Macroeconomic Indicator Forecasting with Deep Neural Networks

Tom Cook & Aaron Smalter Hall

Jul 12, 2018

Motivation

- Forecasting is essential to bank
- We want forecasts to be "good"
- Some dimensions to consider:
  - **Accuracy**
  - **Responsiveness**
  - Also: Precision, Effort to produce, Data Requirements, Model Dependence
- For forecasts of macro indicators, room for improvement in several areas

This Paper

- Deep learning represents opportunity to improve on forecasts
- Upshot:
    - We test several neural network architectures
    - Find most work well in near term
    - Encoder-decoder brings good performance at up to 4 quarters

What Is a Neural Network?

- Developed in 1950s, abandoned by 70's, revitialized in 2000/2010's
- A model comprised of one or more neurons. Each neuron is a linear combination of inputs transformed through an activation function.
- neurons connected in different network structures (architectures)

- Neural networks are universal function approximators
  - We don't need to specify a model, we just need to build an architecture with sufficient capacity to approximate the DGP
- Used to produce breakthroughs in accuracy of ML tasks in last few years
  - We expect similar improvements in performance with forecasting

Architectures

- Fully Connected Architecture
  - Most basic form of neural network
  - Several layers of fully connected nodes; each node in layer gets inputs from all nodes in previous layer
- Recurrent (LSTM)
  - Network built for sequence data. Essentially, a network with 'memory'. Leverages temporal order in data.
- Encoder Decoder
  - Two LSTM cells: encoder cell, decoder cell. Encoder cell generates representation of temporally ordered input data (time series). Decoder cell interprets representation and extrapolates to desired forecast horizon.

More about architectures

Data Choices

- Focus on unemployment
  - Limited revision – limits 'correct vintage' problems
  - Mean-reverting
  - Sufficiently Long
- Input:
  - Civilian unemployment rate (UNRATE)
    - Pulled from FRED (Aug 2017)
  - Monthly values used, quarterly predictions generated
  - last 36 months UNRATE
  - First, Second Differences
- Target:
  - 0-4 quarter predictions
  - 30 runs of each model

Benchmark: SPF

- Survey of Professional Forecasters
  - Usually about 40 responses for given quarter
  - Individual responses available, Median response is SPF point-estimate
  - Generally, SPF forecasts improve on VAR
  - Allows us to see how our model compares to field of experts
  - Variance in SPF provides additional insight for model comparison
- Compare against Individual performance (average participant performance)
- Compare against Combined (ensemble) SPF Forecast

Individual model comparisons

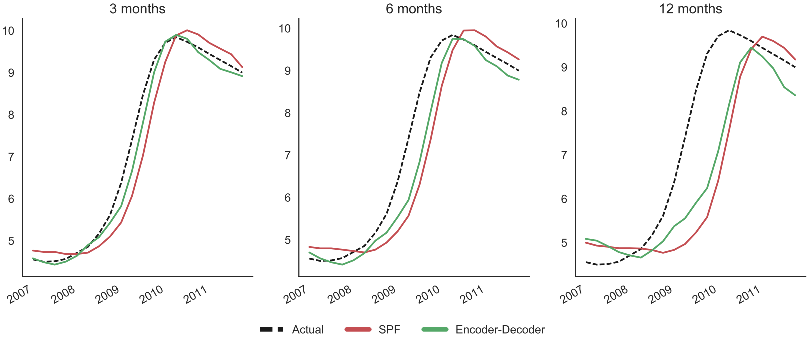
- Assessing Performance:
  - Primarily compare MAE and distribution of MAE over repeated trainings
- Fully Connected and LSTM perform well in short-horizons (0, 3 month)
  - either out-perform or remain competitive with SPF
- Encoder-decoder performs well at all horizons

Architecture Performance vs SPF participants

horiz		Fully Connected	LSTM	Encoder Decoder	DARM	SPF
0 Months	Mean MAE	0.076	0.104	0.044	.117	**0.135**
	St. Dev.	0.027	0.058	0.002		0.063
3 Months	Mean MAE	0.253	0.273	0.184	.328	**0.271**
	St. Dev.	0.029	0.028	0.001		0.093
6 Months	Mean MAE	0.441	0.473	0.305	.493	**0.412**
	St. Dev.	0.041	0.066	0.005		0.157
9 Months	Mean MAE	0.638	0.748	0.461	.658	**0.568**
	St. Dev.	0.048	0.110	0.007		0.228
12 Months	Mean MAE	0.870	1.017	0.620	.907	**0.720**
	St. Dev.	0.049	0.207	0.006		0.291

