High Frequency Inflation Estimation and Forecasting

Sonan Memon
Pakistan Institute of Development Economics, Islamabad

August 2022

Introduction

i. I begin by motivating the utility of high frequency inflation estimation and review recent work from the State Bank of Pakistan for inflation forecasting and now-casting.

ii. I also present stylized facts about the structure of historical and especially recent inflation trends in Pakistan.

iii. However, available data and already used methods cannot achieve high frequency forecasting.

iv. I discuss 3 novel techniques from recent literature including web scrapping, scanner data and synthetic data.

Introduction

i. I generate *synthetic* data using generative ML models (Gaussian Copula and PAR models) and numerical analysis (cubic spline interpolation).

ii. I use cubic splines to estimate monthly inflation rate from quarterly data and unknown weekly inflation rate from actual monthly data.

iii. Meanwhile, I use a probabilistic auto-regressive ML model to forecast future short-run inflation for Pakistan from 2020 to 2023.

iv. I evaluate the accuracy of forecasts by comparing them with forecast error variances and predictions from conventional reduced form vector auto-regressive models (VAR).

Motivation

i. Accurate forecasting of inflation is a concern for market players, central banks, and also governments.

ii. Governments have an incentive to make inflation control a priority and interfere with central bank independence around election periods i.e the *political business cycle*, initiated by Nordhaus (1975).

iii. For instance, Abrams and Butkiewicz (2012) revealed evidence from the Nixon tapes that President Nixon manipulated Arthur Burns and the Federal Reserve into creating a political business cycle.

Motivation

i. While central banks do collect data on consumer price indices, the frequency of collection does not allow accounting for sudden swings in inflation.

ii. However, when for instance, in a matter of few weeks, the Ukraine and Russian crisis changed the inflation landscape, conventional price indices had little forecasting potential.

iii. Similarly, inflation shocks can result from sudden change of central bank governors or government change, terrorism episodes or political turmoil, climate catastrophes, especially in developing economies Vuletin and Zhu (2011).

Research at SBP

i. The State Bank of Pakistan's (SBP) research department has done some work on inflation forecasting by using ML methods (e.g Neural Networks) and monthly year on year (YoY) inflation in Pakistan Hanif et al. (2018).

ii. Similarly, the SBP's research team has worked on now-casting GDP using data on large scale manufacturing growth in Pakistan Hussain et al. (2018) and LASSO type ML methods.

iii. This is in line with emerging methodologies among central banks worldwide, which are all moving toward big data and machine learning methods Doerr et al. (2021).

