Predicting Inflation with Neural Networks

Livia Paranhos
University of Warwick

November 2020

Motivation

- Simple univariate models do well
  (RW, AR, UCSV of Stock and Watson, 2007)

Motivation

- Simple univariate models do well
  (RW, AR, UCSV of Stock and Watson, 2007)
- Possible role for nonlinearities, especially when uncertainty is high

Motivation

- Simple univariate models do well
  (RW, AR, UCSV of Stock and Watson, 2007)
- Possible role for nonlinearities, especially when uncertainty is high
- Difficult to determine what variables are important (Giannone et al., 2018)

Motivation

- Simple univariate models do well
  (RW, AR, UCSV of Stock and Watson, 2007)
- Possible role for nonlinearities, especially when uncertainty is high
- Difficult to determine what variables are important (Giannone et al., 2018)
- Machine learning models can address these issues
  - Big-data environment & highly nonlinear

Motivation

- Simple univariate models do well (RW, AR, UCSV of Stock and Watson, 2007)
- Possible role for nonlinearities, especially when uncertainty is high
- Difficult to determine what variables are important (Giannone et al., 2018)
- Machine learning models can address these issues
  - Big-data environment & highly nonlinear
  - Linear shrinkage methods (LASSO, adaLASSO, RR) Inoue and Kilian (2008), Medeiros and Mendes (2016)
  - Nonlinear methods (random forests, SVM, neural networks) Nakamura (2005), Sermpinis et al. (2014), Chakraborty and Joseph (2017), Medeiros et al. (2019)

This paper

1. Applies a **recurrent** neural network to forecast inflation
   - Algorithm designed to model time series data

This paper

1. Applies a **recurrent** neural network to forecast inflation
   - Algorithm designed to model time series data
2. Relevance of macroeconomic data in the prediction (compared to CPI-only information)
   - Real, nominal & financial data (excluding CPI) *versus* CPI data

This paper

1. Applies a **recurrent** neural network to forecast inflation
   - Algorithm designed to model time series data
2. Relevance of macroeconomic data in the prediction (compared to CPI-only information)
   - Real, nominal & financial data (excluding CPI) *versus* CPI data
3. (By-product)
   - Common components
   - Sensitivity analysis of initial values

Main findings

1. **Accuracy.** Neural nets present better performance than usual benchmarks, especially at the one and two-year horizon forecasts

Main findings

1. **Accuracy.** Neural nets present better performance than usual benchmarks, especially at the one and two-year horizon forecasts
2. **Recurrent neural net.** At least as good as the traditional feed-forward neural network at medium-long horizons

Main findings

1. **Accuracy.** Neural nets present better performance than usual benchmarks, especially at the one and two-year horizon forecasts
2. **Recurrent neural net.** At least as good as the traditional feed-forward neural network at medium-long horizons
3. **Other predictors vs CPI.** Macroeconomic information is important during periods of high uncertainty (Stock and Watson (2009), Medeiros et al. (2019))

Main findings

1. **Accuracy.** Neural nets present better performance than usual benchmarks, especially at the one and two-year horizon forecasts
2. **Recurrent neural net.** At least as good as the traditional feed-forward neural network at medium-long horizons
3. **Other predictors vs CPI.** Macroeconomic information is important during periods of high uncertainty (Stock and Watson (2009), Medeiros et al. (2019))
4. **No sparcity.** All groups of predictors seem to be important to predict inflation (Giannone et al., 2018)

Econometric framework

Consider two sets of predictive variables for $t = 1, ..., T$

$\mathbf{x_t} = (x_{1t}, ..., x_{Nt})'$: pool of economic predictors
$\mathbf{y_t} = (y_{1t}, ..., y_{Mt})'$: CPI and its components

- the set $\mathbf{y_t}$ is not contained in $\mathbf{x_t}$

Econometric framework

Consider two sets of predictive variables for $t = 1, ..., T$

$\mathbf{x_t} = (x_{1t}, ..., x_{Nt})'$: pool of economic predictors
$\mathbf{y_t} = (y_{1t}, ..., y_{Mt})'$: CPI and its components