Error Distributions

Encoder Decoder Response Comparision[1]



[1] Two-period gaussian-window smoother applied

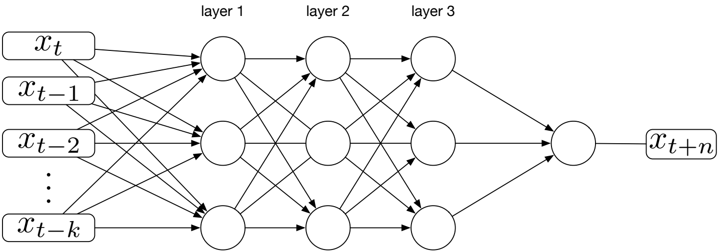
Conclusions

- All models compare well to SPF in short term
  - Suggests robustness to modeling type for short-term forecasting
- Encoder Decoder performed better than SPF at all horizons
  - Improvement in performance appears to come from improvement in responsiveness to shifts in economic conditions

Fully Connected Architecture

- Several layers of fully connected nodes
- Each node in layer gets inputs from all nodes in previous layer
- Each node produces a linear combination of inputs transformed through a transformation function
- Layers can have varying numbers of nodes
- As tested, we include residual connection and dropout

Example fully-connected architecture for univariate timeseries forecast



Recurrent Architecture (LSTM)

- Comes from sequence-based research (temporal/order dependent)
  - e.g. translation, language processing
  - Intuition: memory of past data built into architecture
- Essentially: each node i gets input from i^{th} member of the sequence along with output from node $i-1$
- We use a special recurrent structure - Long Short Term Memory (LSTM) cells
  - Capable of long-run memory
  - Ability to forget allows new parts of sequence to remain relevant

Diagram of LSTM architecture

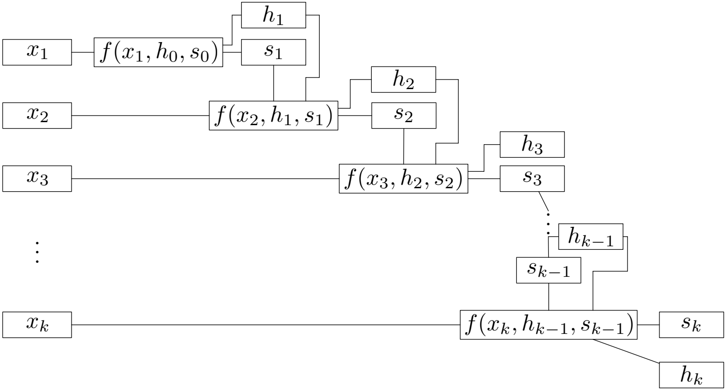
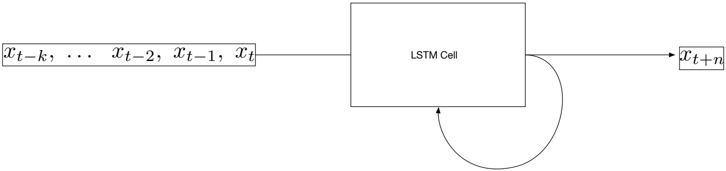
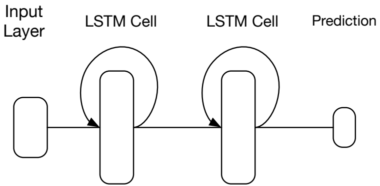


