01_univariate_time_series_regression

September 29, 2021

1 Recurrent Neural Networks

1.1 Univariate Time Series Regression

This notebook demonstrates how to forecast the S&P 500 index using a Recurrent Neural Network.

1.2 Imports & Settings

[1]:
```
import warnings
warnings.filterwarnings('ignore')
```

[2]:
```
%matplotlib inline

from pathlib import Path

import numpy as np
import pandas as pd
import pandas_datareader.data as web
from scipy.stats import spearmanr

from sklearn.metrics import mean_squared_error
from sklearn.preprocessing import MinMaxScaler

import tensorflow as tf
from tensorflow.keras.callbacks import ModelCheckpoint, EarlyStopping
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, LSTM
from tensorflow import keras

import matplotlib.pyplot as plt
import seaborn as sns
```

[3]:
```
gpu_devices = tf.config.experimental.list_physical_devices('GPU')
if gpu_devices:
    print('Using GPU')
    tf.config.experimental.set_memory_growth(gpu_devices[0], True)
else:
    print('Using CPU')
```

```
Using CPU
```

[4]:
```
sns.set_style('whitegrid')
np.random.seed(42)
```

[5]:
```
results_path = Path('results', 'univariate_time_series')
if not results_path.exists():
    results_path.mkdir(parents=True)
```

1.3 Get Data

We obtain data for 2010-2018 from the Federal Reserve Bank's Data Service FRED using the pandas_datareader library in introduced in Chapter 2 on Market and Fundamental Data.

[11]:
```
sp500 = web.DataReader('SP500', 'fred', start='2010', end='2020').dropna()
ax = sp500.plot(title='S&P 500',
           legend=False,
           figsize=(14, 4),
           rot=0)
ax.set_xlabel('')
sns.despine()
```

1.4 Preprocessing

[12]:
```
scaler = MinMaxScaler()
```

[13]:
```
sp500_scaled = pd.Series(scaler.fit_transform(sp500).squeeze(),
                         index=sp500.index)
sp500_scaled.describe()
```

[13]:
```
count    2229.000000
mean        0.451605
std         0.254561
min         0.000000
25%         0.238076
```

```
50%          0.447456
75%          0.659023
max          1.000000
dtype: float64
```

1.5 Generating recurrent sequences from our time series

Our time series is a sequence of numbers indexed by time:

$$x_0, x_1, x_2, ..., x_T$$

where $\{x_t\}$ is the numerical value in period t and T is the total length of the series.

To apply a RNN for regression of classification, we use a sliding window to construct a rolling set of input/output pairs for our model to learn from as animated below.

We will generate sequences of 63 trading days, approximately three months, and use a single LSTM layer with 20 hidden units to predict the index value one timestep ahead. The input to every LSTM layer must have three dimensions, namely: - **Samples**: One sequence is one sample. A batch contains one or more samples. - **Time Steps**: One time step is one point of observation in the sample. - **Features**: One feature is one observation at a time step.

Our S&P 500 sample has 2,264 observations or time steps. We will create overlapping sequences using a window of 63 observations each. For a simpler window of size T = 5, we obtain input-output pairs as shown in the following table:

Input	Output
$\langle x_1, x_2, x_3, x_4, x_5 \rangle$	x_6
$\langle x_2, x_3, x_4, x_5, x_6 \rangle$	x_7
$\vdots$	$\vdots$
$\langle x_{T-5}, x_{T-4}, x_{T-3}, x_{T-2}, x_{T-1} \rangle$	x_T

Generally speaking, for window size S, the relationship takes the form

$$x_t = f(x_{t-1}, x_{t-2}, ..., x_{t-S}) \quad \forall t = S, S+1, ..., T$$

Each of the $T-S$ lagged input sequence or vector is of length S with a corresponding scalar output.

