Semiparametric Ultra-High Dimensional Model Averaging of Nonlinear Dynamic Time Series

Degui LI

Department of Mathematics, University of York

Joint work with Jia Chen, Oliver Linton and Zudi Lu

1. Introduction
2. KSIS+PMAMAR
3. PCA+PMAMAR
4. Empirical application

Suppose that Y_t, $t = 1, \ldots, n$, are n observations collected from a stationary time series process and $\mathbf{X}_t = (\mathbf{Z}_t^\intercal, \mathbf{Y}_{t-1}^\intercal)^\intercal$ with $\mathbf{Z}_t = \left(Z_{t1}, Z_{t2}, \ldots, Z_{tp_n}\right)^\intercal$ and $\mathbf{Y}_{t-1} = (Y_{t-1}, Y_{t-2}, \ldots, Y_{t-d_n})^\intercal$, where $\mathbf{Z}_t$ are exogenous regressors.

Suppose that Y_t, $t = 1, \ldots, n$, are n observations collected from a stationary time series process and $\mathbf{X}_t = (\mathbf{Z}_t^\intercal, \mathbf{Y}_{t-1}^\intercal)^\intercal$ with $\mathbf{Z}_t = \left(Z_{t1}, Z_{t2}, \ldots, Z_{tp_n}\right)^\intercal$ and $\mathbf{Y}_{t-1} = (Y_{t-1}, Y_{t-2}, \ldots, Y_{t-d_n})^\intercal$, where $\mathbf{Z}_t$ are exogenous regressors.

The main interest is to study the multivariate regression function:

$$m(\mathbf{x}) = \mathrm{E}(Y_t \big| \mathbf{X}_t = \mathbf{x}). \tag{1}$$

It is well known that when the dimension of $\mathbf{X}_t$ is very small (univariate or bivariate), $m(\mathbf{x})$ can be well estimated by some commonly-used nonparametric methods such as the kernel method, the local polynomial method and the spline method.

However, if the dimension is large, to address the so-called "curse of dimensionality", various nonparametric and semiparametric models such as additive models, varying-coefficient models and partially linear models have been proposed in the literature for the dynamic time series data (c.f., Fan and Yao, 2003; Teräsvirta, Tjøstheim and Granger, 2010).

In this talk, we assume that the dimension of the exogenous variables $\mathbf{Z}_t$ may be diverging at certain exponential rate of n, which indicates that the dimension of the potential explanatory variables $\mathbf{X}_t$ can be diverging at an exponential rate, i.e.,

$$p_n + d_n = O(\exp\{n^{\delta_0}\})$$

for some positive constant δ_0.

A motivated example

We apply the proposed semiparametric model averaging methods to forecast inflation in the UK. The data were collected from the Office for National Statistics (ONS) and the Bank of England (BoE) websites and included quarterly observations on CPI and some other economics variables over the period Q1 1997 to Q4 2013.

All the variables are seasonally adjusted. We use 53 predictor series measuring aggregate real activity and other economic indicators to forecast CPI. Given the possible time persistence of CPI, we also add its 4 lags as predictors.

Data from Q1 1997 to Q4 2012 are used as the training set and those between Q1 2013 and Q4 2013 are used for forecasting.

The dimension of the candidate covariates is $53 + 4 = 57$, which is comparable to $16 \times 4 = 64$ observations in the training set.

Our aims

- Propose two types of semiparametric dimension reduction with the penalised estimation method involved.

Our aims

- Propose two types of semiparametric dimension reduction with the penalised estimation method involved.
- Examine the nonlinear forecasting performance after dimension reduction.

Step 1 of KSIS+PMAMAR

- Apply Kernel Sure Independence Screening (KSIS) technique for the nonlinear time series setting which screens out the regressors whose marginal regression (or autoregression) functions do not make significant contribution to estimating the joint multivariate regression function.