Diagram of consolidated (rolled) LSTM architecture

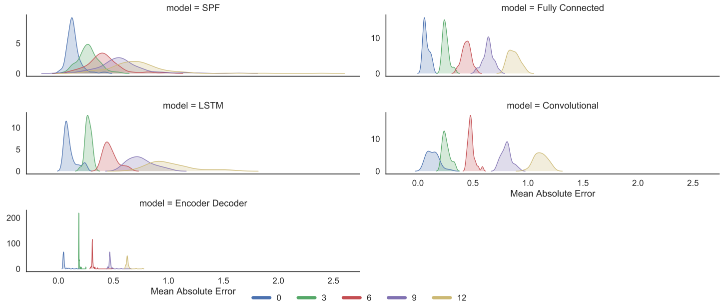


Encoder Decoder

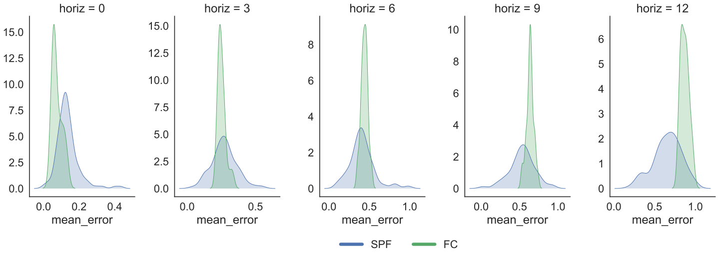
- Inspired by variable-length input, variable-length output tasks (e.g. translation)
  - Sequence-to-sequence models
  - Intuition: Assists with longer-horizon forecasts. Allows model to iteratively extrapolate short-term prediction.
  - Alternatively: Separation of data understanding and forecasting with the data
  - Essentially: Multiple LSTM cells, stacked



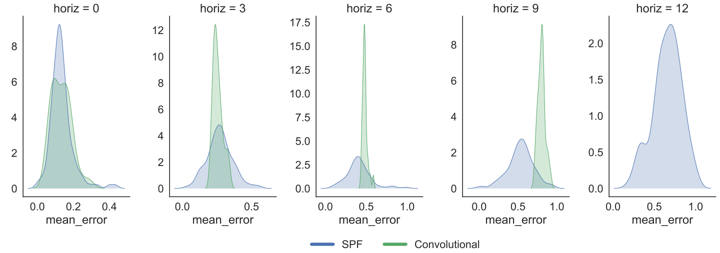
All Error Distributions



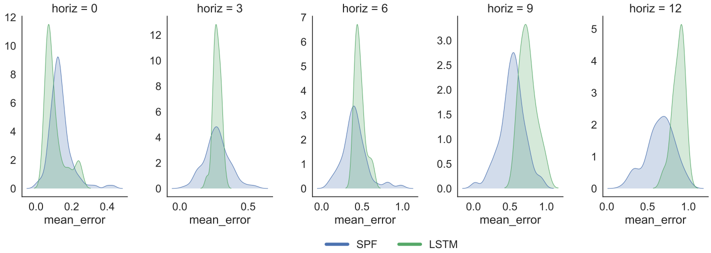
FC vs SPF Participant Distributions



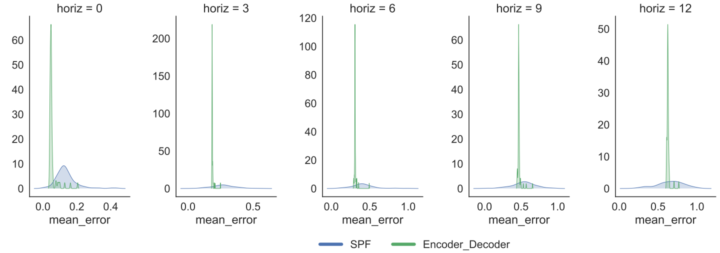
CONV vs SPF Participant Distributions



LSTM v SPF participant distributions



Encoder Decoder vs SPF Participant Distributions



Ensemble Comparisons

- Use perceptron to combine individual runs into single prediction
  - Essentially linear model with individual run predictions as input
- SPF presents single prediction as median forecast

Ensemble Comparisons

Forecast Horizon	Fully Connected	LSTM	Encoder Decoder	SPF
0 month	0.043	0.048	0.041	**0.10**
3 month	0.215	0.242	0.184	**0.23**
6 month	0.414	0.401	0.301	**0.36**
9 month	0.548	0.664	0.459	**0.50**
12 month	0.779	0.900	0.618	**0.63**

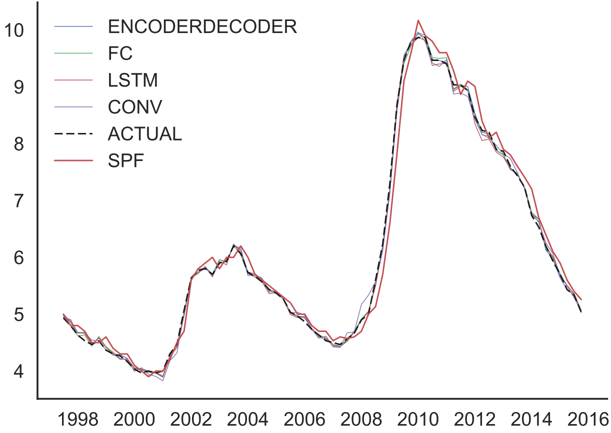
SPF errors for comparision

- SPF is a panel of experts. Each panelist is surveyed multiple times.
- $e = 1/n \sum_i 1/T_i \sum_t |y_t - \hat{y}_{ti}|$
- This tells us average mean absolute error we would expect from single expert over time.
- SPF reports slightly different error statistics – the mean average error of all experts over time – this is the ensemble forecast error.