- the set $\mathbf{y_t}$ is not contained in $\mathbf{x_t}$

Let $\mathbf{z_t}$ be the set collecting the predictors at time t

- $\mathbf{z_t} = \mathbf{x_t}, \mathbf{y_t}$ or $(\mathbf{x_t}, \mathbf{y_t})'$

And let $\mathbf{z_t^L}$ be the set collecting the current and lagged values of $\mathbf{z_t}$

- $\mathbf{z_t^L} = (\mathbf{z_t}, \mathbf{z_{t-1}}, ..., \mathbf{z_{t-(L-1)}})'$

Econometric framework (cont'd)

I suppose that inflation, $y_t \in \mathbb{R}$, evolves nonlinearly wrt $\mathbf{z_t^L}$ through a function G, such that

$$y_{t+h} = G(\mathbf{z_t^L}; \Theta_h) + \varepsilon_{t+h}$$

G is a *neural network*

Econometric framework (cont'd)

I suppose that inflation, $y_t \in \mathbb{R}$, evolves nonlinearly wrt $\mathbf{z_t^L}$ through a function G, such that

$$y_{t+h} = G(\mathbf{z_t^L}; \Theta_h) + \varepsilon_{t+h}$$

G is a *neural network*

Fitting the unknown function $G : \mathbf{z_t^L} \rightarrow y_{t+h}$ to the data corresponds to estimating Θ_h given a network architecture, $\mathcal{A}_G$, by minimizing

$$\mathcal{L} = \frac{1}{T} \sum_{t=1}^{T} \Big(y_{t+h} - G(\mathbf{z_t^L}; \Theta_h) \Big)^2$$

Econometric framework (cont'd)

I suppose that inflation, $y_t \in \mathbb{R}$, evolves nonlinearly wrt $\mathbf{z_t^L}$ through a function G, such that

$$y_{t+h} = G(\mathbf{z_t^L}; \Theta_h) + \varepsilon_{t+h}$$

G is a *neural network*

Fitting the unknown function $G : \mathbf{z_t^L} \to y_{t+h}$ to the data corresponds to estimating Θ_h given a network architecture, $\mathcal{A}_G$, by minimizing

$$\mathcal{L} = \frac{1}{T} \sum_{t=1}^{T} \Big(y_{t+h} - G(\mathbf{z_t^L}; \Theta_h) \Big)^2$$

- $\mathcal{A}_G$: neural net model & tunning parameters (hyperparameters)

Econometric framework (cont'd)

I suppose that inflation, $y_t \in \mathbb{R}$, evolves nonlinearly wrt $\mathbf{z_t^L}$ through a function G, such that

$$y_{t+h} = G(\mathbf{z_t^L}; \Theta_h) + \varepsilon_{t+h}$$

G is a *neural network*

Fitting the unknown function $G : \mathbf{z_t^L} \rightarrow y_{t+h}$ to the data corresponds to estimating Θ_h given a network architecture, $\mathcal{A}_G$, by minimizing

$$\mathcal{L} = \frac{1}{T} \sum_{t=1}^{T} \Big(y_{t+h} - G(\mathbf{z_t^L}; \Theta_h) \Big)^2$$

- $\mathcal{A}_G$: neural net model & tunning parameters (hyperparameters)
- Universal approximation theorem (Cybenko, 1989): simple neural net model can approximate any continuous function up to an arbitrary degree of accuracy

