Weierstrass Institute for
Applied Analysis and Stochastics

Leibniz
Association

Manifold-based time series forecasting

Vladimir Spokoiny

(joint work with N. Puchkin and A. Timofeev)

Mohrenstrasse 39 · 10117 Berlin · Germany · Tel. +49 30 20372 0 · www.wias-berlin.de

15.07.2021

A learner observes $Y_1, \dots, Y_T$ generated from a model

$$Y_t = X_t + \varepsilon_t, \quad 1 \leqslant t \leqslant T,$$

where

- $\{X_t : 1 \leqslant t \leqslant T\} \subset \mathbb{R}^D$ is a hidden Markov chain on a low-dimensional manifold $\mathcal{M}^*$, dim$(\mathcal{M}^*) = d \ll D$
- $\varepsilon_1, \dots, \varepsilon_T$ are independent zero-mean innovations, $\mathbb{E}(\varepsilon_t | X_t) = 0$ for all t from 1 to T

A learner observes $Y_1, \dots, Y_T$ generated from a model

$$Y_t = X_t + \varepsilon_t, \quad 1 \leqslant t \leqslant T,$$

where

- $\{X_t : 1 \leqslant t \leqslant T\} \subset \mathbb{R}^D$ is a hidden Markov chain on a low-dimensional manifold $\mathcal{M}^*$, dim$(\mathcal{M}^*) = d \ll D$
- $\varepsilon_1, \dots, \varepsilon_T$ are independent zero-mean innovations, $\mathbb{E}(\varepsilon_t | X_t) = 0$ for all t from 1 to T

Goal: predict Y_{T+1}

Let $g : \mathbb{R}^d \to \mathbb{R}^p$ be a smooth function and let $\mathbf{\Phi}$ be a $(p \times d)$-matrix.

Central subspace model:

$$Z_t = g(\mathbf{\Phi}^T Z_{t-1}) + \xi_t, \quad 1 \leqslant t \leqslant T$$

Take

$$X_t = (Z_{t-1}, g(\mathbf{\Phi}^T Z_{t-1})) \in \mathbb{R}^{2p}, \quad Y_t = (Z_{t-1}, Z_t), \quad \varepsilon_t = (0, \xi_t).$$

Then $Y_t = X_t + \varepsilon_t$ and X_t lies on the graph of $g \circ \mathbf{\Phi}^T$ which is a d-dimensional submanifold in $\mathbb{R}^D$ with $D = 2p$

A natural extension of the central subspace model is

$$Z_t = g(f(Z_{t-1})) + \xi_t, \quad 1 \leqslant t \leqslant T,$$

where $g : \mathbb{R}^d \to \mathbb{R}^D$ and $f : \mathbb{R}^d \to \mathbb{R}^D$ are smooth functions

Consider $X_t = (Z_{t-1}, g(f(Z_{t-1}))) \in \mathbb{R}^{2p}$, $Y_t = (Z_{t-1}, Z_t)$, $\varepsilon_t = (0, \xi_t)$.

Then X_t lies on the graph of $g \circ f$, which is a d-dimensional submanifold in $\mathbb{R}^D$ with $D = 2p$.

A standard univariate autoregressive model of order τ is given by

$$Z_t = \sum_{i=1}^{\tau} a_i Z_{t-i} + \xi_t, \quad 1 \leqslant t \leqslant T.$$

Fix $D > \tau$ and apply a sliding window technique:
$Y_t = (Z_t, \dots, Z_{t-D+1}) \in \mathbb{R}^D$.

Then the autoregressive model can be rewritten in the form
$Y_t = \boldsymbol{A} Y_{t-1} + \varepsilon_t$, where rank$(\boldsymbol{A}) < D$, $\varepsilon_t = (\xi_t, 0, \dots, 0) \in \mathbb{R}^D$.

In this case, $X_t = \boldsymbol{A} Y_{t-1}$ lies on $\mathrm{Im}(\boldsymbol{A})$.

As rank$(\boldsymbol{A}) < D$, $Im(A)$ is a linear subspace in $\mathbb{R}^D$ of dimension rank$(\boldsymbol{A})$.

$$Y_t = -1 - \cos\big(0.2\pi(Y_{t-1} - 2Y_{t-2})\big) + 0.2\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0,1)$$

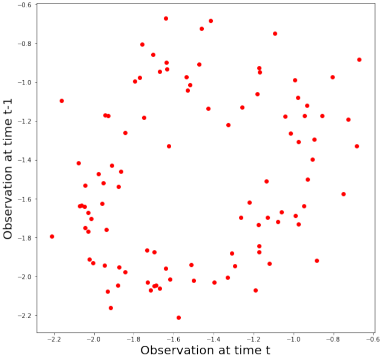


Figure: Example of hidden manifold time series model. The vectors (Y_t, Y_{t-1}) lie around 1-dimensional manifold