We can use the function create_univariate_rnn_data() to stack sequences selected using a rolling windows:

[14]:
```python
def create_univariate_rnn_data(data, window_size):
    n = len(data)
    y = data[window_size:]
    data = data.values.reshape(-1, 1) # make 2D
    X = np.hstack(tuple([data[i: n-j, :] for i, j in␣
 ↪enumerate(range(window_size, 0, -1))]))
    return pd.DataFrame(X, index=y.index), y
```

We apply this function to the rescaled stock index for a window_size=63 to obtain a two-dimensional dataset of shape number of samples x number of timesteps:

[15]:
```
window_size = 63
```

[16]:
```
X, y = create_univariate_rnn_data(sp500_scaled, window_size=window_size)
```

[17]:
```
X.head()
```

[17]:
```
                   0         1         2         3         4         5   \
DATE
2011-05-24  0.097240  0.096633  0.103069  0.106498  0.096740  0.097726
2011-05-25  0.096633  0.103069  0.106498  0.096740  0.097726  0.108250
2011-05-26  0.103069  0.106498  0.096740  0.097726  0.108250  0.103663
2011-05-27  0.106498  0.096740  0.097726  0.108250  0.103663  0.098515
2011-05-31  0.096740  0.097726  0.108250  0.103663  0.098515  0.103976

                   6         7         8         9   …        53        54  \
DATE                                                 …
2011-05-24  0.108250  0.103663  0.098515  0.103976  …  0.120484  0.113439
2011-05-25  0.103663  0.098515  0.103976  0.103135  …  0.113439  0.116508
2011-05-26  0.098515  0.103976  0.103135  0.091499  …  0.116508  0.111426
2011-05-27  0.103976  0.103135  0.091499  0.095782  …  0.111426  0.107549
2011-05-31  0.103135  0.091499  0.095782  0.092097  …  0.107549  0.107320

                  55        56        57        58        59        60  \
DATE
2011-05-24  0.116508  0.111426  0.107549  0.107320  0.112785  0.114149
2011-05-25  0.111426  0.107549  0.107320  0.112785  0.114149  0.109324
2011-05-26  0.107549  0.107320  0.112785  0.114149  0.109324  0.101897
2011-05-27  0.107320  0.112785  0.114149  0.109324  0.101897  0.101388
2011-05-31  0.112785  0.114149  0.109324  0.101897  0.101388  0.103345

                  61        62
DATE
2011-05-24  0.109324  0.101897
2011-05-25  0.101897  0.101388
2011-05-26  0.101388  0.103345
2011-05-27  0.103345  0.105783
2011-05-31  0.105783  0.108310

[5 rows x 63 columns]
```

[18]:
```
y.head()
```

[18]:
```
DATE
2011-05-24    0.101388
```

```
2011-05-25    0.103345
2011-05-26    0.105783
2011-05-27    0.108310
2011-05-31    0.114897
dtype: float64
```

[19]:
```python
X.shape
```

[19]:
```
(2166, 63)
```

1.6 Train-test split

To respect the time series nature of the data, we set aside the data at the end of the sample as hold-out or test set. More specifically, we'll use the data for 2018.

[20]:
```python
ax = sp500_scaled.plot(lw=2, figsize=(14, 4), rot=0)
ax.set_xlabel('')
sns.despine()
```

[21]:
```python
X_train = X[:'2018'].values.reshape(-1, window_size, 1)
y_train = y[:'2018']

# keep the last year for testing
X_test = X['2019'].values.reshape(-1, window_size, 1)
y_test = y['2019']
```

[22]:
```python
n_obs, window_size, n_features = X_train.shape
```

[23]:
```python
y_train.shape
```

[23]:
```
(1914,)
```

1.7 Keras LSTM Layer

Keras has several built-in RNN layers with various configuration options described in detail in the documentation.

```
LSTM(units,
     activation='tanh',
     recurrent_activation='hard_sigmoid',
     use_bias=True,
     kernel_initializer='glorot_uniform',
     recurrent_initializer='orthogonal',
     bias_initializer='zeros',
     unit_forget_bias=True,
     kernel_regularizer=None,
     recurrent_regularizer=None,
     bias_regularizer=None,
     activity_regularizer=None,
     kernel_constraint=None,
     recurrent_constraint=None,
     bias_constraint=None,
     dropout=0.0,
     recurrent_dropout=0.0,
     implementation=1,
     return_sequences=False,
     return_state=False,
     go_backwards=False,
     stateful=False,
     unroll=False)
```