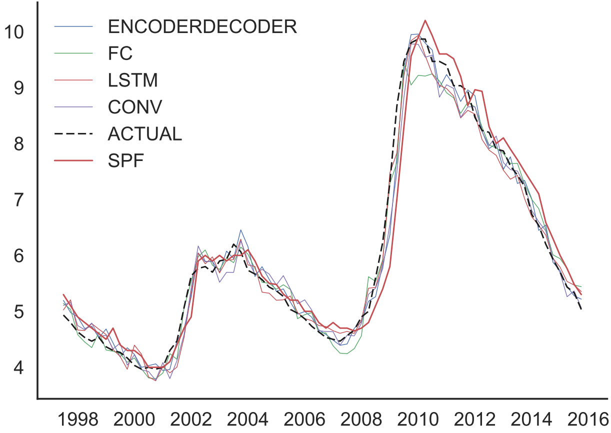
DARM errors

- The directed autoregressive model:
- $\widehat{\text{UNRATE}}_{t+n} = \sum_{i=1}^{k} \beta_i \text{UNRATE}_{t-i}$
- To be contrasted with iterative AR model in which next-step ahead is forecast and then extrapolated to desired forecast horizon.
- DARM error statistics are as reported in the SPF error statistics report.

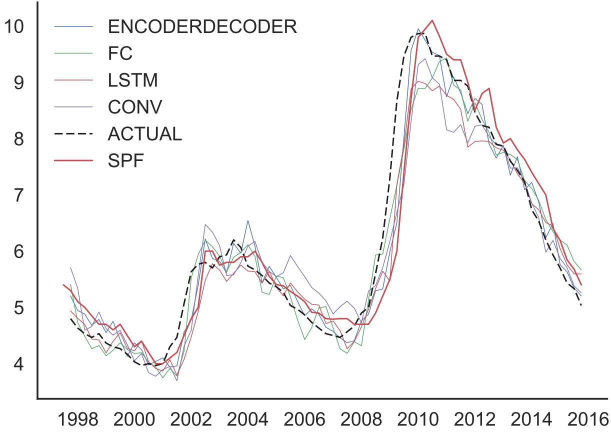
All vs SPF performance (Current Quarter)



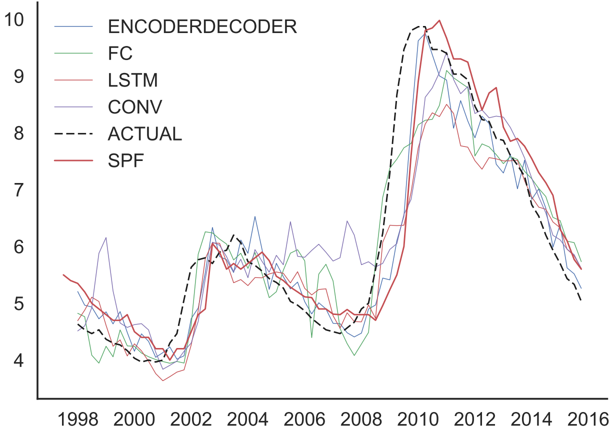
All vs SPF performance (3 months)



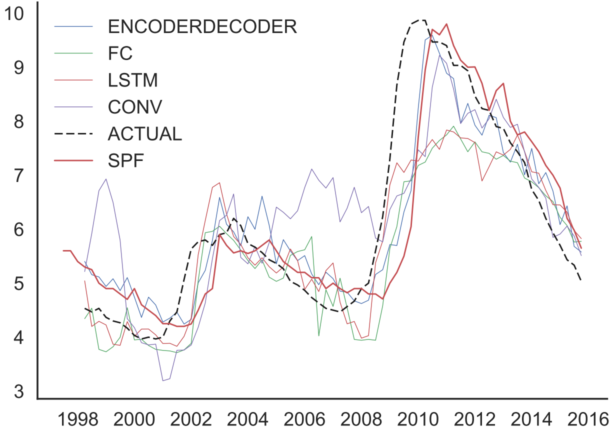
All v SPF performance (6 months)