SBP Now-Casting

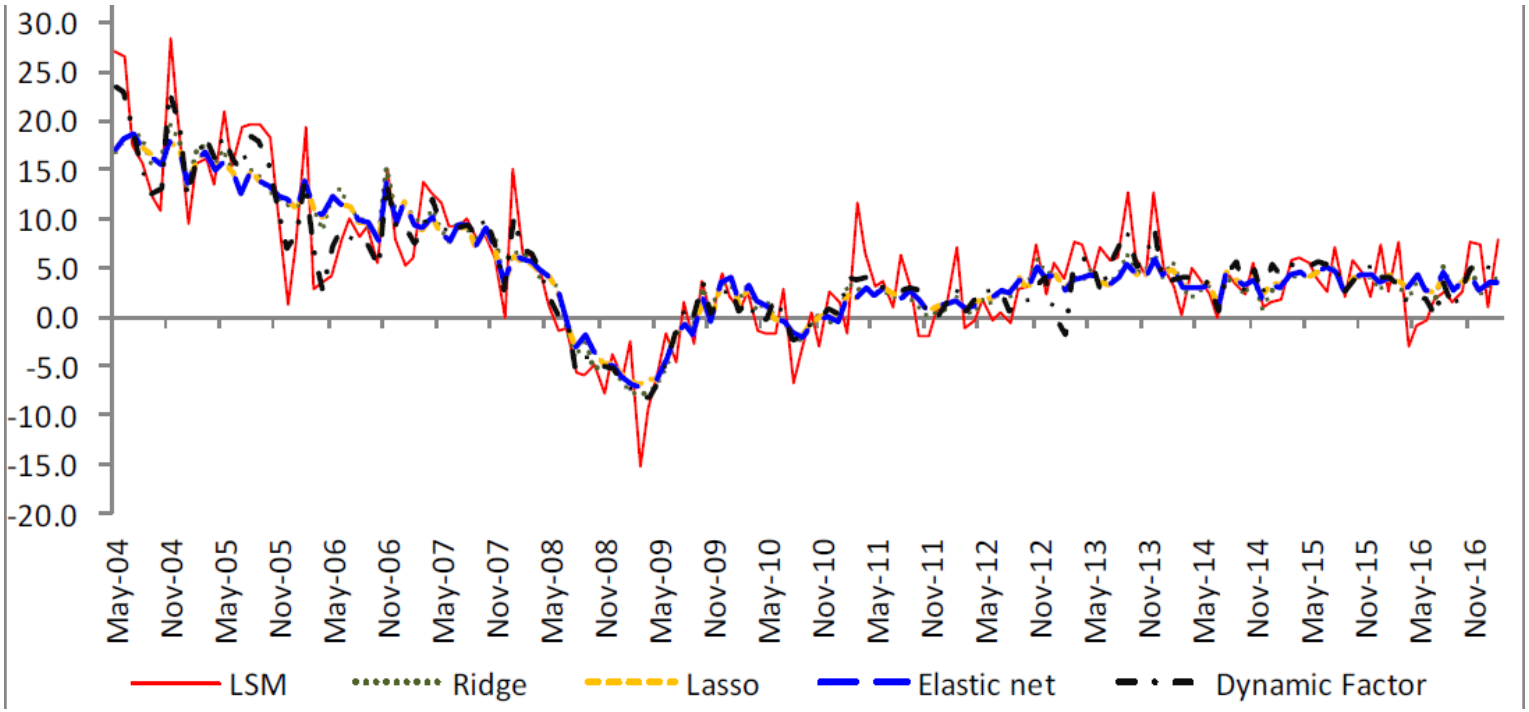


Figure: Source is Hussain et al. (2018)

Modern Forecasting Methods

i. *Web scrapping*: e.g the daily consumer price index (CPI) produced by the Billion Prices Project of Cavallo and Rigobon (2016) offers a glimpse of the direction taken by consumer price inflation in real time.

ii. Another branch of emerging literature uses *scanner data* from super markets data Beck et al. (2020) on prices.

iii. *Synthetic data* is artificially generated to mimic key information of the actual data and provide the ability to draw valid statistical inferences. It overcomes privacy, confidentiality and cost of data collection constraints (see Raghunathan (2021)).

