Kernel Three Pass Regression Filter

Rajveer Jat and Daanish Padha

University of California Riverside

Introduction

We propose a novel supervised, non-linear, yet computationally fast method of forecasting a single time series using a high-dimensional set of predictors extending [1]. The method excludes irrelevant information and operates within an enhanced framework capable of handling nonlinear dependencies. The method is computationally efficient and demonstrates strong empirical performance, particularly over longer forecast horizons.

Model

Notations: $\boldsymbol{y}$ is target variable, $\boldsymbol{X}$ is predictor matrix, $\boldsymbol{Z}$ is the matrix of proxies for $\boldsymbol{y}$. $\mathbf{F}_t$ is matrix of factors, $\mathcal{K}(\cdot,\cdot)$ is a kernel function. $\boldsymbol{J}_T \equiv \boldsymbol{I}_T - \frac{1}{T}\iota_T\iota_T'$ is the demeaning matrix, where $\boldsymbol{I}_T$ is the T-dimensional identity matrix and ι_T the T-vector of ones.

Data Transformation Let $\varphi : \boldsymbol{X} \to \mathcal{F}$ denote a transformation of the original data into a higher-dimensional space (Hilbert space) containing the original set of predictors and their non-linear transformations. $\mathcal{F}$ is M dimensional space and $\boldsymbol{X}$ is N dimensional input, $M >> N$. Number of sample size is T.

The Procedure

Pass	Description
1.	Run time series regression of $\varphi_j(\mathbf{x})$ on $\boldsymbol{Z}$ for $j = 1,\ldots,M$, $\varphi_j(\boldsymbol{x}_t) = \tilde{\phi}_{0,j} + \boldsymbol{z}_t'\tilde{\boldsymbol{\phi}}_j + \hat{v}_{1jt}$, retain slope estimate $\tilde{\boldsymbol{\phi}}_j$.
2.	Run cross-section regression of $\varphi(\boldsymbol{x}_t)$ on $\tilde{\boldsymbol{\phi}}$ for $t = 1,\ldots,T$, $\varphi_j(\boldsymbol{x}_t) = \tilde{\boldsymbol{\phi}}_j'\hat{\boldsymbol{F}}_t + \hat{v}_{2jt}$, retain slope estimate $\hat{\boldsymbol{F}}_t$.
3.	Run time series regression of y_{t+h} on predictive factors $\hat{\boldsymbol{F}}_t$, $\hat{y}_{t+h} = \hat{\beta}_0 + \hat{\boldsymbol{F}}'\hat{\boldsymbol{\beta}}$, delivers the forecast.