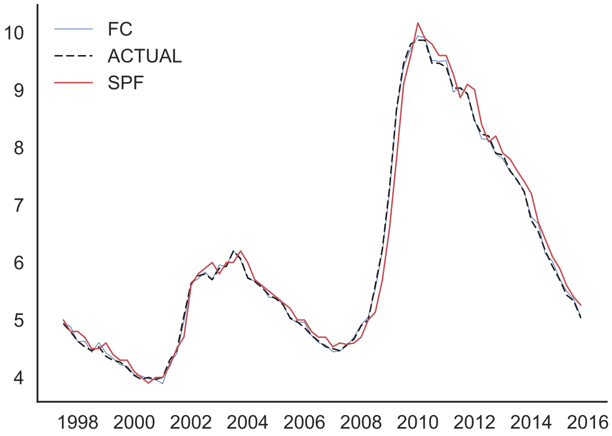
All v SPF performance (9 months)



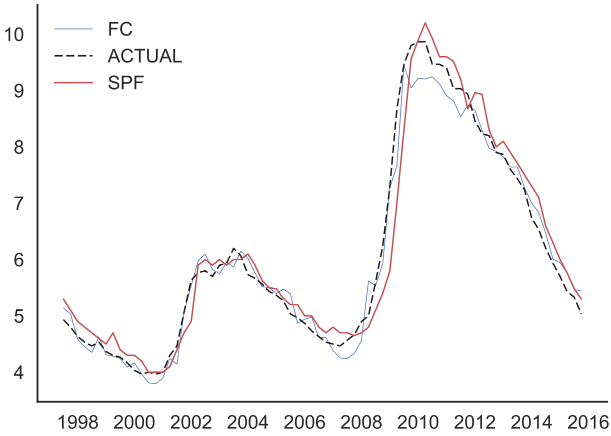
All vs SPF performance (12 Months)



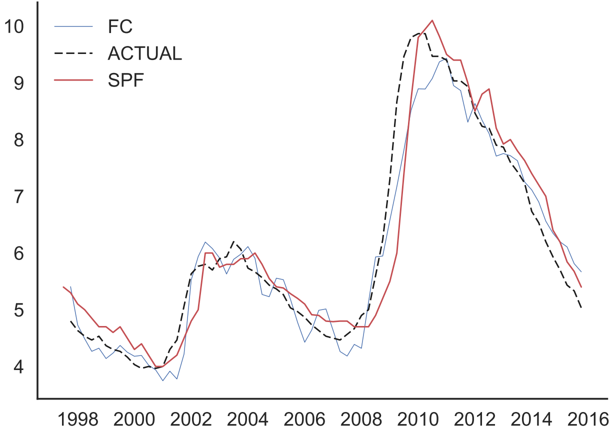
FC vs SPF performance (Current Quarter)



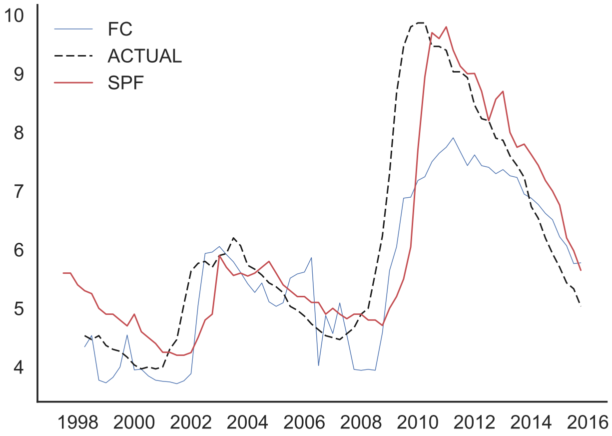
FC vs SPF performance (3 months)



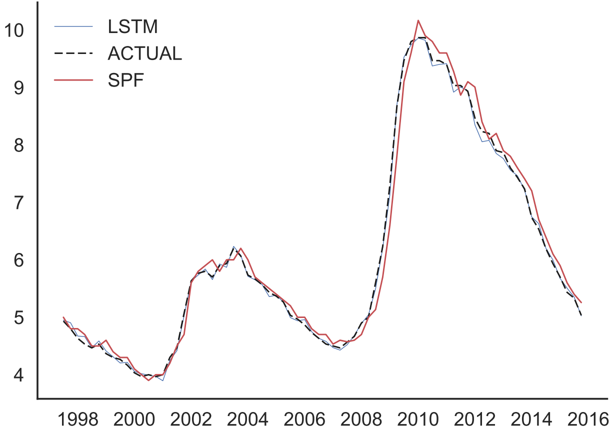
FC v SPF performance (6 months)