Monthly Inflation in Pakistan

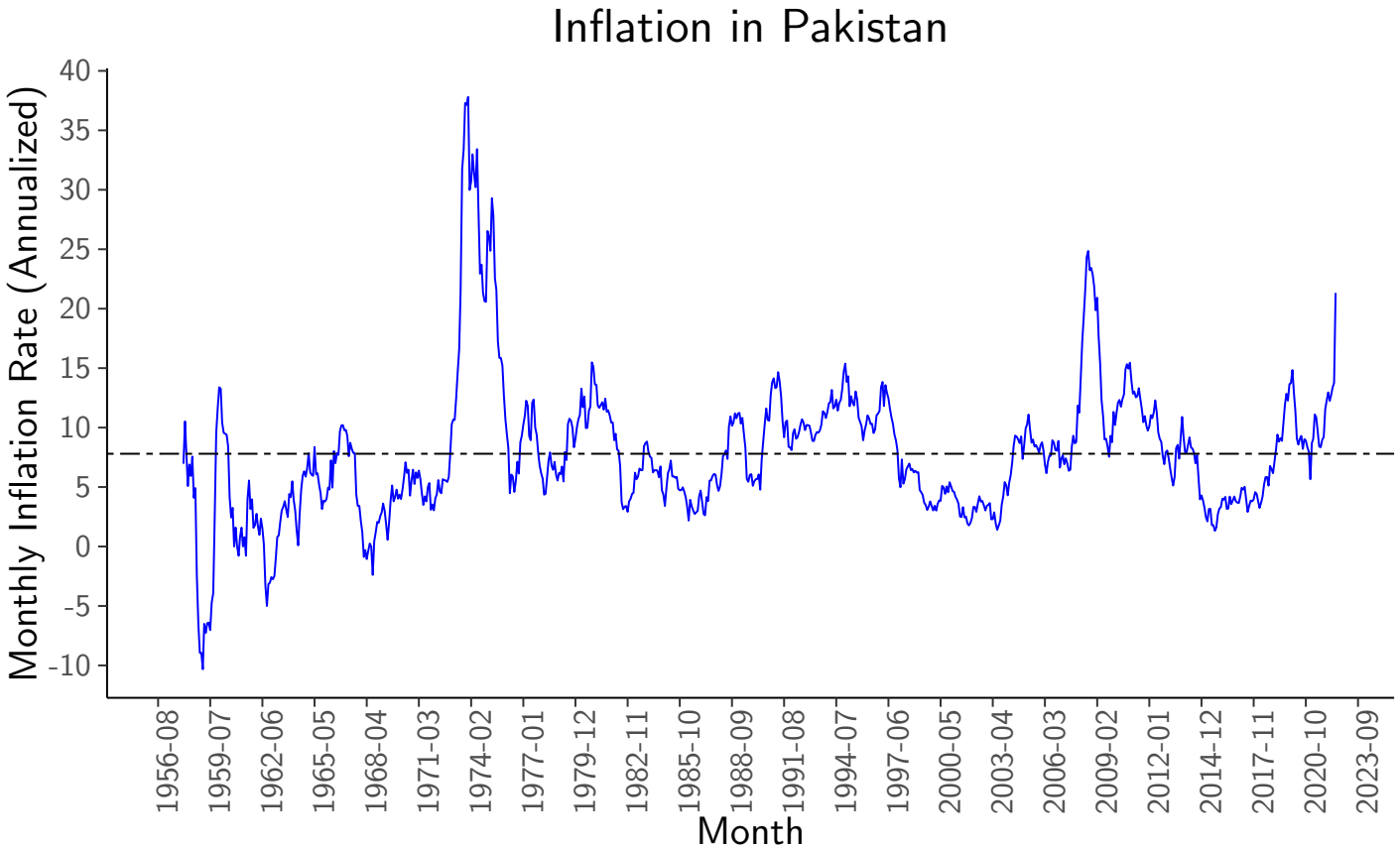


Figure: Data is From IMF

Recent Monthly Inflation in Pakistan

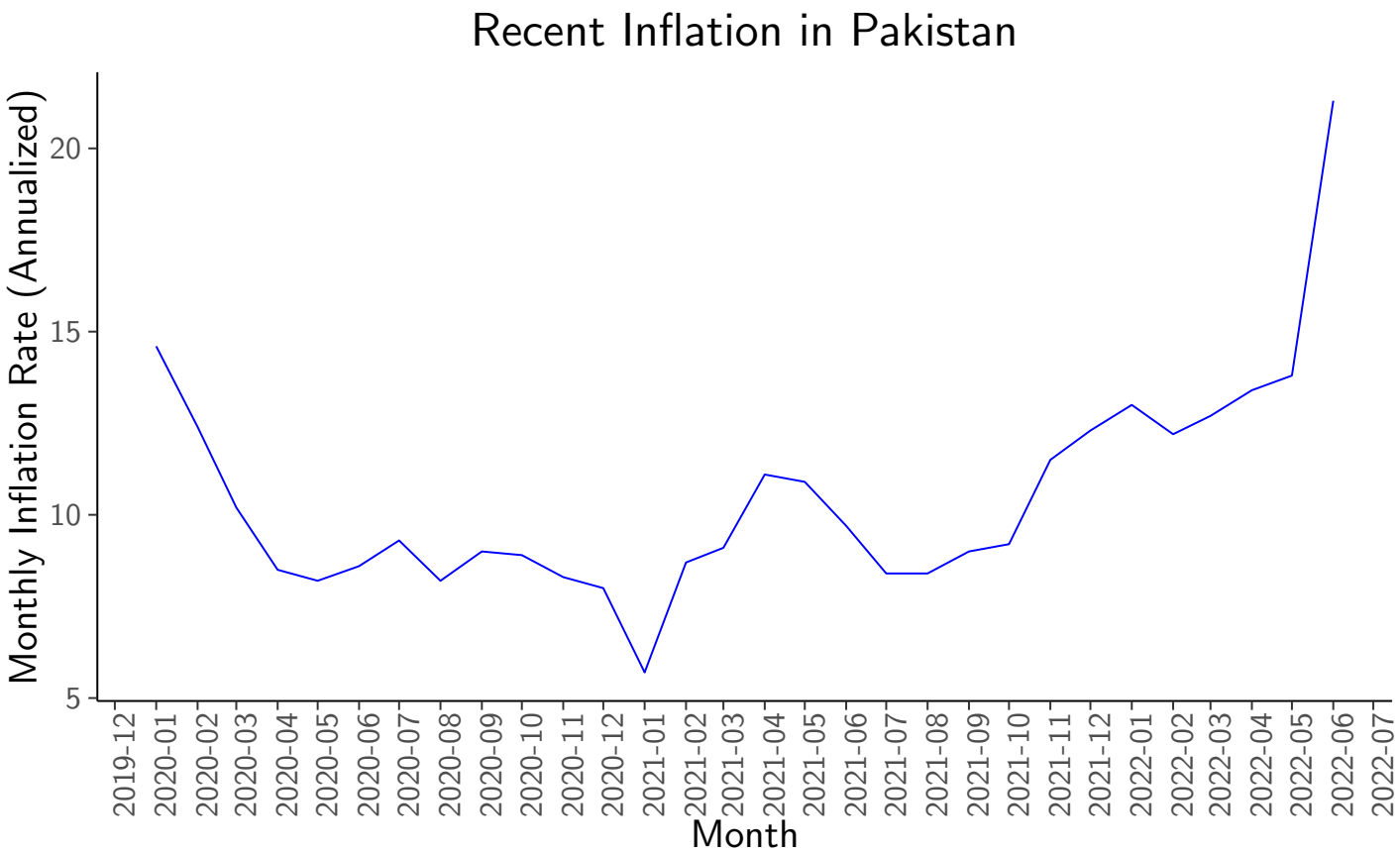


Figure: Data is From State Bank of Pakistan

Annual Inflation Trends

Table: Annual National Inflation (December 2017 to December 2021)

Year on Year (%)	Dec 2017	Dec 2018	Dec 2019	Dec 2020	Dec 2021
Categories					
Headline Inflation	4.6	5.4	12.6	8	12.3
Food Inflation	3.8	0.6	17.9	12.9	10.6
Core Measure (NFNE)	5.5	7.64	7.7	6.4	8.5
Clothing	3.6	6.3	9.8	9.7	11.2
Health	10.9	7.1	11.3	8.1	9.4
Transport	4.5	18.4	14.7	-3.5	24.1
Education	12.4	9.8	6	1.3	2.8

Note: Data is from State Bank of Pakistan.

Methodology: Synthetic Data From Copula

i. A copula C in space $\mathbb{R}^n$ is a multivariate CDF (cumulative density function) supported by the unit hyperplane $[0,1]^N$ with the property that all of its marginals are uniformly distributed on $[0,1]$. Formally, C is the function of the form below, where $0 \leq s_n \leq 1$ and $u_n \sim U[0,1], \forall n$.

$$C(s_1, s_2, s_3, ..., s_N) = \mathbb{P}\{u_1 \leq s_1, ..., u_N \leq s_N\} \qquad (1)$$

ii. Synthetic data from probabilistic auto-regressive models.

iii. Cubic spline interpolation.