Fix an integer b, construct patches

$$\{Y_t = (Z_{t-1}, Z_{t-2}, \dots, Z_{t-b}) \in \mathbb{R}^{pb} : b+1 \leqslant t \leqslant T\}$$

and consider a set of pairs

$$S_T = \{(Y_t, Z_t) : b+1 \leqslant t \leqslant T\}$$

Assumption: high-dimensional vectors $Y_{b+1}, \dots, Y_{T+1} \in \mathbb{R}^D$, $D = bp$, lie around a low-dimensional manifold

Project the points $Y_{b+1}, \ldots, Y_{T+1}$ onto a manifold and denote the projections by $\widehat{X}_{b+1}, \ldots, \widehat{X}_T$, respectively

Assumption: close vectors $\widehat{X}_i$ and $\widehat{X}_j$ correspond to similar $Z_i - Z_{i-1}$ and $Z_j - Z_{j-1}$, i. e. similar historical behaviours induce similar increments

Prediction at the moment $T+1$ is determined by the weighted k-nearest neighbors rule:

$$\widehat{Z}_{T+1} = Z_T + \frac{\sum\limits_{t=b+1}^{T} w_t (Z_t - Z_{t-1})}{\sum\limits_{t=b+1}^{T} w_t},$$

where the weights are defined by the formula

$$w_t = e^{-(T+1-t)/\tau} \mathcal{K}\left(\frac{\|\widehat{X}_{T+1} - \widehat{X}_t\|}{h_k} \right)$$

Here h_k is the distance to k-th nearest neighbor, $\tau > 0$ is a discounting parameter and $\mathcal{K}(\cdot)$ is a localizing kernel

In [Osher et al., 2017], the authors suggested recovering a manifold solving the following optimization problem (approximately):

$$\frac{1}{T}\sum_{t=1}^{T} d^2(Y_t, \mathcal{M}) + \lambda \mathsf{dim}(\mathcal{M}) \to \min_{\mathcal{M}}$$

Numerical approximation is based on the identity

$$\mathsf{dim}(\mathcal{M}) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{D}\|\nabla_{\mathcal{M}} X_t^{(j)}\|^2,$$

where $X_t^{(j)}$ is the j-th component of X_t, $\nabla_{\mathcal{M}}$ stands for differentiation with respect to intrinsic coordinates on $\mathcal{M}$.

Approximation of the target functional:

$$\frac{1}{T}\sum_{t=1}^{T} d^2(Y_t,\mathcal{M}) + \lambda \text{dim}(\mathcal{M})$$
$$= \frac{1}{T^2 h^d}\sum_{i=1}^{T}\sum_{j=1}^{T}\|Y_i - X_j\|^2 w_{ij}$$
$$+ \frac{\lambda}{2T^2 h^d}\sum_{i=1}^{n}\sum_{j=1}^{n}\|X_i - X_j\|^2 w_{ij} + o(h),$$

where $w_{ij} = e^{-\|X_i - X_j\|^2/(4h^2)}$. One can use the proximal gradient descent method to find an approximate solution [Osher et al., 2017]

Another structure adaptive procedure is offered in [Puchkin and Spokoiny, 2019]. The idea originates from [Hristache et al., 2001b]:

- The method iteratively estimates projections of $Y_1, \dots, Y_T$ onto a manifold and projectors onto the tangent spaces
- Estimates of projections of $Y_1, \dots, Y_T$ onto the manifold are weighted averages
- On early stage, the points $Y_1, \dots, Y_T$ may lie quite far from $\mathcal{M}^*$, so the method averages over large isotropic neighborhoods to capture the global structure
- On final steps, the method averages over cylindric neighborhoods stretched along normal directions

Initialization: for each i, compute the weighted sample covariance

$$\widetilde{\mathbf{\Sigma}}_i = \sum_{j=1}^{T} \widetilde{w}_{ij}(Y_j - Y_i)(Y_j - Y_i)^T, \quad 1 \leqslant i, j \leqslant T,$$

with the weights

$$\widetilde{w}_{ij} = \exp\left(-h_0^{-2}\|(Y_i - Y_j)\|^2\right), \quad 1 \leqslant i, j \leqslant T,$$

where $\sigma \ll h_0 \ll \varkappa$.

Define $\widehat{\mathbf{\Pi}}_i^{(0)}$ as a projector onto a linear span of eigenvectors, corresponding to the d largest eigenvalues of $\widetilde{\mathbf{\Sigma}}_i$.