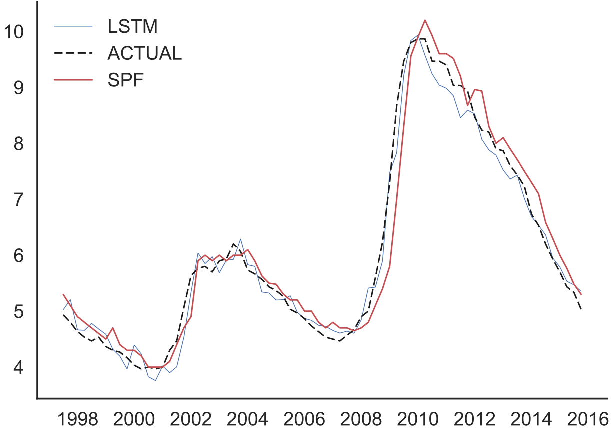
FC vs SPF performance (12 Months)



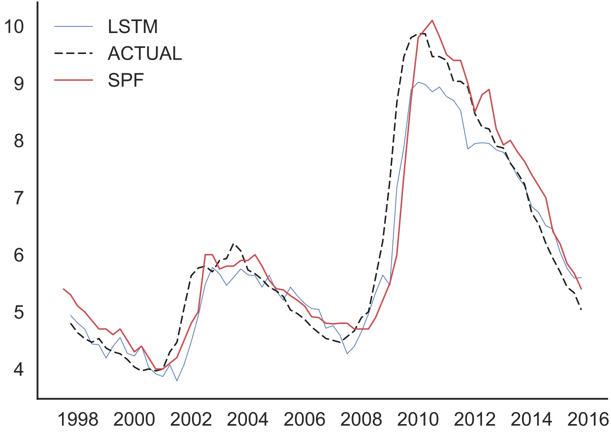
LSTM vs SPF performance (Current Quarter)



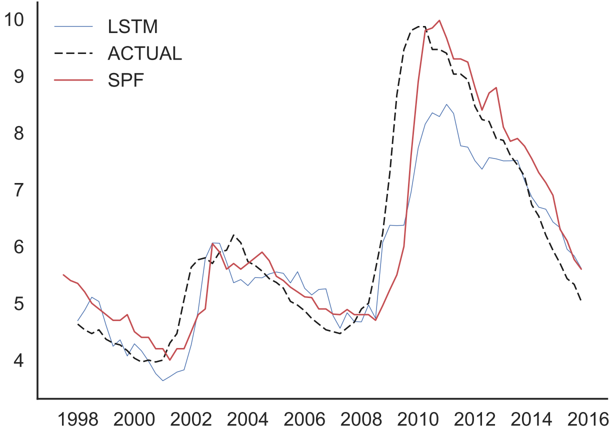
LSTM vs SPF performance (3 months)



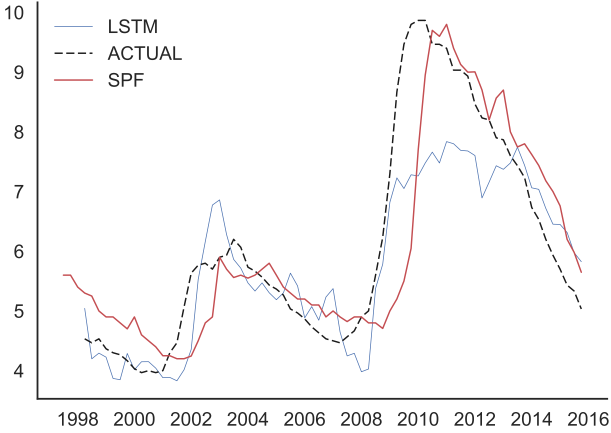
LSTM v SPF performance (6 months)



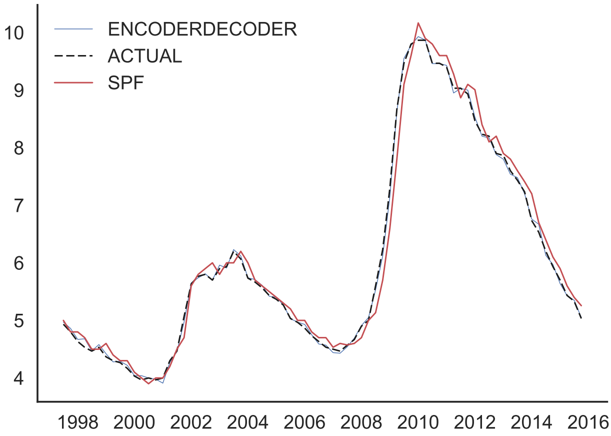
LSTM v SPF performance (9 months)