Synthetic Data From Gaussian Copula

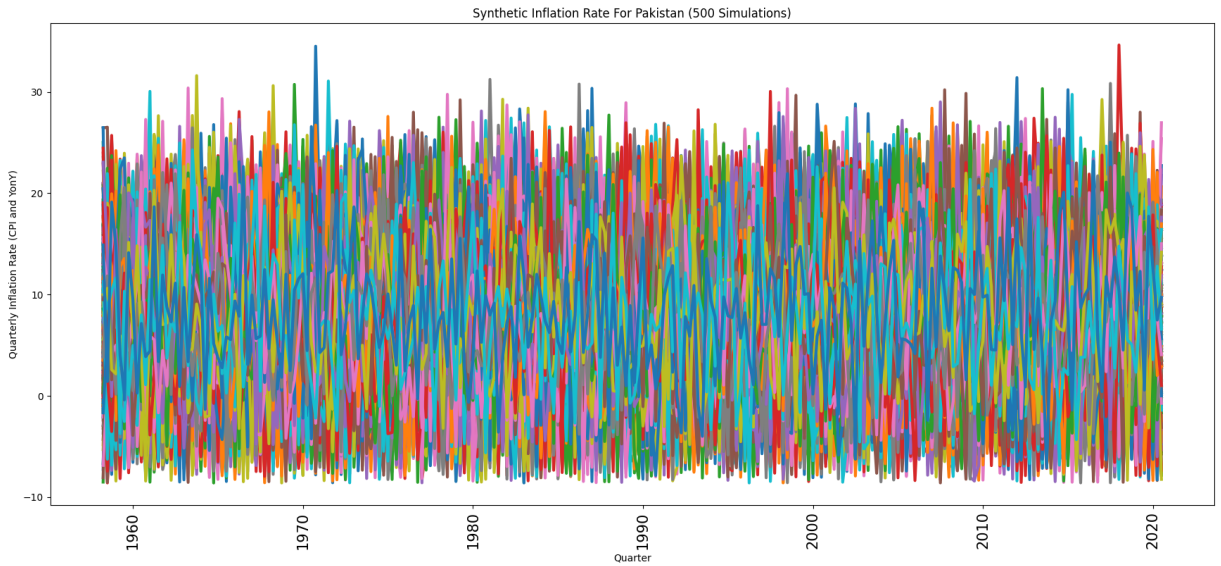


Figure: Author's Simulations

Methodology: Synthetic Data From PAR

i. Probability Auto-Regressive Model (PAR) is a synthetic data creation methodology which is well suited for time series models and accounts for the auto-correlation structure of time series data.

ii. Assume that we are given access to a data set $\mathcal{D}$, consisting of n-dimensional data points x. For simplicity, let us assume that the data points are binary, i.e. $x \in (0,1)^N$. By the chain rule of probability, we can factorize the joint distribution over the n-dimensions as:

$$p(x) = \prod_{i=1}^{n} p(x_i | x_1, x_2, .., x_{i-1}) \tag{2}$$

Synthetic Data From PAR Model

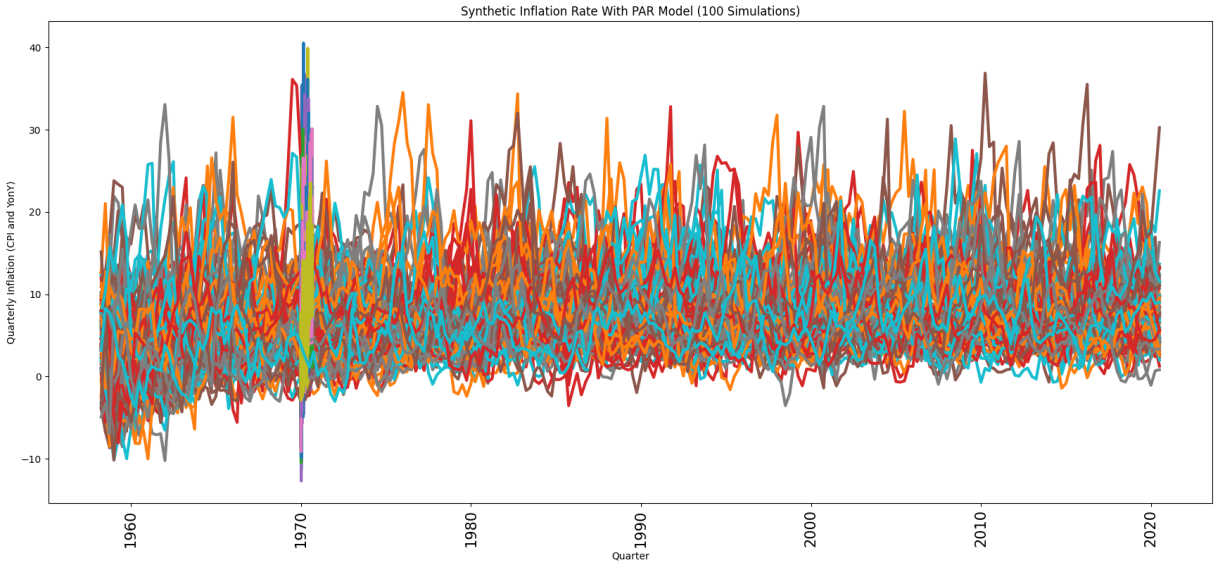


Figure: Author's Simulations

Methodology: Cubic Spline Interpolation

i. Cubic spline interpolation is an interpolation method which uses cubic polynomials to connect the existing data nodes, which allows estimation of unknown and high frequency intermediate data points.

ii. For a mathematical and formal review of cubic spline interpolation, you can refer to Burden et al. (2015).