The Models

1. Feed forward (FF) model [*multilayer perceptron*]

   (FF-pool) $\mathbf{z_t} = \mathbf{x_t}$

   (FF-cpi) $\mathbf{z_t} = \mathbf{y_t}$

2. Long-short term memory (LSTM) model [*recurrent* neural net]

   (LSTM-pool) $\mathbf{z_t} = \mathbf{x_t}$

   (LSTM-all) $\mathbf{z_t} = (\mathbf{x_t}, \mathbf{y_t})'$

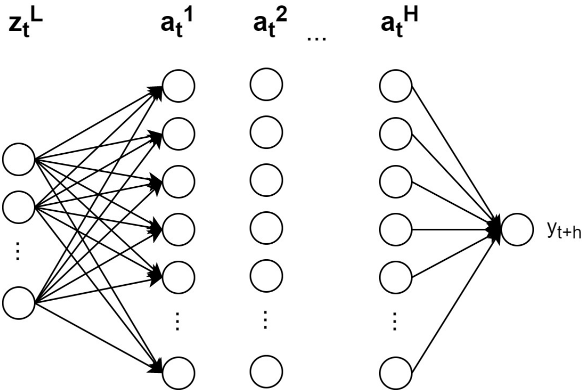
3. FF-LSTM model

   $\mathbf{z_t} = (\mathbf{x_t}, \mathbf{y_t})'$

The Models

1. The Feed forward (FF) model
2. The LSTM model
3. The FF-LSTM model

The feed forward model

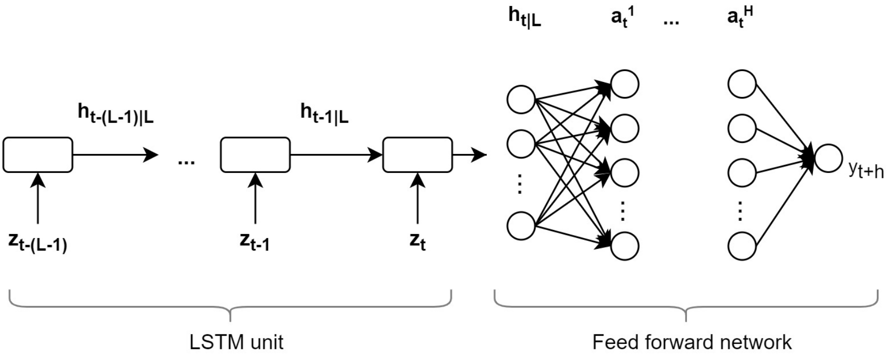


FF equations

The Models

1. The Feed forward (FF) model
2. The LSTM model
3. The FF-LSTM model

The LSTM model



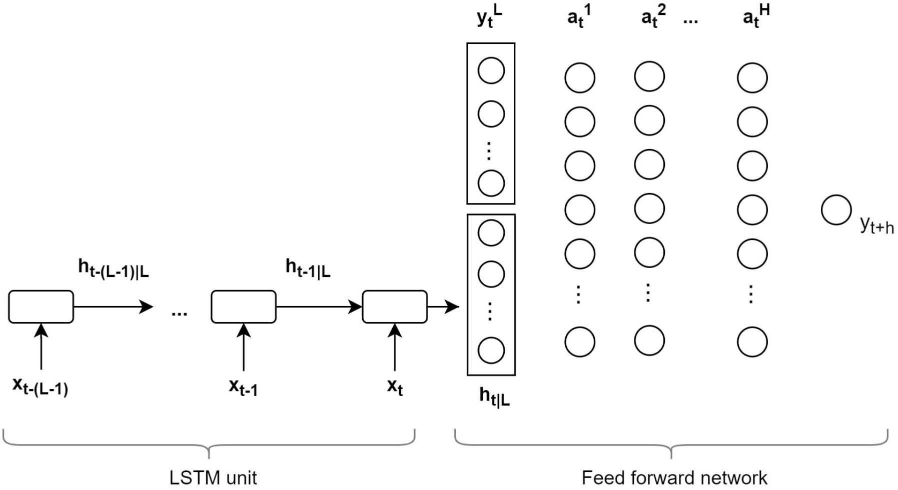
LSTM equations

LSTM graph

The Models

1. The Feed forward (FF) model
2. The LSTM model
3. The FF-LSTM model

The FF-LSTM model



▸ FF-LSTM equations

Data

- FRED-MD data base, downloaded in November, 2019 (128 series, 730 observations)
- Data set $\mathbf{y_t}$ ($M = 10$)
  - CPI: all items, apparel, transportation, medical care, commodities, durables, services, all items less food, all items less shelter, all items less medical care
- Data set $\mathbf{x_t}$ ($N = 118$)
  - Output&Income, Labour market, Housing, Consumption, Money&Credit, Interest&ER, Prices, Stock market

Optimal tuning parameters

	Tuning parameters				
	lags	nodes	layers	factors	# par.
FF-cpi	24	128	4	n/a	80K
FF-pool	48	128	3	n/a	760K
LSTM-pool	48	128	4	2	50K
LSTM-all	48	128	4	2	50K
FF-LSTM	24, 48	128	4	2	80K