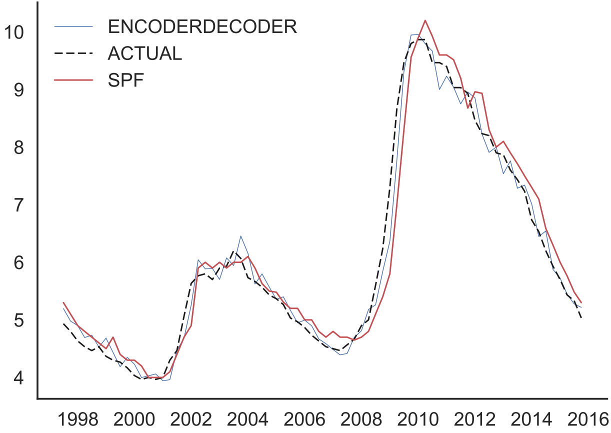
LSTM vs SPF performance (12 Months)



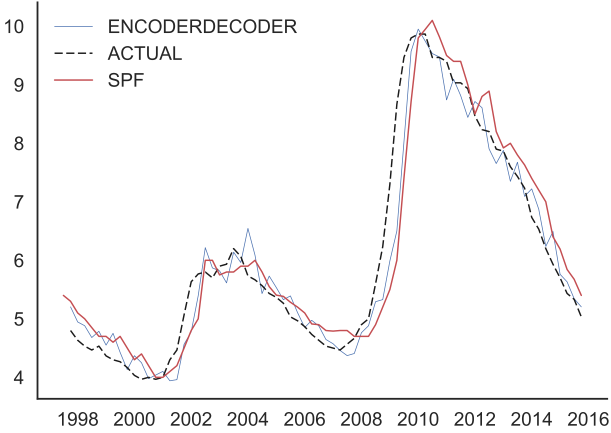
Encoder Decoder vs SPF performance (Current Quarter)



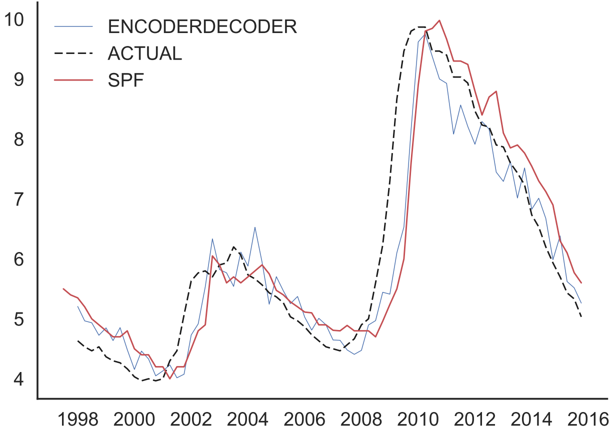
Encoder Decoder vs SPF performance (3 months)



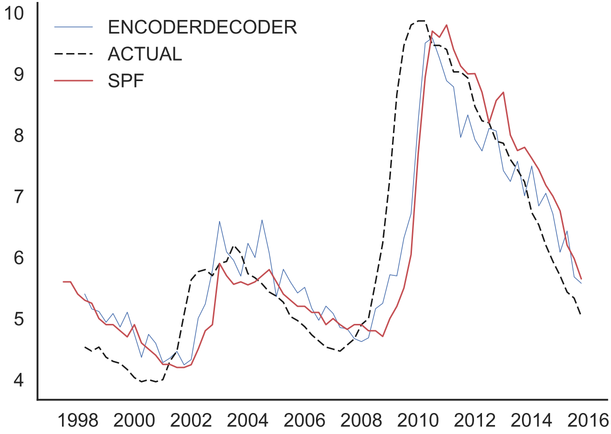
Encoder Decoder v SPF performance (6 months)



