pmatrix} A_1^+ & A_2^+ \\ A_3^+ & A_4^+ \end{pmatrix}$$

S^+ and B^+ Moore-Penrose pseudo-inverse of S and B.

- **Remark:** y is typically centered, **X** standardized, β_0 not constrained

The model described in the paper

- Choosing λ_1 and λ_2, L? $\rightarrow$ **Rolling origin forecasting** on grid of values
- model learning on **training** set –> **forecasting** on test set –> **change the origin** –> **repeat** –> average errors

224 *I. Kaastra, M. Boyd / Neurocomputing 10 (1996) 215–236*

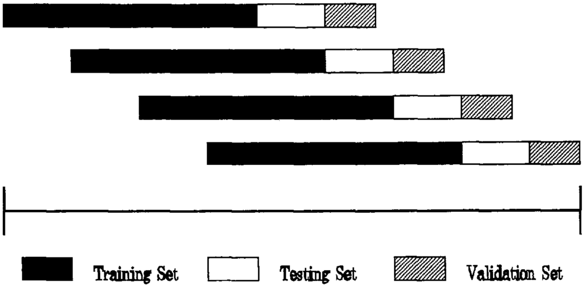


Fig. 3. Walk-forward sliding windows testing routine.

Plan

Plan

- RVFL models description
- The model described in the paper
- **Example of a Dynamic Nelson Siegel (DNS) model for spot interest rates**

Example of a Dynamic Nelson Siegel (DNS) model for spot interest rates

The DNS model for spot rates

$$y_t(\tau) = \alpha_{1,t} + \alpha_{2,t}\left(\frac{1 - e^{-\tau/\mu}}{e^{-\tau/\mu}}\right) + \alpha_{3,t}\left(\frac{1 - e^{-\tau/\mu}}{e^{-\tau/\mu}} - e^{-\tau/\mu}\right)$$

with

- τ = time to maturity and t = observation date.
- μ control the position of the hump, used in the cross-validation
- $\alpha_{i,t}, i = 1, \ldots, 3$: the 3 regression **parameters to be <u>estimated</u> using market data + forecast simultaneously**

Example of a Dynamic Nelson Siegel (DNS) model for spot interest rates

Typical estimation of $\alpha_{i,t}, i = 1, \ldots, 3$ in DNS

- Observed market spot rates $y_t^M(\tau)$ at time t for maturity τ
- Fix parameter μ
- Solve for $\alpha_{i,t}, i = 1, \ldots, 3$ (ordinary least squares)

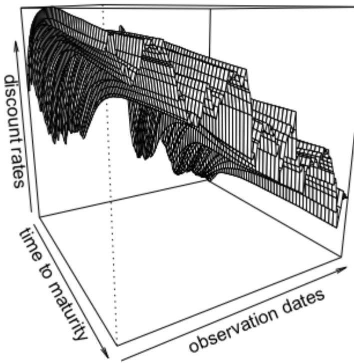
$$\begin{pmatrix} 1 & \frac{1-e^{-\tau_1/\mu}}{e^{-\tau_1/\mu}} & \frac{1-e^{-\tau_1/\mu}}{e^{-\tau_1/\mu}} - e^{-\tau_1/\mu} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\tau_m/\mu}}{e^{-\tau_m/\mu}} & \frac{1-e^{-\tau_m/\mu}}{e^{-\tau_m/\mu}} - e^{-\tau_m/\mu} \end{pmatrix} \begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \end{pmatrix} = \begin{pmatrix} y_t^M(\tau_1) \\ \vdots \\ y_t^M(\tau_m) \end{pmatrix}$$

- Obtain three time series of parameters
- Forecast the three time series

Example of a Dynamic Nelson Siegel (DNS) model for spot interest rates

The market data $y_t^M(\tau)$

- **Spot zero rates** data from BundesBank website
- Observed from 2002 to 2015, on a **monthly** frequency



Example of a Dynamic Nelson Siegel (DNS) model for spot interest rates

- **Summary** of the spot zero rates data (in %)

Maturity	Min	Median	Max
1	(0.116)	2.045	5.356
5	0.170	2.863	5.146
15	0.711	3.954	5.758
30	0.805	3.962	5.784
50	0.749	3.630	5.467

- **Aim**: benchmarking RVFL against competitor models for forecasting $\alpha_{i,t}, i = 1, \ldots, 3$: VAR and ARIMA
- **How**: Rolling origin forecasting = **training** the model –> **forecasting** the spot discount rates –> **change the origin** –> **repeat**

Example of a Dynamic Nelson Siegel (DNS) model for spot interest rates

Benchmark models

- VAR(p), p = number of lags, chosen by cross validation:
  Example on 2 variables and 1 lag

  $$\alpha_{1,t} = c_1 + \phi_{11,1}\alpha_{1,t-1} + \phi_{12,1}\alpha_{2,t-1} + e_{1,t}$$

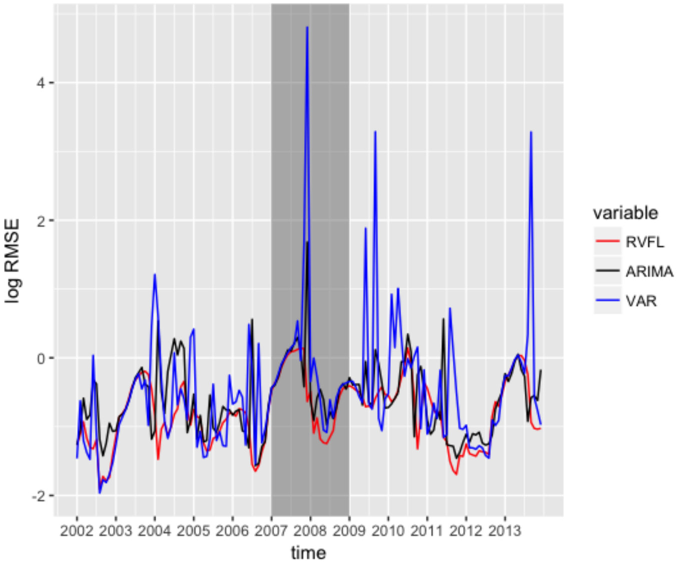
  $$\alpha_{2,t} = c_2 + \phi_{21,1}\alpha_{1,t-1} + \phi_{22,1}\alpha_{2,t-1} + e_{2,t}$$

  applied **simultaneously** to each series $\alpha_{i,t}, i = 1, \ldots, 3$
- ARIMA(p, d, q); B lag operator: $B\alpha_{i,t} = \alpha_{i,t-1}$

  $$(1 - \phi_1 B - \ldots \phi_p B^p)(1 - B)^d \alpha_{i,t} = c + (1 - \theta_1 B - \ldots \theta_q B^q)e_t$$

  applied **seperately** to each series $\alpha_{i,t}, i = 1, \ldots, 3$

Example of a Dynamic Nelson Siegel (DNS) model for spot interest rates



Example of a Dynamic Nelson Siegel (DNS) model for spot interest rates

Comparison of results

Method	Min	Median	Max
RVFL	0.1487	**0.4491**	**1.1535**
ARIMA	0.2089	0.5187	5.3798
VAR	**0.1402**	0.5549	122.22

Example of a Dynamic Nelson Siegel (DNS) model for spot interest rates

Stressed forecast curves