▸ Variable selection

Out-of-sample performance

Table: Loss ratios wrt the AR(1) model over 2006M08-2019M10

		Horizon (months)				
Model	Data	1	3	6	12	24
RMSE						
UCSV	CPI	1.13	1.05	1.03	1.02	1.00
FADL	Pool	1.05	1.09	1.08	1.01	1.00
FF-cpi	CPI	1.07	1.01	1.01	0.98	0.91
FF-pool	Pool	1.09	**0.92**	**0.90**	0.94	0.99
LSTM-pool	Pool	1.00	0.93	1.03	0.93	0.92
LSTM-all	All	**0.98**	0.94	1.04	**0.92**	0.91
FF-LSTM	All	1.06	0.99	0.99	0.97	**0.89**

Initial values

Predictions

RMSE over time

Fluctuation test (Giacomini and Rossi (2010))

- Test for equal forecast accuracy robust to instability

RMSE over time

Fluctuation test (Giacomini and Rossi (2010))

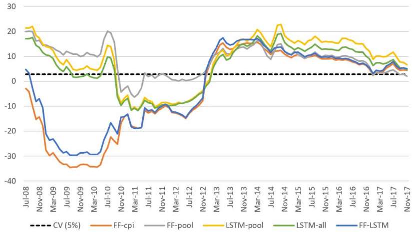
- Test for equal forecast accuracy robust to instability

In this application

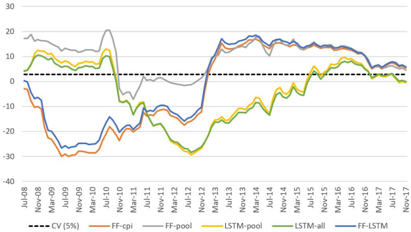
- Statistic above the critical value implies the candidate model is superior to the benchmark for a specific time window
- Rolling window of $m = 48$ observations across the out-of-sample

RMSE over time (cont'd)

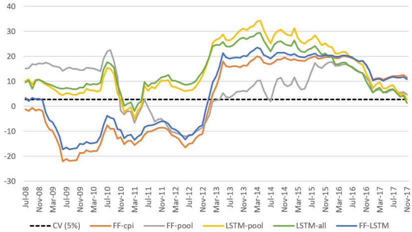
(a) Horizon 3



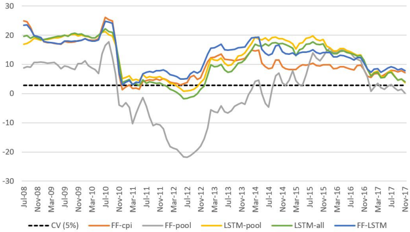
(b) Horizon 6



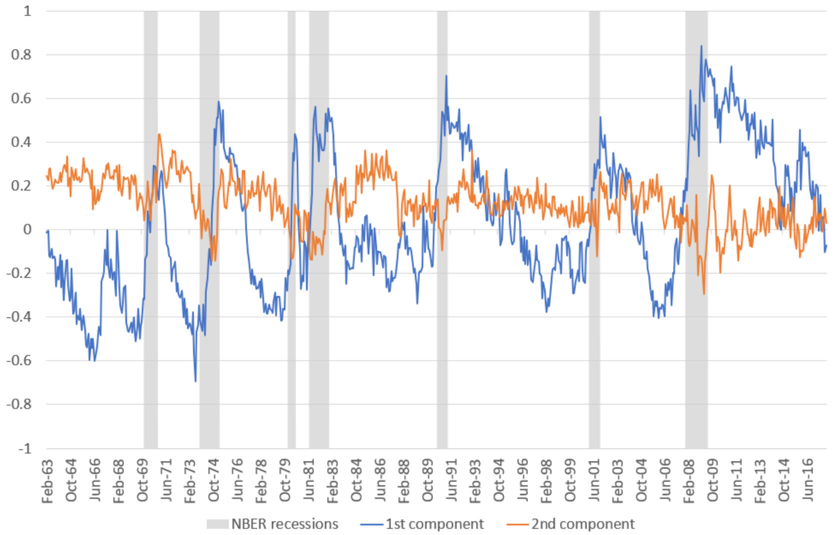
(c) Horizon 12



(d) Horizon 24



Common components (FF-LSTM, horizon 24)



Conclusions

- Proposes to forecast inflation using **recurrent** neural networks
  - Highly nonlinear
  - Suitable for time series analysis
- Relevance of macroeconomic predictors
- (Recurrent) Neural networks are promising to forecast inflation at **medium-long** horizons

References

Chakraborty, C., and A. Joseph. (2017). “Machine learning at central banks.” Bank of England Working Papers 674.