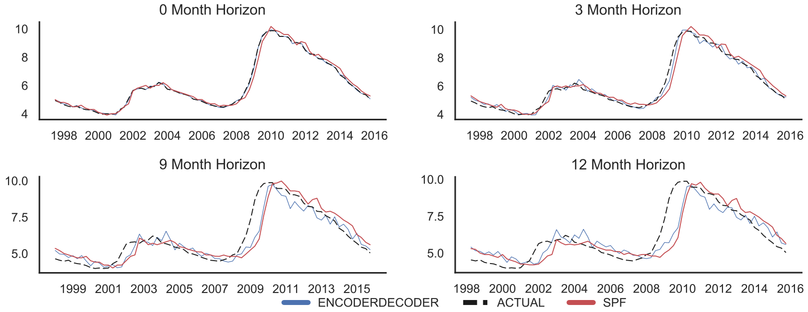
Encoder Decoder v SPF performance (9 months)



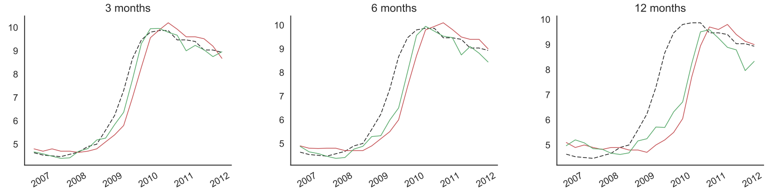
Encoder Decoder vs SPF performance (12 Months)



Encoder Decoder Performance

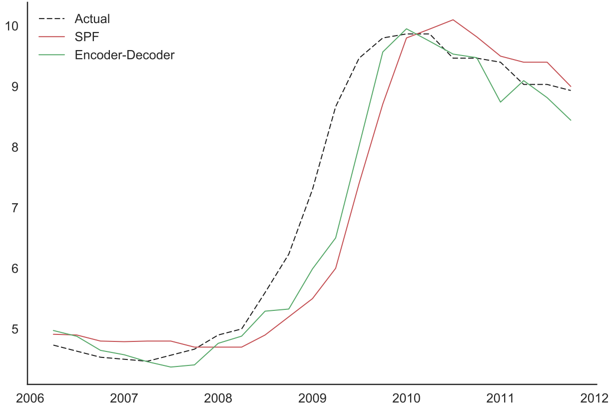


encoder decoder reaction (no smoothing)

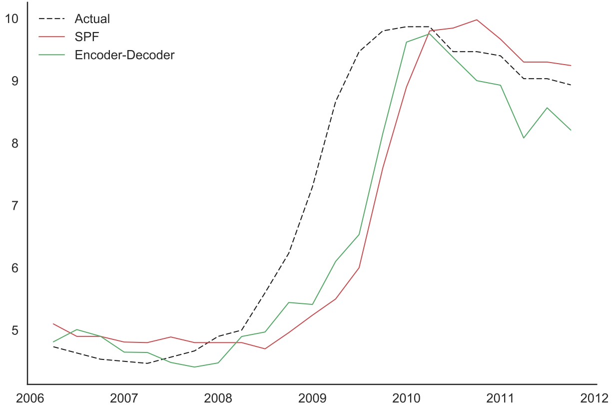


encoder decoder reaction 3 months (no smoothing)

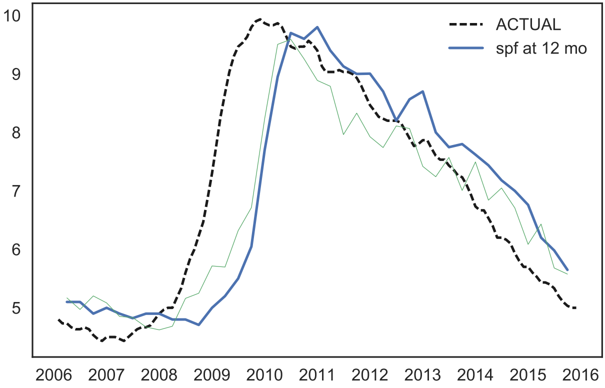
encoder decoder reaction 6 months (no smoothing)



encoder decoder reaction 9 months (no smoothing)



encoder decoder reaction 12 months (no smoothing)



response table

	Encoder Decoder	SPF
Unemployment Nadir (Q1 2007)		
3 Month Horizon Model	Q1 2007	Q3 2007
6 Month Horizon Model	Q2 2007	Q3 2007
9 Month Horizon Model	Q3 2007	Q2 2008
12 Month Horizon Model	Q1 2008	Q3 2008

Unemployment Apex: Q1 2010		
3 Month Horizon Model	Q4 2009	Q1 2010
6 Month Horizon Model	Q4 2009	Q2 2010
9 Month Horizon Model	Q1 2010	Q3 2010
12 Month Horizon Model	Q2 2010	Q4 2010