Step 2 of KSIS+PMAMAR

- Consider a semiparametric method of Model Averaging MArginal Regression (MAMAR) for the regressors and autoregressors that survive the screening procedure, and propose a penalised MAMAR method to further select the regressors and determine the optimal combination of the significant marginal regression and autoregression functions.

KSIS

Literature on SIS

- Linear models: Fan and Lv (2008);
- Additive models: Fan, Feng and Song (2011);
- Varying coefficient models: Fan, Ma and Dai (2014) and Liu, Li and Wu (2014);
- Linear SIS+model averaging: Ando and Li (2014).

KSIS–1

For notational simplicity, we let

$$X_{tj} = \begin{cases} Z_{tj}, & j = 1, 2, \ldots, p_n, \\ Y_{t-(j-p_n)}, & j = p_n + 1, p_n + 2, \ldots, p_n + d_n. \end{cases}$$

KSIS–1

For notational simplicity, we let

$$X_{tj} = \begin{cases} Z_{tj}, & j = 1, 2, \ldots, p_n, \\ Y_{t-(j-p_n)}, & j = p_n + 1, p_n + 2, \ldots, p_n + d_n. \end{cases}$$

For $j = 1, \ldots, p_n + d_n$, the kernel smoother of marginal regression

$$m_j(x_j) := \mathrm{E}(Y_t | X_{tj} = x_j)$$

is

$$\hat{m}_j(x_j) = \frac{\sum_{t=1}^{n} Y_t K_{tj}(x_j)}{\sum_{t=1}^{n} K_{tj}(x_j)}, \quad K_{tj}(x_j) = K\left(\frac{X_{tj} - x_j}{h_1}\right).$$

KSIS–2

We consider ranking the importance of the covariates by calculating the correlation between the response variable and marginal regression:

$$\mathrm{cor}(j) = \frac{\mathrm{cov}(j)}{\sqrt{\mathrm{v}(Y) \cdot \mathrm{v}(j)}} = \left[\frac{\mathrm{v}(j)}{\mathrm{v}(Y)}\right]^{1/2}, \qquad (2)$$

where $\mathrm{v}(Y) = \mathrm{var}(Y_t)$, $\mathrm{v}(j) = \mathrm{var}(m_j(X_{tj}))$ and

$$\mathrm{cov}(j) = \mathrm{cov}(Y_t, m_j(X_{tj})) = \mathrm{var}(m_j(X_{tj})) = \mathrm{v}(j).$$

Equation (2) indicates that the value of $\mathrm{cor}(j)$ is non-negative for all j and the ranking of $\mathrm{cor}(j)$ is equivalent to the ranking of $\mathrm{v}(j)$ as $\mathrm{v}(Y)$ is positive and invariant across j.

KSIS–3

The sample version of $\text{cor}(j)$ can be constructed as

$$\hat{\text{cor}}(j) = \frac{\hat{\text{cov}}(j)}{\sqrt{\hat{\text{v}}(Y) \cdot \hat{\text{v}}(j)}} = \left[\frac{\hat{\text{v}}(j)}{\hat{\text{v}}(Y)}\right]^{1/2},$$

where

$$\hat{\text{v}}(Y) = \frac{1}{n}\sum_{t=1}^{n} Y_t^2 - \left(\frac{1}{n}\sum_{t=1}^{n} Y_t\right)^2,$$

$$\hat{\text{cov}}(j) = \hat{\text{v}}(j) = \frac{1}{n}\sum_{t=1}^{n} \hat{m}_j^2(X_{tj}) - \left[\frac{1}{n}\sum_{t=1}^{n} \hat{m}_j(X_{tj})\right]^2,$$

KSIS criterion

The screened sub-model can be determined by,

$$\hat{\mathcal{S}} = \{j = 1, 2, \ldots, p_n + d_n : \ \hat{\mathrm{v}}(j) \geq \rho_n\},$$

where ρ_n is a pre-determined positive number.

KSIS criterion

The screened sub-model can be determined by,

$$\hat{\mathcal{S}} = \{j = 1, 2, \ldots, p_n + d_n : \ \hat{\mathrm{v}}(j) \geq \rho_n\},$$

where ρ_n is a pre-determined positive number.