Cybenko, G. (1989). “Approximation by superposition of a sigmoidal function.” Mathematics of Control, Signals and Systems 2, 303–314.

Giacomini, R., and B. Rossi. (2010). “Forecast Comparisons in Unstable Environments.” Journal of Applied Econometrics 25, 595–620.

Giannone, D., M. Lenza, and G. Primiceri. (2018). “Economic predictions with big data: The illusion of sparcity.” Working paper, Northwestern University .

Inoue, A., and L. Kilian. (2008). “How Useful Is Bagging in Forecasting Economic Time Series? A Case Study of U.S. Consumer Price Inflation.” Journal of the American Statistical Association 103, 511–522.

Medeiros, M., and E. Mendes. (2016). “L1-regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic errors.” Journal of Econometrics 191, 255–271.

Medeiros, M. C., G. Vasconcelos, A. Veiga, and E. Zilberman. (2019). “Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods.” Journal of Business & Economic Statistics 1–45.

Nakamura, E.. (2005). “Inflation forecasting using a neural network.” Economics Letters 86, 373–378.

Sermpinis, G., C. Stasinakis, K. Theofilatos, and A. Karathanasopoulos. (2014). “Inflation and Unemployment Forecasting with Genetic Support Vector Regression.” Journal of Forecasting 33, 471–487.

Stock, J. H., and M. W. Watson. (2007). “Why Has U.S. Inflation Become Harder to Forecast?” Journal of Money, Credit and Banking. 39, 7, 1849–1849.

Stock, J. H., and M. W. Watson. (2009). “Phillips Curve Inflation Forecasts.” In “Fuhrer J, Kodrzycki Y, Little J, Olivei G Understanding Inflation and the Implications for Monetary Policy.”, 99–202, Cambridge: MIT Press.

The feed forward model

$$G(\mathbf{z_t^L}; \Theta_h) = g_{FF}\left(\mathbf{z_t^L}; \theta_h\right)$$

and

$$g_{FF}(\mathbf{z_t^L}; \theta_h) = \mathbf{W_{H+1} a_t^H} + b_{H+1}$$
$$\mathbf{a_t^i} = ReLu(\mathbf{W_i a_t^{i-1}} + \mathbf{b_i}), \quad i = 1, 2, ..., H$$
$$\mathbf{a_t^0} = \mathbf{z_t^L}$$

where

$\mathbf{a_t^i}$: $n \times 1$ hidden layer vectors

$\theta_h = (\{\mathbf{W_i}\}_{i=1}^H, \{\mathbf{b_i}\}_{i=1}^H)'$: model parameters

$ReLu : \mathbb{R} \to \mathbb{R}$: rectified linear unit function $f(z) = max\{0, z\}$

$$\mathcal{A}_G = \{G(\mathbf{z_t^L}; \Theta_h), L, n, H\}$$

Go Back

The LSTM model

$$G(\mathbf{z_t^L};\Theta_h) = g_{FF}\Big(\mathbf{h_{t|L}}(\mathbf{z_t};\phi_h);\theta_h\Big)$$

where
$t|L \equiv t|t, t-1, ..., t-(L-1)$

Go Back

The LSTM model

$$G(\mathbf{z_t^L};\Theta_h) = g_{FF}\Big(\mathbf{h_{t|L}}(\mathbf{z_t};\phi_h);\theta_h\Big)$$

$$\mathbf{f_{t|L}} = sigmoid(\mathbf{W_f'z_t} + \mathbf{U_f h_{t-1|L}} + \mathbf{b_f})$$

$$\mathbf{i_{t|L}} = sigmoid(\mathbf{W_i'z_t} + \mathbf{U_i h_{t-1|L}} + \mathbf{b_i})$$

$$\mathbf{o_{t|L}} = sigmoid(\mathbf{W_o'z_t} + \mathbf{U_o h_{t-1|L}} + \mathbf{b_o})$$