Interpolated Weekly Inflation Series

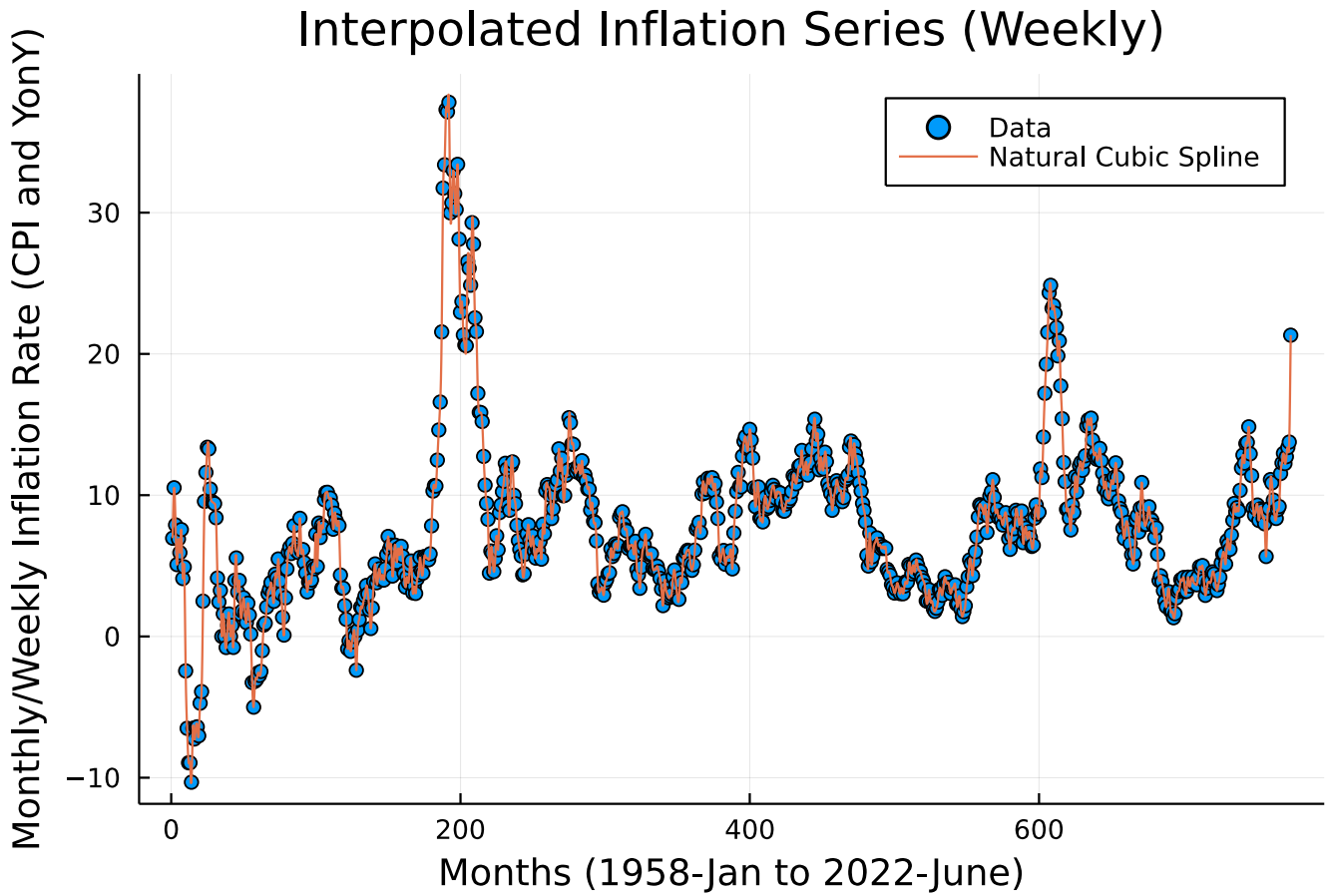


Figure: Using Monthly Inflation to Interpolate Weekly Inflation

Fan Charts

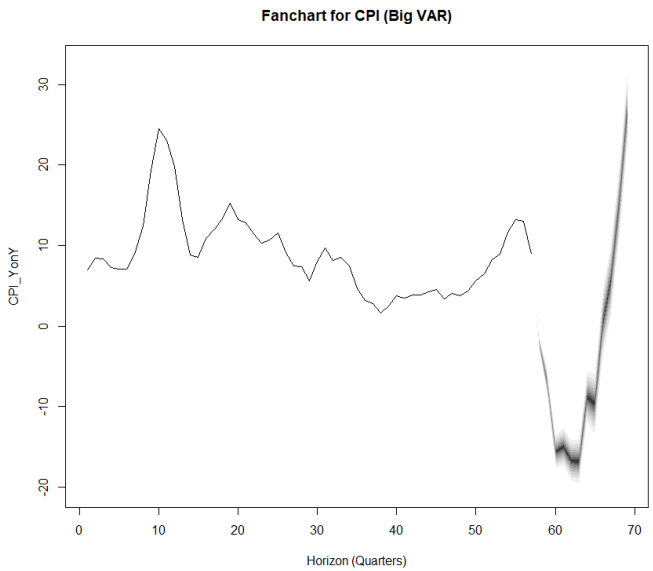
i. The following fan charts are based on forecasts from simple reduced form VAR models for Pakistan.

ii. I estimate two versions of The VAR model; the first one is the Big VAR which includes variables such as CPI, short term external debt measure, M2, tax revenue, imports and SR rate (short term interest rate) variables for Pakistan.

iii. As a robustness check, I also estimate a simpler VAR with the variables CPI and imports only. The models are estimated for the period from 2006Q2 to 2020Q2 due to restrictions on availability of data.