Architecture

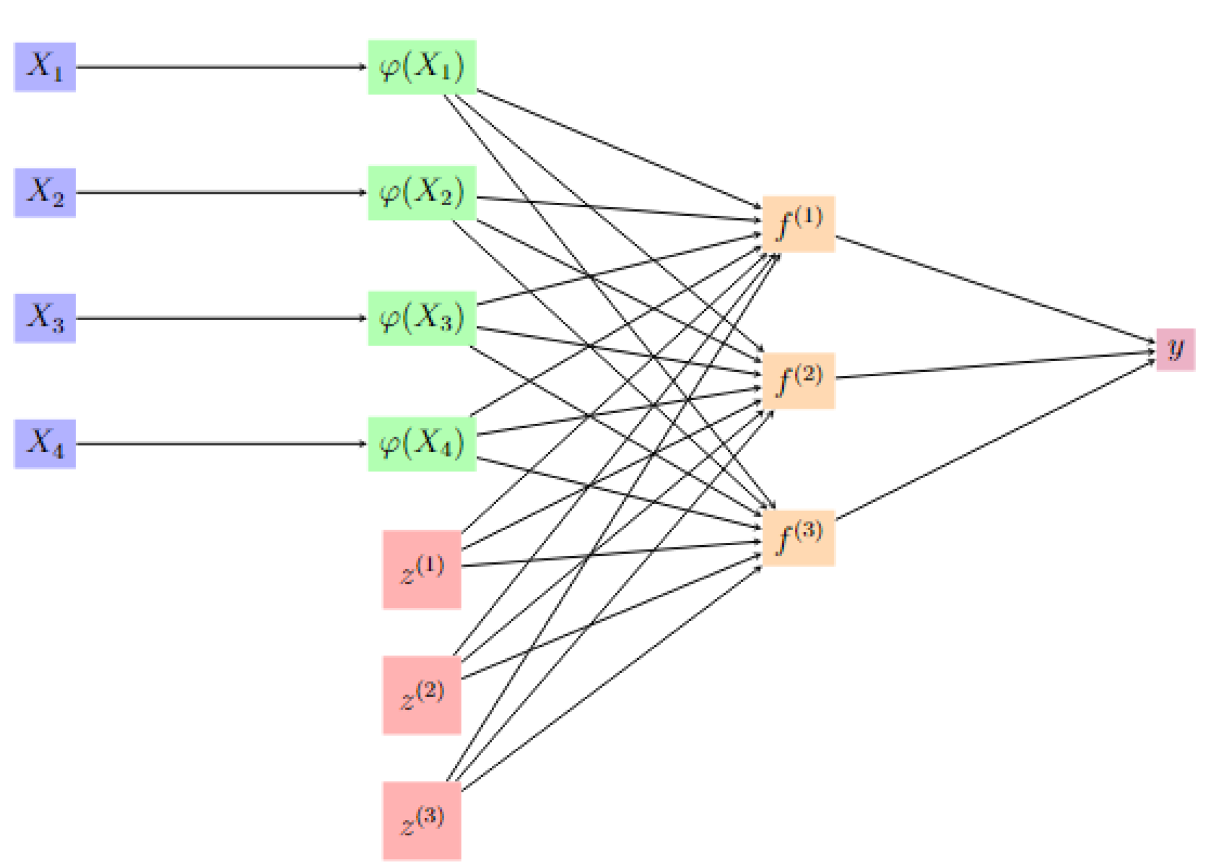

Fig. 1: Kernel 3PRF with T=4 and L=3 relevant factors

Closed Form Expression of Forecast

$$\hat{\boldsymbol{y}} = \iota\bar{y} + \boldsymbol{J}_T\mathcal{K}(\boldsymbol{X},\boldsymbol{X}')\boldsymbol{J}_T\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{J}_T\mathcal{K}(\boldsymbol{X},\boldsymbol{X}')\boldsymbol{J}_T\mathcal{K}(\boldsymbol{X},\boldsymbol{X}')\boldsymbol{J}_T\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{J}_T\mathcal{K}(\boldsymbol{X},\boldsymbol{X}')\boldsymbol{J}_T\boldsymbol{y}$$