where
$t|L \equiv t|t, t-1, ..., t-(L-1)$

Go Back

The LSTM model

$$G(\mathbf{z_t^L};\Theta_h) = g_{FF}\Big(\mathbf{h_{t|L}}(\mathbf{z_t};\phi_h);\theta_h\Big)$$

$$\mathbf{f_{t|L}} = sigmoid(\mathbf{W'_f z_t} + \mathbf{U_f h_{t-1|L}} + \mathbf{b_f})$$

$$\mathbf{i_{t|L}} = sigmoid(\mathbf{W'_i z_t} + \mathbf{U_i h_{t-1|L}} + \mathbf{b_i})$$

$$\mathbf{o_{t|L}} = sigmoid(\mathbf{W'_o z_t} + \mathbf{U_o h_{t-1|L}} + \mathbf{b_o})$$

$$\mathbf{c_{t|L}} = \mathbf{f_{t|L}} \odot \mathbf{c_{t-1|L}} + \mathbf{i_{t|L}} \odot tanh(\mathbf{W'_c z_t} + \mathbf{U_c h_{t-1|L}} + \mathbf{b_c})$$

where
$t|L \equiv t|t, t-1, ..., t-(L-1)$

Go Back

The LSTM model

$$G(\mathbf{z_t^L};\Theta_h) = g_{FF}\Big(\mathbf{h_{t|L}}(\mathbf{z_t};\phi_h);\theta_h\Big)$$

$$\mathbf{f_{t|L}} = sigmoid(\mathbf{W_f'z_t} + \mathbf{U_f h_{t-1|L}} + \mathbf{b_f})$$

$$\mathbf{i_{t|L}} = sigmoid(\mathbf{W_i'z_t} + \mathbf{U_i h_{t-1|L}} + \mathbf{b_i})$$

$$\mathbf{o_{t|L}} = sigmoid(\mathbf{W_o'z_t} + \mathbf{U_o h_{t-1|L}} + \mathbf{b_o})$$

$$\mathbf{c_{t|L}} = \mathbf{f_{t|L}} \odot \mathbf{c_{t-1|L}} + \mathbf{i_{t|L}} \odot tanh(\mathbf{W_c'z_t} + \mathbf{U_c h_{t-1|L}} + \mathbf{b_c})$$

$$\mathbf{h_{t|L}} = \mathbf{o_{t|L}} \odot tanh(\mathbf{c_{t|L}})$$

$$\mathbf{h_0} = \mathbf{0}, \quad \mathbf{c_0} = \mathbf{0}$$

where

$t|L \equiv t|t, t-1, ..., t-(L-1)$

$\mathbf{f}, \mathbf{i}, \mathbf{o}, \mathbf{c}, \mathbf{h} \in \mathbb{R}^{s\times 1}$

Go Back

The LSTM model

$$G(\mathbf{z_t^L};\Theta_h) = g_{FF}\Big(\mathbf{h_{t|L}}(\mathbf{z_t};\phi_h);\theta_h\Big)$$

$$\mathbf{f_{t|L}} = sigmoid(\mathbf{W'_f z_t} + \mathbf{U_f h_{t-1|L}} + \mathbf{b_f})$$

$$\mathbf{i_{t|L}} = sigmoid(\mathbf{W'_i z_t} + \mathbf{U_i h_{t-1|L}} + \mathbf{b_i})$$

$$\mathbf{o_{t|L}} = sigmoid(\mathbf{W'_o z_t} + \mathbf{U_o h_{t-1|L}} + \mathbf{b_o})$$

$$\mathbf{c_{t|L}} = \mathbf{f_{t|L}} \odot \mathbf{c_{t-1|L}} + \mathbf{i_{t|L}} \odot tanh(\mathbf{W'_c z_t} + \mathbf{U_c h_{t-1|L}} + \mathbf{b_c})$$

$$\mathbf{h_{t|L}} = \mathbf{o_{t|L}} \odot tanh(\mathbf{c_{t|L}})$$

$$\mathbf{h_0} = \mathbf{0}, \quad \mathbf{c_0} = \mathbf{0}$$

where

$t|L \equiv t|t, t-1, ..., t-(L-1)$

$\mathbf{f}, \mathbf{i}, \mathbf{o}, \mathbf{c}, \mathbf{h} \in \mathbb{R}^{s\times 1}$