1.8 Define the Model Architecture

Having created input/output pairs out of our time series and cut this into training/testing sets, we can now begin setting up our RNN. We use Keras to quickly build a two hidden layer RNN of the following specifications

- layer 1 uses an LSTM module with 20 hidden units (note here the input_shape = (window_size,1))
- layer 2 uses a fully connected module with one unit
- the 'mean_squared_error' loss should be used (remember: we are performing regression here)

This can be constructed using just a few lines - see e.g., the general Keras documentation and the LSTM documentation in particular for examples of how to quickly use Keras to build neural network models. Make sure you are initializing your optimizer given the keras-recommended approach for RNNs

[24]:
```
rnn = Sequential([
    LSTM(units=10,
         input_shape=(window_size, n_features), name='LSTM'),
    Dense(1, name='Output')
])
```

The summary shows that the model has 1,781 parameters:

[25]:
```
rnn.summary()
```

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
LSTM (LSTM)                  (None, 10)                480
_________________________________________________________________
Output (Dense)               (None, 1)                 11
=================================================================
Total params: 491
Trainable params: 491
Non-trainable params: 0
_________________________________________________________________
```

1.9 Train the Model

We train the model using the RMSProp optimizer recommended for RNN with default settings and compile the model with mean squared error for this regression problem:

[26]:
```
optimizer = keras.optimizers.RMSprop(lr=0.001,
                                     rho=0.9,
                                     epsilon=1e-08,
                                     decay=0.0)
```

[27]:
```
rnn.compile(loss='mean_squared_error',
            optimizer=optimizer)
```

We define an EarlyStopping callback and train the model for up to 100 episodes.

[28]:
```
rnn_path = (results_path / 'rnn.h5').as_posix()
checkpointer = ModelCheckpoint(filepath=rnn_path,
                               verbose=1,
                               monitor='val_loss',
                               save_best_only=True)
```

[29]:
```
early_stopping = EarlyStopping(monitor='val_loss',
                              patience=20,
                              restore_best_weights=True)
```

[30]:
```
lstm_training = rnn.fit(X_train,
                        y_train,
                        epochs=150,
                        batch_size=20,
                        shuffle=True,
                        validation_data=(X_test, y_test),
                        callbacks=[early_stopping, checkpointer],
                        verbose=1)
```

```
Epoch 1/150
```

```
95/96 [============================>.] - ETA: 0s - loss: 0.0162
Epoch 00001: val_loss improved from inf to 0.00766, saving model to
results/univariate_time_series/rnn.h5
96/96 [==============================] - 1s 12ms/step - loss: 0.0161 - val_loss:
0.0077
Epoch 2/150
95/96 [============================>.] - ETA: 0s - loss: 5.0726e-04
Epoch 00002: val_loss improved from 0.00766 to 0.00135, saving model to
results/univariate_time_series/rnn.h5
96/96 [==============================] - 1s 9ms/step - loss: 5.0613e-04 -
val_loss: 0.0014
Epoch 3/150
94/96 [============================>.] - ETA: 0s - loss: 4.2700e-04
Epoch 00003: val_loss did not improve from 0.00135
96/96 [==============================] - 1s 9ms/step - loss: 4.2515e-04 -
val_loss: 0.0033
Epoch 4/150
94/96 [============================>.] - ETA: 0s - loss: 4.0140e-04
Epoch 00004: val_loss did not improve from 0.00135
96/96 [==============================] - 1s 10ms/step - loss: 3.9946e-04 -
val_loss: 0.0022
Epoch 5/150
91/96 [===========================>..] - ETA: 0s - loss: 3.7595e-04
Epoch 00005: val_loss did not improve from 0.00135
96/96 [==============================] - 1s 9ms/step - loss: 3.6440e-04 -
val_loss: 0.0033
Epoch 6/150
96/96 [==============================] - ETA: 0s - loss: 3.4672e-04
Epoch 00006: val_loss improved from 0.00135 to 0.00068, saving model to
results/univariate_time_series/rnn.h5
96/96 [==============================] - 1s 9ms/step - loss: 3.4672e-04 -
val_loss: 6.7836e-04
Epoch 7/150
95/96 [============================>.] - ETA: 0s - loss: 3.1172e-04
Epoch 00007: val_loss did not improve from 0.00068
96/96 [==============================] - 1s 9ms/step - loss: 3.1417e-04 -
val_loss: 0.0047
Epoch 8/150
93/96 [============================>.] - ETA: 0s - loss: 3.2092e-04
Epoch 00008: val_loss did not improve from 0.00068
96/96 [==============================] - 1s 9ms/step - loss: 3.1923e-04 -
val_loss: 0.0014
Epoch 9/150
95/96 [============================>.] - ETA: 0s - loss: 2.9657e-04
Epoch 00009: val_loss improved from 0.00068 to 0.00043, saving model to
results/univariate_time_series/rnn.h5
96/96 [==============================] - 1s 9ms/step - loss: 2.9568e-04 -
val_loss: 4.3396e-04
```