- The initial guesses $\widehat{\mathbf{\Pi}}_1^{(0)}, \dots, \widehat{\mathbf{\Pi}}_T^{(0)}$ of $\mathbf{\Pi}(X_1), \dots, \mathbf{\Pi}(X_T)$, the number of iterations $K+1$, an initial bandwidth h_0, the threshold τ and constants $a > 1$ and $\gamma > 0$ are given.
- k from 0 to K do
  - Compute the weights $w_{ij}^{(k)}$ according to the formula

$$w_{ij}^{(k)} = \exp\left(-h_k^{-2} \|\widehat{\mathbf{\Pi}}_i^{(k)}(Y_i - Y_j)\|^2\right) \mathbb{1}\left(\|Y_i - Y_j\| \leqslant \tau\right), \quad 1$$

  - Update $\widehat{X}_i^{(k)}$ for each i.
  - If $k < K$, update $\widehat{\mathbf{\Pi}}_i^{(k+1)}$ for each i and set $h_{k+1} = a^{-1} h_k$.
- Return the estimates $\widehat{X}_1 = \widehat{X}_1^{(K)}, \dots, \widehat{X}_T = \widehat{X}_T^{(K)}$

Estimates update:

$$\widehat{X}_i^{(k)} = \left(\sum_{j=1}^{T} w_{ij}^{(k)} Y_j \right) \Big/ \left(\sum_{j=1}^{T} w_{ij}^{(k)} \right), \quad 1 \leqslant i \leqslant T$$

Projectors update:

- For each i from 1 to T, define a set $\mathcal{J}_i^{(k)} = \{ j : \| \widehat{X}_j^{(k)} - \widehat{X}_i^{(k)} \| \leqslant \gamma h_k \}$ and compute the matrices

$$\widehat{\mathbf{\Sigma}}_i^{(k)} = \sum_{j \in \mathcal{J}_i^{(k)}} (\widehat{X}_j^{(k)} - \widehat{X}_i^{(k)})(\widehat{X}_j^{(k)} - \widehat{X}_i^{(k)})^T, \quad 1 \leqslant i \leqslant T$$

- If $k < K$, for each i from 1 to T, define $\widehat{\mathbf{\Pi}}_i^{(k+1)}$ as a projector onto a linear span of eigenvectors of $\widehat{\mathbf{\Sigma}}_i^{(k)}$, corresponding to the largest d eigenvalues

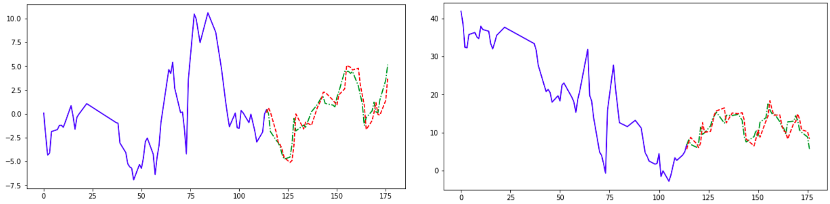
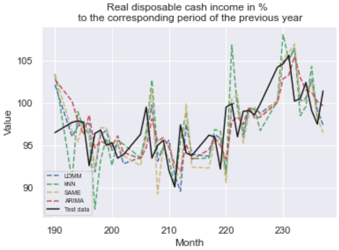
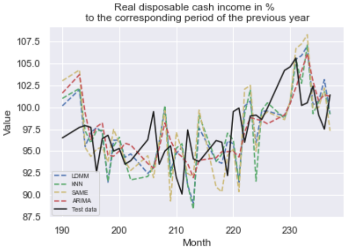


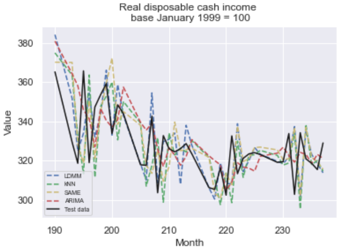
Figure: An example of prediction of stock prices for two companies. The blue line corresponds to the historical data. The red line and the green line correspond to the prediction and the actual value, respectively



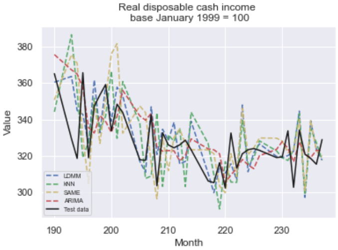
(a) One month ahead



(b) Two months ahead



(c) One month ahead



(d) Two months ahead

(e) One month ahead

(f) Two months ahead

(g) One month ahead

(h) Two months ahead

Hristache, M., Juditsky, A., Polzehl, J., and Spokoiny, V. (2001a).
Structure adaptive approach for dimension reduction.
Ann. Statist., 29(6):1537–1566.

Hristache, M., Juditsky, A., and Spokoiny, V. (2001b).
Direct estimation of the index coefficient in a single-index model.
Ann. Statist., 29(3):595–623.

Osher, S., Shi, Z., and Zhu, W. (2017).
Low dimensional manifold model for image processing.
SIAM J. Img. Sci., 10:1669–1690.

Puchkin, N. and Spokoiny, V. (2019).
Structure-adaptive manifold estimation.
working paper or preprint.