The above criterion is equivalent to

$$\hat{\mathcal{S}} = \{j = 1, 2, \ldots, p_n + d_n : \ \hat{\mathrm{cor}}(j) \geq \rho_n^{\diamond}\},$$

where $\rho_n^{\diamond} = \rho_n^{1/2} / \sqrt{\hat{\mathrm{v}}(Y)}$.

Asymptotic theory for KSIS–1

Define the index set of “true” candidate models as

$$\mathcal{S} = \{j = 1, 2, \ldots, p_n + d_n : \ \mathrm{v}(j) \neq 0\}.$$

Asymptotic theory for KSIS–2

Theorem 1(i) *Suppose that the conditions A1–A5 in the paper are satisfied. For any small* $\delta_1 > 0$*, there exists a positive constant* δ_2 *such that*

$$\mathsf{P}\left(\max_{1\le j\le p_n+d_n}\left|\hat{\mathsf{v}}(j)-\mathsf{v}(j)\right| > \delta_1 n^{-2(1-\theta_1)/5}\right) = O\left(M(n)\exp\left\{-\delta_2 n^{(1-\theta_1)/5}\right\}\right),$$

where $M(n) = (p_n + d_n)n^{(17+18\theta_1)/10}$ *and* $1/6 < \theta_1 < 1$ *is defined by* $h_1 = n^{-\theta_1}$.

Asymptotic theory for KSIS–3

Theorem 1(ii) *If we choose the pre-determined tuning parameter* $\rho_n = \delta_1 n^{-2(1-\theta_1)/5}$ *and assume*

$$\min_{j \in \mathcal{S}} \mathrm{v}(j) \geq 2\delta_1 n^{-2(1-\theta_1)/5},$$

then we have

$$\mathrm{P}\left(\mathcal{S} \subset \hat{\mathcal{S}}\right) \geq 1 - O\left(M_{\mathcal{S}}(n) \exp\left\{-\delta_2 n^{(1-\theta_1)/5}\right\}\right),$$

where $M_{\mathcal{S}}(n) = |\mathcal{S}| n^{(17+18\theta_1)/10}$ *with* $|\mathcal{S}|$ *being the cardinality of* $\mathcal{S}$.

Remark on the asymptotic theory

As $p_n + d_n = O(\exp\{n^{\delta_0}\})$, in order to ensure the validity of Theorem 1(i), we need to impose the restriction $\delta_0 < (1 - \theta_1)/5$, which reduces to $\delta_0 < 4/25$ if the order of the optimal bandwidth in kernel smoothing (i.e., $\theta_1 = 1/5$) is used.

PMAMAR

MAMAR approximation–1

We denote the chosen covariates (after KSIS in the first step) by $\mathbf{X}_t^* = \left(X_{t1}^*, X_{t2}^*, \ldots, X_{tq_n}^*\right)^{\intercal}$ which may include both exogenous variables and lags of Y_t, where q_n might be divergent but is smaller than the sample size n.

MAMAR approximation–1

We denote the chosen covariates (after KSIS in the first step) by $\mathbf{X}_t^* = \left(X_{t1}^*, X_{t2}^*, \ldots, X_{tq_n}^*\right)^{\intercal}$ which may include both exogenous variables and lags of Y_t, where q_n might be divergent but is smaller than the sample size n.

Define

$$m^*(\mathbf{x}) = \mathrm{E}(Y_t | \mathbf{X}_t^* = \mathbf{x}), \quad (3)$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_{q_n})^{\intercal}$.

MAMAR approximation–2

We approximate the conditional regression function $m^*(\mathbf{x})$ by an affine combination of one-dimensional conditional component regressions

$$m_j^*(x_j) = \mathsf{E}(Y_t | X_{tj}^* = x_j), \quad j = 1, \ldots, q_n.$$

Each marginal regression $m_j^*(\cdot)$ can be treated as a "nonlinear candidate model".