$\Theta_h = (\phi_h, \theta_h)'$, with

$\phi_h = (\mathbf{W_{(j)}}, \mathbf{U_{(j)}}, \mathbf{b_{(j)}})'$, $j = f, i, o, c$

$\theta_h = (\{\mathbf{W_i}\}_{i=1}^{H}, \{\mathbf{b_i}\}_{i=1}^{H})'$, given n and H

Go Back

The LSTM model

$$G(\mathbf{z_t^L};\Theta_h) = g_{FF}\Big(\mathbf{h_{t|L}}(\mathbf{z_t};\phi_h);\theta_h\Big)$$

$$\mathbf{f_{t|L}} = sigmoid(\mathbf{W_f'z_t} + \mathbf{U_f h_{t-1|L}} + \mathbf{b_f})$$

$$\mathbf{i_{t|L}} = sigmoid(\mathbf{W_i'z_t} + \mathbf{U_i h_{t-1|L}} + \mathbf{b_i})$$

$$\mathbf{o_{t|L}} = sigmoid(\mathbf{W_o'z_t} + \mathbf{U_o h_{t-1|L}} + \mathbf{b_o})$$

$$\mathbf{c_{t|L}} = \mathbf{f_{t|L}} \odot \mathbf{c_{t-1|L}} + \mathbf{i_{t|L}} \odot tanh(\mathbf{W_c'z_t} + \mathbf{U_c h_{t-1|L}} + \mathbf{b_c})$$

$$\mathbf{h_{t|L}} = \mathbf{o_{t|L}} \odot tanh(\mathbf{c_{t|L}})$$

$$\mathbf{h_0} = \mathbf{0}, \quad \mathbf{c_0} = \mathbf{0}$$

where

$t|L \equiv t|t, t-1, ..., t-(L-1)$

$\mathbf{f}, \mathbf{i}, \mathbf{o}, \mathbf{c}, \mathbf{h} \in \mathbb{R}^{s\times 1}$

$\Theta_h = (\phi_h, \theta_h)'$, with

$\phi_h = (\mathbf{W_{(j)}}, \mathbf{U_{(j)}}, \mathbf{b_{(j)}})'$, $j = f, i, o, c$

$\theta_h = (\{\mathbf{W_i}\}_{i=1}^{H}, \{\mathbf{b_i}\}_{i=1}^{H})'$, given n and H

$\mathcal{A}_G = \{G(\mathbf{z_t^L};\Theta_h), s, L, n, H\}$

Go Back

The FF-LSTM model

$$G(\mathbf{z_t^L}; \Theta_h) = g_{FF}\Big((\mathbf{h_{t|L}}(\mathbf{z_t}; \phi_h), \mathbf{z_t^L})'; \theta_h\Big)$$

where

$t|L \equiv t|t, t-1, ..., t-(L-1)$

$\Theta_h = (\phi_h, \theta_h)'$,

$\mathcal{A}_G = \{G(\mathbf{z_t^L}; \Theta_h), s, L, n, H\}$

g_{FF} receives the input vector $(\mathbf{h_{t|L}}(\cdot), \mathbf{z_t^L})' \in \mathbb{R}^{(s+ML)\times 1}$

Go Back

Sensitivity to initial values (I)

- The predictions at each point in time are averages over 1400 predictions with distinct initial values
- Non-convexity increases the sensitivity of the algorithm to initial values

Counterfactual exercise

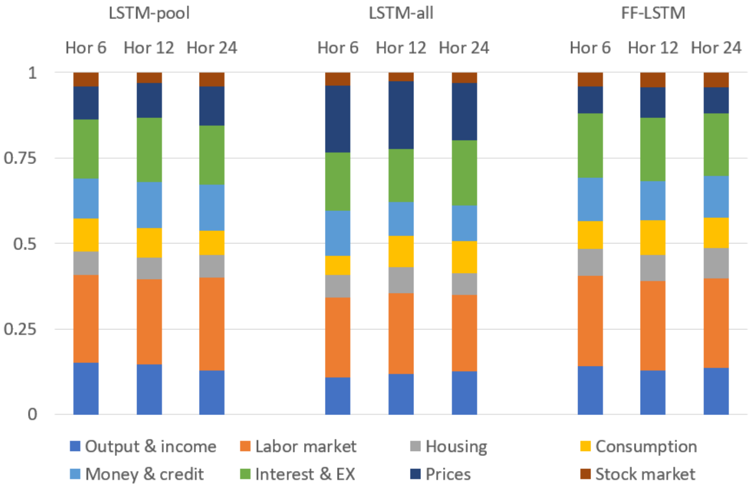
- Compute the performance of all 1400 prediction series over the out-of-sample set
- Compare the series with minimum error out-of-sample with the average prediction series