```
Epoch 10/150
91/96 [===========================>..] - ETA: 0s - loss: 2.8456e-04
Epoch 00010: val_loss did not improve from 0.00043
96/96 [==============================] - 1s 9ms/step - loss: 2.8585e-04 -
val_loss: 0.0016
Epoch 11/150
92/96 [===========================>..] - ETA: 0s - loss: 2.5980e-04
Epoch 00011: val_loss improved from 0.00043 to 0.00032, saving model to
results/univariate_time_series/rnn.h5
96/96 [==============================] - 1s 9ms/step - loss: 2.6074e-04 -
val_loss: 3.1798e-04
Epoch 12/150
91/96 [===========================>..] - ETA: 0s - loss: 2.6164e-04
Epoch 00012: val_loss did not improve from 0.00032
96/96 [==============================] - 1s 9ms/step - loss: 2.5868e-04 -
val_loss: 4.8836e-04
Epoch 13/150
96/96 [==============================] - ETA: 0s - loss: 2.5184e-04
Epoch 00013: val_loss did not improve from 0.00032
96/96 [==============================] - 1s 9ms/step - loss: 2.5184e-04 -
val_loss: 4.2231e-04
Epoch 14/150
95/96 [============================>.] - ETA: 0s - loss: 2.4671e-04
Epoch 00014: val_loss did not improve from 0.00032
96/96 [==============================] - 1s 10ms/step - loss: 2.4586e-04 -
val_loss: 4.4436e-04
Epoch 15/150
91/96 [===========================>..] - ETA: 0s - loss: 2.3177e-04
Epoch 00015: val_loss did not improve from 0.00032
96/96 [==============================] - 1s 9ms/step - loss: 2.3762e-04 -
val_loss: 4.7206e-04
Epoch 16/150
92/96 [===========================>..] - ETA: 0s - loss: 2.2798e-04
Epoch 00016: val_loss did not improve from 0.00032
96/96 [==============================] - 1s 9ms/step - loss: 2.2959e-04 -
val_loss: 3.2628e-04
Epoch 17/150
92/96 [===========================>..] - ETA: 0s - loss: 2.2682e-04
Epoch 00017: val_loss did not improve from 0.00032
96/96 [==============================] - 1s 10ms/step - loss: 2.2815e-04 -
val_loss: 0.0013
Epoch 18/150
96/96 [==============================] - ETA: 0s - loss: 2.1929e-04
Epoch 00018: val_loss did not improve from 0.00032
96/96 [==============================] - 1s 9ms/step - loss: 2.1929e-04 -
val_loss: 0.0022
Epoch 19/150
91/96 [===========================>..] - ETA: 0s - loss: 2.1801e-04
```