MAMAR approximation–2

We approximate the conditional regression function $m^*(\mathbf{x})$ by an affine combination of one-dimensional conditional component regressions

$$m_j^*(x_j) = \mathsf{E}(Y_t | X_{tj}^* = x_j), \quad j = 1, \ldots, q_n.$$

Each marginal regression $m_j^*(\cdot)$ can be treated as a "nonlinear candidate model".

A weighted average of $m_j^*(x_j)$ is used to approximate $m^*(\mathbf{x})$, i.e.,

$$m^*(\mathbf{x}) \approx w_0 + \sum_{j=1}^{q_n} w_j m_j^*(x_j),$$

where w_j, $j = 0, 1, \ldots, q_n$, are to be determined later and can be seen as the weights for different "candidate models".

MAMAR approximation–3

By replacing $m_j^*(X_{tj}^*)$, $j = 1, \ldots, q_n$, by their corresponding nonparametric estimates $\hat{m}_j^*(X_{tj}^*)$, we have the following "approximate linear model":

$$Y_t \approx w_0 + \sum_{j=1}^{q_n} w_j \hat{m}_j^*(X_{tj}^*).$$

The above MAMAR approximation is introduced in Li, Linton and Lu (2015) and recently applied by Chen *et al* (2015) in the dynamic portfolio choice with many conditioning variables.

PMAMAR–1

For $j = 1, \ldots, q_n$, we estimate the marginal regression functions $m_j^*(\cdot)$ by the kernel smoothing method:

$$\hat{m}_j^*(x_j) = \frac{\sum_{t=1}^n Y_t \overline{K}_{tj}(x_j)}{\sum_{t=1}^n \overline{K}_{tj}(x_j)}, \quad \overline{K}_{tj}(x_j) = K\Big(\frac{X_{tj}^* - x_j}{h_2}\Big).$$

Then, for $j = 1, \ldots, q_n$, we let

$$\hat{\mathcal{M}}(j) = \left[\hat{m}_j^*(X_{1j}^*), \ldots, \hat{m}_j^*(X_{nj}^*)\right]^{\intercal} =: \mathcal{S}_n(j)\mathcal{Y}_n$$

be the estimated values of

$$\mathcal{M}(j) = \left[m_j^*(X_{1j}^*), \ldots, m_j(X_{nj}^*)\right]^{\intercal},$$

where $\mathcal{S}_n(j)$ is the $n \times n$ smoothing matrix whose (k, l)-component is $\overline{K}_{lj}(X_{kj}^*)/\big[\sum_{t=1}^n \overline{K}_{tj}(X_{kj}^*)\big]$, and $\mathcal{Y}_n = (Y_1, \ldots, Y_n)^{\intercal}$.

PMAMAR–2

We define the objective function by

$$\mathcal{Q}_n(\mathbf{w}_n) = \left[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_n)\right]^{\intercal}\left[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_n)\right] + n\sum_{j=1}^{q_n} p_\lambda(|w_j|), \quad (4)$$

where

$$\hat{\mathcal{M}}(\mathbf{w}_n) = \left[w_1\mathcal{S}_n(1) + \ldots + w_{q_n}\mathcal{S}_n(q_n)\right]\mathcal{Y}_n = \mathcal{S}_n(\mathcal{Y})\mathbf{w}_n,$$

$\mathcal{S}_n(\mathcal{Y}) = \left[\mathcal{S}_n(1)\mathcal{Y}_n, \ldots, \mathcal{S}_n(q_n)\mathcal{Y}_n\right]$, and $p_\lambda(\cdot)$ is a penalty function with a tuning parameter λ.