Short-Horizon Out of Sample (OOS) Forecasting

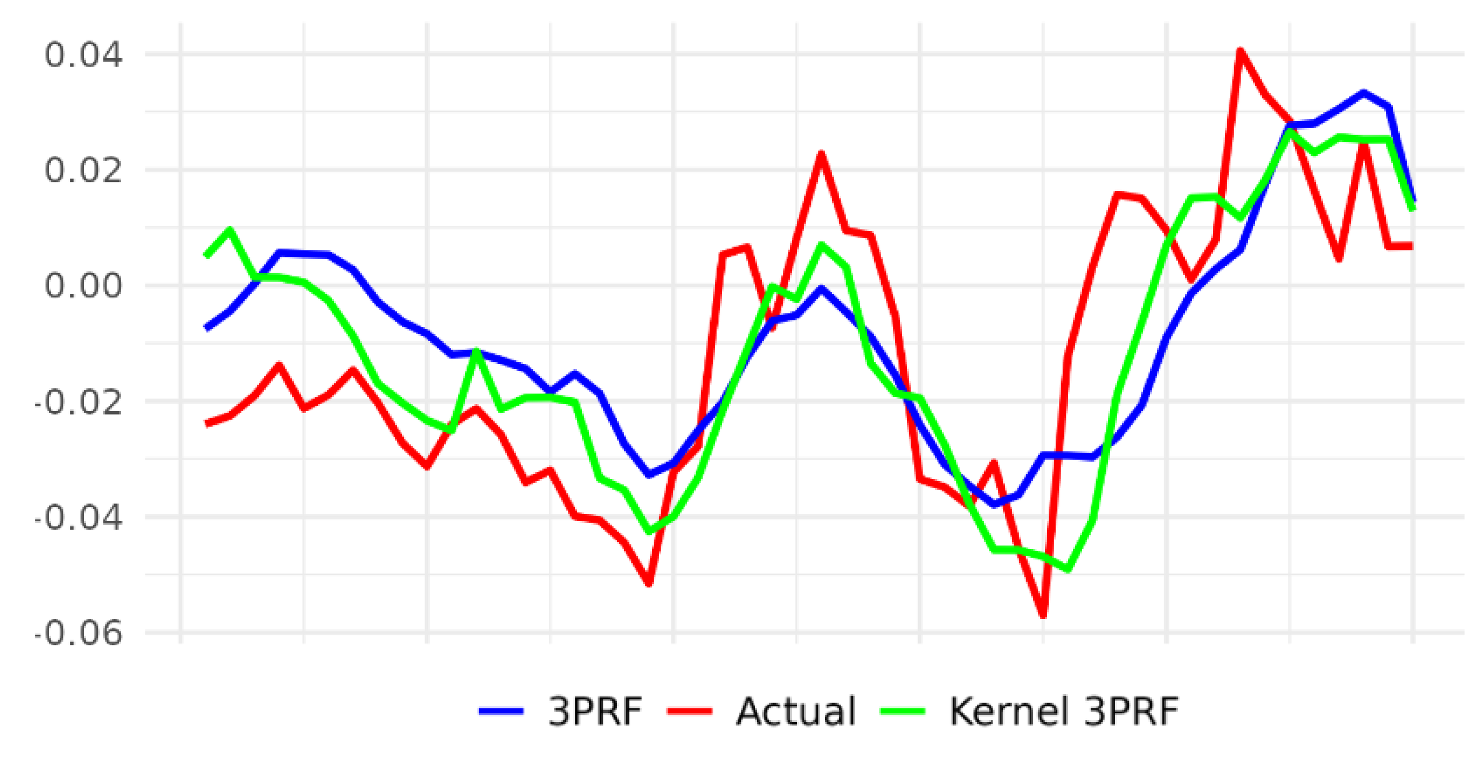

Fig. 2: One-step Ahead Forecast for GDP Deflator

Long-Horizon OOS Forecasting

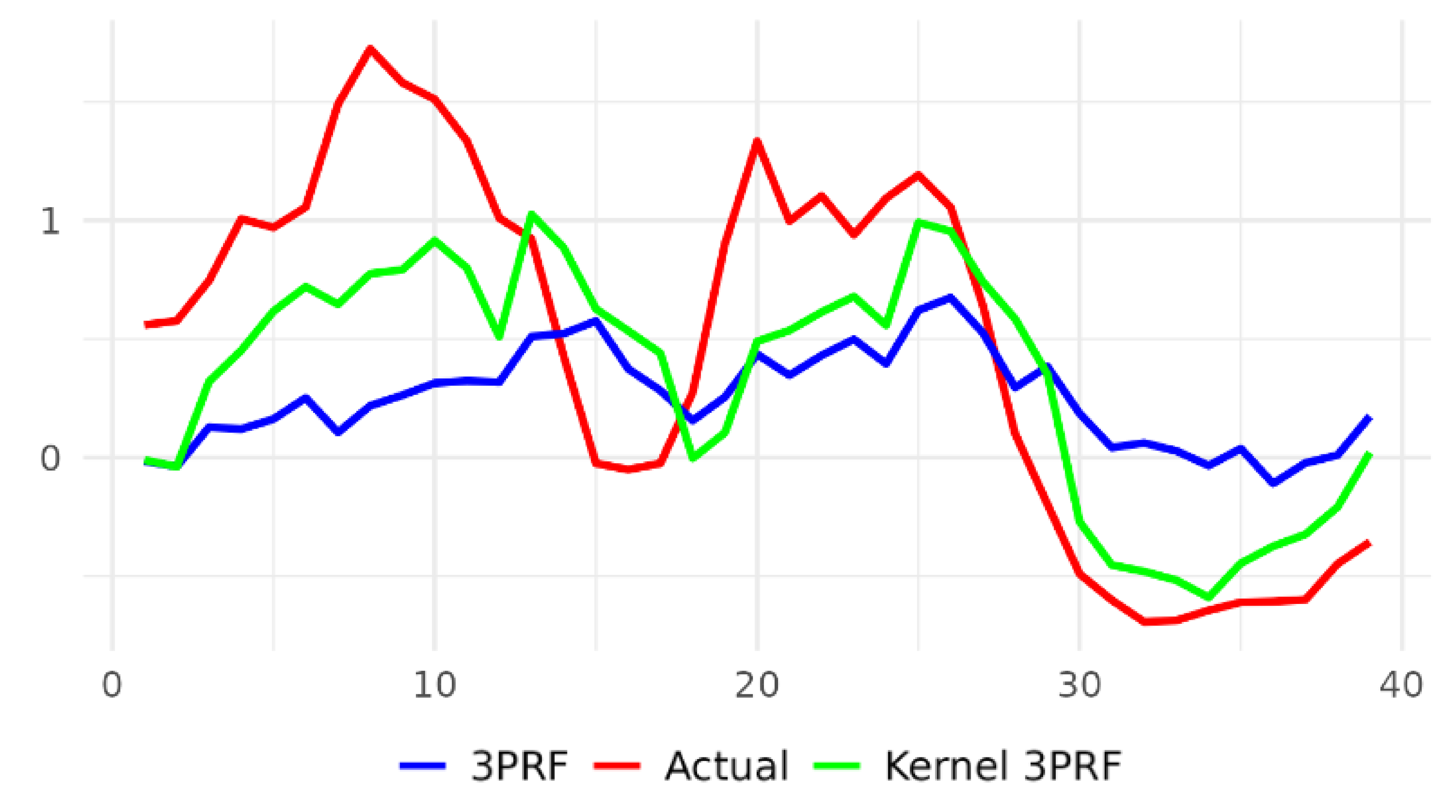

Fig. 3: 12-step Ahead Forecast for S&P500 Index

All-Horizon OOS Forecasting

[PC: Principal Component, Reg: Regression, Sq-PC: square input data then take its PCs. PC-Sq: square of PCs, kPCA: Kernel PCA, 3PRF: [1]'s method, k3PRF: our method]

***Consumer Price Index (CPI)** : h-period ahead Out of Sample R^2*

Method	h=1	h=2	h=4	h=6	h=8	h=10	h=12
AR	**0.704**	**0.706**	**0.565**	0.397	0.211	0.062	-0.038
PC	0.660	0.535	0.154	-0.163	-0.252	-0.248	-0.173
Sq-PC	0.410	0.296	0.049	-0.055	-0.156	-0.200	-0.173
PC-Sq	0.649	0.512	0.186	-0.019	-0.087	-0.187	-0.228
kPCA	0.440	0.380	0.189	-0.043	-0.024	0.042	-0.006
3PRF	0.641	0.566	0.352	0.192	0.241	0.255	0.141
k3PRF	0.676	0.612	0.463	**0.469**	**0.434**	**0.349**	**0.477**

Best Forecasting Methods on 176 US Series

Analysis: Comparative performance of models across a total of $176 \times 8 = 1408$ target-horizon combinations on 176 US time series in our FRED-QD data compiled by McCracken & Ng. Our sample runs from 1964 to 2007.

Best Method Definition:A method is considered 'best' under tolerance level ε if its out-of-sample R^2 is within ε percentage of the top method's R^2. For non-zero tolerance, multiple methods can be 'best'.

Distribution of Best Forecasting Methods Across All Series (Percentage)

Analysis	Tolerance(%)	Methods						