iv. Standard information criteria are used to select lag orders and forecasts are evaluated for 12 quarters after 2020Q2, starting from 2020 Q3 and ending at 2023 Q3.

VAR Fan Charts



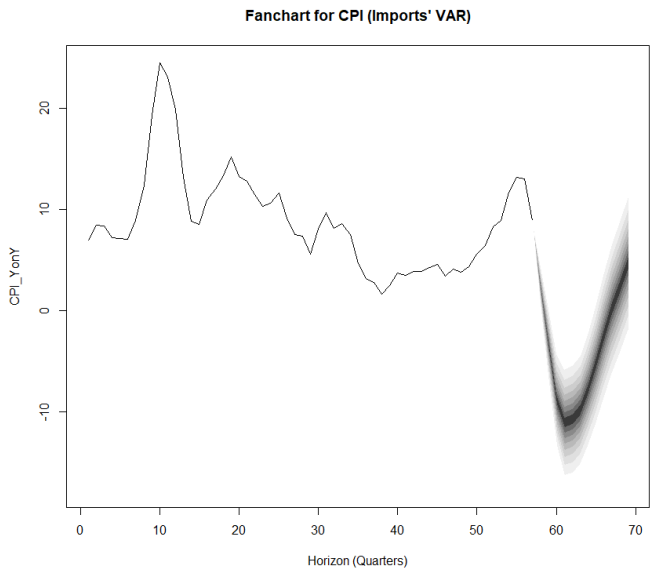


Figure: Fan Charts Using VAR Models (Big and Imports Only)

FEVD (Forecast Error Variance Decomposition)

i. I also present FEVD (forecast error variance decomposition) results for the Big VAR model.

ii. The horizon for FEVD is 12 quarter ahead and it demonstrates that non-CPI variables drive a larger share of the forecast error at longer horizons.

iii. For instance, non-CPI variables explain more than 70% of the forecast error variance at the 8th quarter ahead forecasts and almost 80% of the error variance at 12th quarter ahead.

iv. Meanwhile, for the first 3 quarters, CPI explains more than 60% of the forecast errors.