PMAMAR–2

We define the objective function by

$$\mathcal{Q}_n(\mathbf{w}_n) = \left[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_n)\right]^{\intercal}\left[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_n)\right] + n\sum_{j=1}^{q_n} p_{\lambda}(|w_j|), \quad (4)$$

where

$$\hat{\mathcal{M}}(\mathbf{w}_n) = \left[w_1\mathcal{S}_n(1) + \ldots + w_{q_n}\mathcal{S}_n(q_n)\right]\mathcal{Y}_n = \mathcal{S}_n(\mathcal{Y})\mathbf{w}_n,$$

$\mathcal{S}_n(\mathcal{Y}) = \left[\mathcal{S}_n(1)\mathcal{Y}_n, \ldots, \mathcal{S}_n(q_n)\mathcal{Y}_n\right]$, and $p_{\lambda}(\cdot)$ is a penalty function with a tuning parameter λ.

Our semiparametric estimator of the optimal weights $\mathbf{w}_o$ can be obtained through minimising the objective function $\mathcal{Q}_n(\mathbf{w}_n)$:

$$\hat{\mathbf{w}}_n = \arg\min_{\mathbf{w}_n} \mathcal{Q}_n(\mathbf{w}_n).$$

Penalty functions

- AIC and BIC: $p_\lambda(|z|) = 0.5\lambda^2 I(|z| \neq 0)$ with different values of λ;
- LASSO: $p_\lambda(|z|) = \lambda|z|$;
- SCAD: $p'_\lambda(z) = \lambda \left[I(z \leq \lambda) + \frac{a_0\lambda - z}{(a_0-1)\lambda} I(z > \lambda) \right]$ with $p_\lambda(0) = 0$, where $a_0 > 2$, $\lambda > 0$ and $I(\cdot)$ is the indicator function.

In our numerical studies, we use the SCAD penalty in PMAMAR.

Asymptotic theory for PMAMAR–1

Theorem 2 (i) *Suppose that the conditions A1–A8 are satisfied. There exists a local minimizer* $\hat{\mathbf{w}}_n$ *of the objective function* $\mathcal{Q}_n(\cdot)$ *defined in (4) such that*

$$\|\hat{\mathbf{w}}_n - \mathbf{w}_o\| = O_P\left(\sqrt{q_n}(n^{-1/2} + a_n)\right),$$

where $\|\cdot\|$ *denotes the Euclidean norm and*

$$a_n = \max_{1 \le j \le q_n} \left\{ |p'_\lambda(|w_{oj}|)|,\ |w_{oj}| \neq 0 \right\}.$$

Asymptotic theory for PMAMAR–2

Theorem 2 (ii) *Let $\hat{\mathbf{w}}_n(2)$ be the estimator of $\mathbf{w}_o(2)$ which is composed of all the zero weights and further assume that*

$$\lambda \to 0, \quad \frac{\sqrt{n}\lambda}{\sqrt{q_n}} \to \infty, \quad \liminf_{n\to\infty} \liminf_{\vartheta\to 0+} \frac{p'_\lambda(\vartheta)}{\lambda} > 0.$$

Then, the local minimizer $\hat{\mathbf{w}}_n$ of the objective function $\mathcal{Q}_n(\cdot)$ satisfies $\hat{\mathbf{w}}_n(2) = \mathbf{0}$ with probability approaching one.

Asymptotic theory for PMAMAR–3

Theorem 2 (iii) *If we further assume that the eigenvalues of $\mathbf{\Lambda}_{n1}$ are bounded away from zero and infinity,*

$$\sqrt{n}\mathbf{A}_n\mathbf{\Sigma}_n^{-1/2}\left(\mathbf{\Lambda}_{n1}+\mathbf{\Omega}_n\right)\left[\hat{\mathbf{w}}_n(1)-\mathbf{w}_o(1)-\left(\mathbf{\Lambda}_{n1}+\mathbf{\Omega}_n\right)^{-1}\boldsymbol{\omega}_n\right]\xrightarrow{d}N(\mathbf{0},\mathbf{A}_0),$$

where $\mathbf{0}$ is a null vector whose dimension may change from line to line, $\mathbf{A}_n$ is an $s\times s_n$ matrix such that $\mathbf{A}_n\mathbf{A}_n^{\intercal}\to\mathbf{A}_0$ and $\mathbf{A}_0$ is an $s\times s$ symmetric and non-negative definite matrix, s is a fixed positive integer. The definitions of $\mathbf{\Sigma}_n$, $\mathbf{\Lambda}_{n1}$, $\mathbf{\Omega}_n$ and $\boldsymbol{\omega}_n$ are given in the paper.