		AR(2)	PC	Sq-PC	PC-Sq	kPCA	3PRF	k3PRF
All Horizons								
	0	48.22	0.21	0.85	1.42	2.98	6.47	39.56
	5	50.07	1.14	1.35	1.99	3.34	9.16	43.54
	10	52.41	2.27	2.13	3.34	4.26	13.07	48.37
	20	55.68	5.68	3.69	7.74	6.75	23.30	62.57
Short-horizon								
	0	84.09	0.14	0.43	0.57	0.43	1.70	12.64
	5	87.07	1.42	0.71	1.56	0.57	5.11	18.75
	10	90.77	3.27	1.70	3.84	1.28	9.23	26.14
	20	94.32	8.38	3.41	10.37	3.55	20.03	48.72
Long-horizon								
	0	12.36	0.28	1.28	2.27	5.54	11.79	66.48
	5	13.07	0.85	1.99	2.41	6.11	13.21	68.32
	10	14.06	1.28	2.56	2.84	7.24	16.90	70.60
	20	17.05	2.98	3.98	5.11	9.94	26.56	76.42
Excluding AR								
	0	-	1.42	1.56	2.84	5.47	13.00	75.71
	5	-	2.84	2.06	4.76	5.75	17.97	78.76
	10	-	5.26	3.27	7.74	7.03	25.99	81.53
	20	-	11.08	5.89	14.35	11.43	41.34	86.08

Conclusion

We demonstrate that this approach is a reliable forecasting tool, with its improved performance stemming from two key features: it captures non-linear relationships by transforming input data into a higher-dimensional space. It operates as a supervised method, filtering out irrelevant factors.

Forecast Convergence Rate

Under Assumptions (given in the paper, almost same as [1]), we have

$$\hat{y}_{t+h} - \mathbb{E}_t y_{t+h} = O_p(\min\{M, T\})$$

Miscellaneous

We use the rolling window method to find OOS R^2. We use cross-validation to select tuning parameters.

References

[1] Bryan Kelly and Seth Pruitt. "The three-pass regression filter: A new approach to forecasting using many predictors". In: *Journal of Econometrics* 186.2 (2015), pp. 294–316.