Macroeconomic Indicator Forecasting with Deep Neural Networks

Tom Cook & Aaron Smalter Hall

Jul 12, 2018
1 Introduction

2 Method: Neural Networks

3 Data

4 Results
<table>
<thead>
<tr>
<th>Introduction</th>
<th>Method: Neural Networks</th>
<th>Data</th>
<th>Results</th>
</tr>
</thead>
</table>

Macroeconomic Indicator Forecasting with Deep Neural Networks
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, revitalized 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.
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

<table>
<thead>
<tr>
<th>horiz</th>
<th>Fully Connected</th>
<th>LSTM</th>
<th>Encoder Decoder</th>
<th>DARM</th>
<th>SPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 Months</td>
<td>Mean MAE</td>
<td>0.076</td>
<td>0.104</td>
<td>0.044</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td>St. Dev.</td>
<td>0.027</td>
<td>0.058</td>
<td>0.002</td>
<td>0.063</td>
</tr>
<tr>
<td>3 Months</td>
<td>Mean MAE</td>
<td>0.253</td>
<td>0.273</td>
<td>0.184</td>
<td>0.328</td>
</tr>
<tr>
<td></td>
<td>St. Dev.</td>
<td>0.029</td>
<td>0.028</td>
<td>0.001</td>
<td>0.093</td>
</tr>
<tr>
<td>6 Months</td>
<td>Mean MAE</td>
<td>0.441</td>
<td>0.473</td>
<td>0.305</td>
<td>0.493</td>
</tr>
<tr>
<td></td>
<td>St. Dev.</td>
<td>0.041</td>
<td>0.066</td>
<td>0.005</td>
<td>0.157</td>
</tr>
<tr>
<td>9 Months</td>
<td>Mean MAE</td>
<td>0.638</td>
<td>0.748</td>
<td>0.461</td>
<td>0.658</td>
</tr>
<tr>
<td></td>
<td>St. Dev.</td>
<td>0.048</td>
<td>0.110</td>
<td>0.007</td>
<td>0.228</td>
</tr>
<tr>
<td>12 Months</td>
<td>Mean MAE</td>
<td>0.870</td>
<td>1.017</td>
<td>0.620</td>
<td>0.907</td>
</tr>
<tr>
<td></td>
<td>St. Dev.</td>
<td>0.049</td>
<td>0.207</td>
<td>0.006</td>
<td>0.291</td>
</tr>
</tbody>
</table>
Encoder Decoder Response Comparison\(^1\)

\(^1\)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
<table>
<thead>
<tr>
<th>Architectures</th>
<th>Error Distributions</th>
<th>Ensemble Models</th>
<th>V SPF</th>
<th>FC</th>
<th>LSTM</th>
<th>ENCODER-DECODER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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
<table>
<thead>
<tr>
<th>Architectures</th>
<th>Error Distributions</th>
<th>Ensemble Models</th>
<th>V SPF</th>
<th>FC</th>
<th>LSTM</th>
<th>ENCODEC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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
  - 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

Tom Cook & Aaron Smalter Hall
Macroeconomic Indicator Forecasting with Deep Neural Networks

Jul 12, 2018
Slide 21
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

<table>
<thead>
<tr>
<th>Forecast Horizon</th>
<th>Fully Connected</th>
<th>LSTM</th>
<th>Encoder Decoder</th>
<th>SPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 month</td>
<td>0.043</td>
<td>0.048</td>
<td>0.041</td>
<td>0.10</td>
</tr>
<tr>
<td>3 month</td>
<td>0.215</td>
<td>0.242</td>
<td>0.184</td>
<td>0.23</td>
</tr>
<tr>
<td>6 month</td>
<td>0.414</td>
<td>0.401</td>
<td>0.301</td>
<td>0.36</td>
</tr>
<tr>
<td>9 month</td>
<td>0.548</td>
<td>0.664</td>
<td>0.459</td>
<td>0.50</td>
</tr>
<tr>
<td>12 month</td>
<td>0.779</td>
<td>0.900</td>
<td>0.618</td>
<td>0.63</td>
</tr>
</tbody>
</table>
SPF errors for comparison

- SPF is a panel of experts. Each panelist is surveyed multiple times.
- \( e = \frac{1}{n} \sum_i \frac{1}{T_i} \sum_t |y_t - \hat{y}_{ti}| \)
- 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:
- \[ \hat{\text{UNRATE}}_{t+n} = \sum_{i=1}^{k} \beta_i \text{UNRATE}_{t-i} \]
- 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.
<table>
<thead>
<tr>
<th>Architectures</th>
<th>Error Distributions</th>
<th>Ensemble Models</th>
<th>V SPF</th>
<th>FC</th>
<th>LSTM</th>
<th>ENCDEC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>V SPF</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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)

Graph showing LSTM vs SPF performance from 1998 to 2016.
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

0 Month Horizon

3 Month Horizon

9 Month Horizon

12 Month Horizon

ENCODERDECODER ACTUAL SPF
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)
### Unemployment Nadir (Q1 2007)

<table>
<thead>
<tr>
<th>Model</th>
<th>Encoder</th>
<th>Decoder</th>
<th>SPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Month Horizon Model</td>
<td>Q1 2007</td>
<td>Q3 2007</td>
<td>Q3 2007</td>
</tr>
<tr>
<td>6 Month Horizon Model</td>
<td>Q2 2007</td>
<td>Q3 2007</td>
<td>Q3 2007</td>
</tr>
<tr>
<td>9 Month Horizon Model</td>
<td>Q3 2007</td>
<td>Q2 2008</td>
<td>Q2 2008</td>
</tr>
<tr>
<td>12 Month Horizon Model</td>
<td>Q1 2008</td>
<td>Q3 2008</td>
<td>Q3 2008</td>
</tr>
</tbody>
</table>

### Unemployment Apex: Q1 2010

<table>
<thead>
<tr>
<th>Model</th>
<th>Encoder</th>
<th>Decoder</th>
<th>SPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Month Horizon Model</td>
<td>Q4 2009</td>
<td>Q1 2010</td>
<td>Q1 2010</td>
</tr>
<tr>
<td>6 Month Horizon Model</td>
<td>Q4 2009</td>
<td>Q2 2010</td>
<td>Q2 2010</td>
</tr>
<tr>
<td>9 Month Horizon Model</td>
<td>Q1 2010</td>
<td>Q3 2010</td>
<td>Q3 2010</td>
</tr>
<tr>
<td>12 Month Horizon Model</td>
<td>Q2 2010</td>
<td>Q4 2010</td>
<td>Q4 2010</td>
</tr>
</tbody>
</table>