Simulation for KSIS+PMAMAR

Example 1

The sample size is set to be $n = 100$, and the numbers of candidate exogenous covariates and lagged terms are $(p_n, d_n) = (30, 10)$ and $(p_n, d_n) = (150, 50)$. The model is defined by

$$\begin{aligned} Y_t &= m_1(Z_{t1}) + m_2(Z_{t2}) + m_3(Z_{t3}) + m_4(Z_{t4}) + m_5(Y_{t-1}) \\ &\quad + m_6(Y_{t-2}) + m_7(Y_{t-3}) + \varepsilon_t \end{aligned}$$

for $t \geq 1$, where, following Meier, van de Geer and Bühlmann (2009), we set

$$m_i(x) = \sin(0.5\pi x), \qquad i = 1, 2, \ldots, 7.$$

The exogenous covariates

$$\mathbf{Z}_t = (Z_{t1}, Z_{t2}, \ldots, Z_{tp_n})^{\intercal}$$

are independently drawn from p_n-dimensional Gaussian distribution with zero mean and covariance matrix $\mathrm{cov}(\mathbf{Z}) = I_{p_n}$ or $C_{\mathbf{Z}}$, whose the main-diagonal entries are 1 and off-diagonal entries are $1/2$.

The error term ε_t are independently generated from the $\mathrm{N}(0, 0.7^2)$ distribution. The real size of exogenous regressors is 4 and the real lag length is 3.

We generate $100 + n$ observations from the process with initial states $Y_{-2} = Y_{-1} = Y_0 = 0$ and discard the first $100 - d_n$ observations.

Summary of the simulation result

- The iterative version of KSIS+PMAMAR performs better in both estimation and prediction than the KSIS+PMAMAR.
- The penGAM is the most conservative in variable selection and on average selects the least number of variables.
- The ISIS suffers from the model misspecification problem.
- When the correlation among the exogenous variables increases, the performance of all approaches worsens.

PCA+PMAMAR

- Impose an approximate factor modelling structure on the ultra-high dimensional exogenous regressors and use the well-known principal component analysis to estimate the latent common factors;
- Apply the PMAMAR method to select the estimated common factors and lags of the response variable which are significant.

Factor models and PCA

PCA–1

Letting

$$\mathbf{B}_n^0 = (\mathbf{b}_1^0, \ldots, \mathbf{b}_{p_n}^0)^\intercal \quad \text{and} \quad \mathbf{U}_t = (u_{t1}, \ldots, u_{tp_n})^\intercal,$$

we assume the approximate factor model:

$$\mathbf{Z}_t = \mathbf{B}_n^0 \mathbf{f}_t^0 + \mathbf{U}_t,$$

where $\mathbf{b}_k^0$ is an r-dimensional vector of factor loadings, $\mathbf{f}_t^0$ is an r-dimensional vector of common factors, and u_{tk} is called an idiosyncratic error.

PCA–2

Denote $\mathcal{Z}_n = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)^{\mathsf{T}}$, the $n \times p_n$ matrix of the observations of the exogenous variables. We then construct

$$\hat{\mathcal{F}}_n = \left(\hat{\mathbf{f}}_1, \ldots, \hat{\mathbf{f}}_n\right)^{\mathsf{T}}$$

as the $n \times r$ matrix consisting of the r eigenvectors (multiplied by $\sqrt{n}$) associated with the r largest eigenvalues of the $n \times n$ matrix

$$\mathcal{Z}_n \mathcal{Z}_n^{\mathsf{T}} / (np_n).$$

PCA–3

Define

$$\mathbf{H} = \hat{\mathbf{V}}^{-1}\left(\hat{\mathcal{F}}_n^{\mathsf{T}}\mathcal{F}_n^0/n\right)\left[(\mathbf{B}_n^0)^{\mathsf{T}}\mathbf{B}_n^0/p_n\right], \quad \mathcal{F}_n^0 = \left(\mathbf{f}_1^0, \ldots, \mathbf{f}_n^0\right)^{\mathsf{T}},$$

and $\hat{\mathbf{V}}$ is the $r \times r$ diagonal matrix of the first r largest eigenvalues of $\mathcal{Z}_n\mathcal{Z}_n^{\mathsf{T}}/(np_n)$ arranged in descending order.