Sensitivity to initial values (II)

	Horizons (months)				
Model	1	3	6	12	24
RMSE					
FF-cpi	0.96	0.97	0.95	0.98	0.97
FF-pool	0.96	0.96	0.95	0.93	0.89
LSTM-pool	0.97	0.93	0.89	0.96	0.93
LSTM-all	0.91	0.99	0.86	0.99	0.96
FF-LSTM	0.96	0.97	0.97	0.96	0.94

Go Back

Variable selection

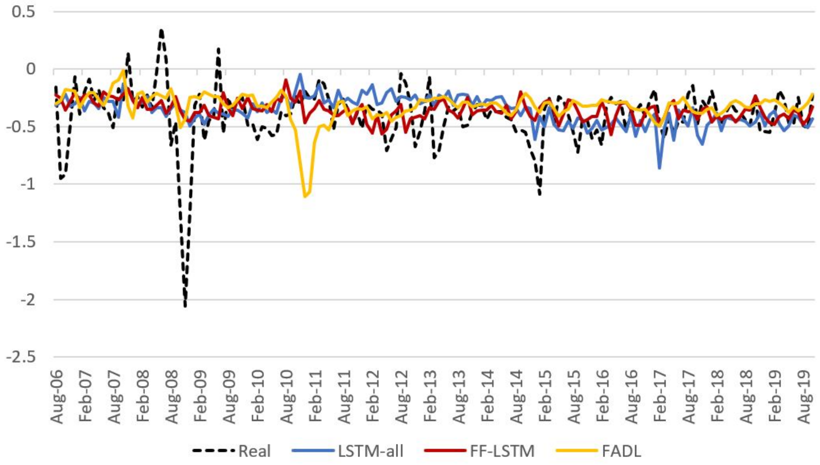


Go Back

A word on identification

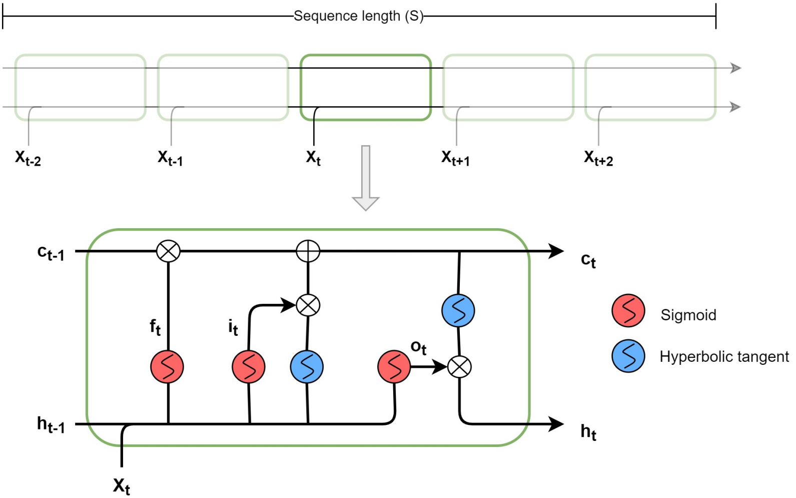
- No guarantee of global optimum
  - Neural networks focus on good prediction accuracy on unseen data, ultimately an empirical question
  - Multiple equilibria and/or flat regions: intrinsic symmetry, mutual dependence of weights
- Non-convexity increases the sensitivity of the learning algorithm to initial values
  - Zero-mean uniform distribution
  - Empirical solution: average out the predictions of large number of repeats
- No identification explains the use of cross-validation as model selection (no probabilistic assumptions)

Predictions at the two-year ahead horizon



Go Back

The LSTM



◂ Go Back