FEVD

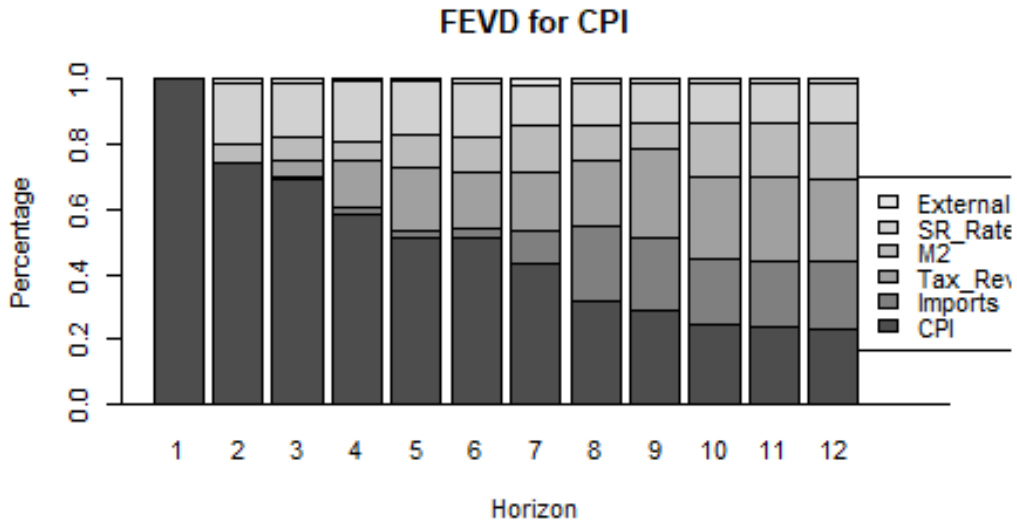


Figure: Forecast Error Variance Decomposition For VAR Big

Forecasts From PAR Models

i. In order to generate forecasts based on PAR models, I use new and currently evolving estimation techniques developed by Alexandrov et al (2019).

ii. To maintain consistency with VAR model from last section, I apply the probabilistic auto-regressive model (PAR) on inflation series of Pakistan, starting from 2006Q2 and ending at 2020Q3.

iii. The top graph in next slide presents 12 quarter ahead forecasts from the PAR model, starting from 2020Q4 to 2023 Q4.

iv. The bottom graph of the same figure presents 36 month ahead forecasts from October 2020 to October 2023.

Forecasts From PAR Model

Figure: 12 Quarter/36 Month Ahead Forecasts From PAR Models

Forecasts From PAR Models

i. The PAR model is more accurate than the VAR forecasts for both the quarterly and monthly specifications of PAR.

ii. However, the forecasts based on quarterly inflation series are more accurate predictors of the hyperinflation crisis of 2021 and 2022.

iii. For 2021 and 2022, the monthly PAR model can predict a rise in inflation which is close to 15% but the quarterly data predicts close to 20% inflation level.

iv. However, monthly PAR model predicts a stabilization and downward adjustment of inflation in 2023, which is close to 10% by October 2023. Given the current crisis, the monthly PAR model may be defeated by the quarterly model.

Forecasts From PAR Models

i. My results are robust to estimation based on entire time series period from 1958 to 2020 for both quarterly and monthly data in addition to variations in size of testing data cuts.

ii. My default cut for training data is 11 quarters in quarterly estimation and 69 months for 22 monthly data.

iii. For robustness, I evaluate the forecasts for training data with 33 (11), 69 (23) and 105 (35) monthly (quarterly) cuts for monthly (quarterly) data respectively.

iv. In all these cases, the forecast accuracy is superior to the VAR models.

Conclusion

i. In this working paper, I review novel methodologies for inflation forecasting, while being motivated by the current inflation crisis of Pakistan.

ii. More specifically, I mainly use probability autoregressive models and cubic spline interpolation.

iii. In addition, I use standard, reduced form, vector auto-regressive models to forecast inflation and compare the forecasting potential of VAR versus my ML model forecasts.

Conclusion

i. It may be fruitful to expand the data availability by collecting web-scrapped and scanner data on high frequencies for Pakistan.

ii. Moreover, another extension could be to use a combination of ML based methods and VAR's (Vector Autoregressive), which would be an extension of the machine learning model used in this paper

Thank you

Q and A