Asymptotic theory for PCA

Theorem 3 (i). *Suppose that the conditions B1–B4 are satisfied, and*

$$n = o(p_n^2), \quad p_n = O\left(\exp\{n^{\delta_*}\}\right), \quad 0 \le \delta_* < 1/3.$$

For the PCA estimation $\hat{\mathbf{f}}_t$*, we have*

$$\max_t \left\| \hat{\mathbf{f}}_t - \mathbf{H}\mathbf{f}_t^0 \right\| = O_P\left(n^{-1/2} + n^{1/4} p_n^{-1/2}\right).$$

PMAMAR with estimated factor regressors

Consider the following multivariate regression function with rotated latent factors and lags of response:

$$m_f^*(\mathbf{x}_1, \mathbf{x}_2) = \mathsf{E}\left(Y_t | \mathbf{H}\mathbf{f}_t^0 = \mathbf{x}_1, \mathbf{Y}_{t-1} = \mathbf{x}_2\right).$$

Apply the PMAMAR with

$$\hat{\mathbf{X}}_{t,f}^* = \left(\hat{\mathbf{f}}_t^\intercal, \mathbf{Y}_{t-1}^\intercal\right)^\intercal = \left(\hat{f}_{t1}, \ldots, \hat{f}_{tr}, \mathbf{Y}_{t-1}^\intercal\right)^\intercal.$$

For $k = 1, \ldots, r$, define

$$m^*_{k,f}(z_k) = \mathrm{E}\left[Y_t | \tilde{f}^0_{tk} = z_k\right], \quad \tilde{f}^0_{tk} = e_r^\intercal(k)\mathbf{H}\mathbf{f}^0_t,$$

where $e_r(k)$ is an r-dimensional column vector with the k-th element being one and zeros elsewhere, $k = 1, \ldots, r$.

We estimate $m^*_{k,f}(z_k)$ by the kernel smoothing method:

$$\hat{m}^*_{k,f}(z_k) = \frac{\sum_{t=1}^n Y_t \widetilde{K}_{tk}(z_k)}{\sum_{t=1}^n \widetilde{K}_{tk}(z_k)}, \quad \widetilde{K}_{tk}(z_k) = K\left(\frac{\hat{f}_{tk} - z_k}{h_3}\right), \quad j = 1, \ldots r,$$

where h_3 is a bandwidth and $\hat{f}_{tk}$ is the k-th element of $\hat{\mathbf{f}}_t$.

Theorem 3(ii) *Suppose that the conditions A5 and B1–B5 are satisfied, and the latent factor* $\mathbf{f}_t^0$ *has a compact support. Then we have*

$$\max_{1\leq k\leq r} \sup_{z_k\in\mathcal{F}_k^*} \left|\hat{m}_{k,f}^*(z_k) - \tilde{m}_{k,f}^*(z_k)\right| = o_P\left(n^{-1/2}\right),$$

where $\mathcal{F}_k^*$ *is the compact support of* $\tilde{f}_{tk}^0$*,* $\tilde{m}_{k,f}^*(z_k)$ *is the infeasible kernel estimation defined as* $\hat{m}_{k,f}^*(z_k)$ *but with* $\hat{f}_{tk}$ *replaced by* $\tilde{f}_{tk}^0$*.*

Relevant literature

- Factor-augmented linear regression and autoregression: Stock and Watson (2002), Bernanke, Boivin and Eliasz, (2005) Bai and Ng (2006), Pesaran, Pick and Timmermann (2011) and Cheng and Hansen (2015).

Simulation for PCA+PMAMAR

How to choose the number of factors

- Set a maximum number, say $r_{\max}$ (which is usually not too large), for the factors. Since the factors extracted from the eigenanalysis are orthogonal to each other, the over-extracted insignificant factors will be discarded in the PMAMAR step.
- Select the first few eigenvectors (corresponding to the first few largest eigenvalues) of $\mathcal{Z}_n \mathcal{Z}_n^{\intercal}/(np_n)$ so that a pre-determined amount, say 95%, of the total variation is accounted for.
- Other commonly-used selection criteria such as BIC can be found in Bai and Ng (2002) and Fan, Liao and Mincheva (2013).

Example 2

The exogenous variables $\mathbf{Z}_t$ are generated via an approximate factor model:

$$\mathbf{Z}_t = \mathbf{B}\mathbf{f}_t + \mathbf{z}_t,$$

where the rows of the $p_n \times r$ loadings matrix $\mathbf{B}$ and the common factors $\mathbf{f}_t$, $t = 1, \cdots, n$, are independently generated from the multivariate $\mathrm{N}(\mathbf{0}, I_r)$ distribution, and the p_n-dimensional error terms $\mathbf{z}_t$, $t = 1, \cdots, n$, are independently drawn from $0.1\mathrm{N}(\mathbf{0}, I_{p_n})$.

We set $p_n = 30$ or 150, $r = 3$, and generate the response variable via

$$\begin{aligned} Y_t &= m_1(f_{t1}) + m_2(f_{t2}) + m_3(f_{t3}) + m_4(Y_{t-1}) \\ &\quad + m_5(Y_{t-2}) + m_6(Y_{t-3}) + \varepsilon_t, \end{aligned}$$

where f_{ti} is the i-th component of $\mathbf{f}_t$, $m_i(\cdot)$, $i = 1, \cdots, 6$, are the same as in Example 1, and ε_t, $t = 1, \cdots, n$, are independently drawn from the $\mathrm{N}(0, 0.7^2)$ distribution.

In this example, we choose the number of candidate lags of Y as $d_n = 10$.

Summary of simulation result

- When $p_n = 30$, the KSIS+PMAMAR outperforms all the other approaches (except the Oracle) in terms of estimation and prediction accuracy.
- When p_n becomes larger than n, the PCA based approaches show their advantage in effective dimension reduction of the exogenous variables, which results in their lower EE and PE.
- The PCA+PMAMAR has a lower EE but higher PE than the PCA+KSIS+PMAMAR. This is due to the fact that without the KSIS step the PCA+PMAMAR selects more false lags of Y.

Example 3

We next apply the proposed semiparametric model averaging methods to forecast inflation in the UK. The data were collected from ONS and BoE websites and included quarterly observations on CPI and some other economics variables over the period Q1 1997 to Q4 2013.

All the variables are seasonally adjusted. We use 53 predictor series measuring aggregate real activity and other economic indicators to forecast CPI. Given the possible time persistence of CPI, we also add its 4 lags as predictors.

Data from Q1 1997 to Q4 2012 are used as the training set and those between Q1 2013 and Q4 2013 are used for forecasting.

Method	IKSIS+PMAMAR	KSIS+PMAMAR	PCA+PMAMAR
PE	0.0360	0.1130	0.0787
Method	penGAM	ISIS	Phillips curve
PE	0.0865	0.3275	1.1900

The Phillips curve specification is:

$$I_{t+1} - I_t = \alpha + \beta(L)U_t + \gamma(L)\Delta I_t + \varepsilon_{t+1},$$

where I_t is the CPI in the t-th quarter,
$\beta(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 L^3$ and $\gamma(L) = \gamma_0 + \gamma_1 L + \gamma_2 L^2 + \gamma_3 L^3$ are lag polynomials with L being the lag operator, U_t is the unemployment rate, and Δ is the first difference operator.

Thank you very much