```
Epoch 00019: val_loss improved from 0.00032 to 0.00024, saving model to
results/univariate_time_series/rnn.h5
96/96 [==============================] - 1s 9ms/step - loss: 2.1470e-04 -
val_loss: 2.4200e-04
Epoch 20/150
96/96 [==============================] - ETA: 0s - loss: 2.1644e-04
Epoch 00020: val_loss improved from 0.00024 to 0.00023, saving model to
results/univariate_time_series/rnn.h5
96/96 [==============================] - 1s 10ms/step - loss: 2.1644e-04 -
val_loss: 2.3101e-04
Epoch 21/150
96/96 [==============================] - ETA: 0s - loss: 2.0451e-04
Epoch 00021: val_loss improved from 0.00023 to 0.00021, saving model to
results/univariate_time_series/rnn.h5
96/96 [==============================] - 1s 9ms/step - loss: 2.0451e-04 -
val_loss: 2.1255e-04
Epoch 22/150
91/96 [===========================>..] - ETA: 0s - loss: 2.0134e-04
Epoch 00022: val_loss did not improve from 0.00021
96/96 [==============================] - 1s 9ms/step - loss: 2.0179e-04 -
val_loss: 2.2027e-04
Epoch 23/150
96/96 [==============================] - ETA: 0s - loss: 1.9941e-04
Epoch 00023: val_loss did not improve from 0.00021
96/96 [==============================] - 1s 9ms/step - loss: 1.9941e-04 -
val_loss: 4.4025e-04
Epoch 24/150
96/96 [==============================] - ETA: 0s - loss: 1.9096e-04
Epoch 00024: val_loss did not improve from 0.00021
96/96 [==============================] - 1s 9ms/step - loss: 1.9096e-04 -
val_loss: 5.4885e-04
Epoch 25/150
96/96 [==============================] - ETA: 0s - loss: 1.8358e-04
Epoch 00025: val_loss did not improve from 0.00021
96/96 [==============================] - 1s 9ms/step - loss: 1.8358e-04 -
val_loss: 2.4444e-04
Epoch 26/150
95/96 [============================>.] - ETA: 0s - loss: 1.8497e-04
Epoch 00026: val_loss did not improve from 0.00021
96/96 [==============================] - 1s 10ms/step - loss: 1.8471e-04 -
val_loss: 4.2620e-04
Epoch 27/150
96/96 [==============================] - ETA: 0s - loss: 1.7350e-04
Epoch 00027: val_loss did not improve from 0.00021
96/96 [==============================] - 1s 9ms/step - loss: 1.7350e-04 -
val_loss: 4.0677e-04
Epoch 28/150
91/96 [===========================>..] - ETA: 0s - loss: 1.6671e-04
```

```
Epoch 00028: val_loss did not improve from 0.00021
96/96 [==============================] - 1s 9ms/step - loss: 1.6921e-04 -
val_loss: 3.8056e-04
Epoch 29/150
94/96 [============================>.] - ETA: 0s - loss: 1.6757e-04
Epoch 00029: val_loss did not improve from 0.00021
96/96 [==============================] - 1s 10ms/step - loss: 1.7095e-04 -
val_loss: 3.5113e-04
Epoch 30/150
93/96 [============================>.] - ETA: 0s - loss: 1.5892e-04
Epoch 00030: val_loss improved from 0.00021 to 0.00018, saving model to
results/univariate_time_series/rnn.h5
96/96 [==============================] - 1s 10ms/step - loss: 1.5886e-04 -
val_loss: 1.7758e-04
Epoch 31/150
92/96 [===========================>..] - ETA: 0s - loss: 1.5301e-04
Epoch 00031: val_loss improved from 0.00018 to 0.00016, saving model to
results/univariate_time_series/rnn.h5
96/96 [==============================] - 1s 10ms/step - loss: 1.5573e-04 -
val_loss: 1.5858e-04
Epoch 32/150
95/96 [============================>.] - ETA: 0s - loss: 1.4994e-04
Epoch 00032: val_loss improved from 0.00016 to 0.00016, saving model to
results/univariate_time_series/rnn.h5
96/96 [==============================] - 1s 10ms/step - loss: 1.5008e-04 -
val_loss: 1.5702e-04
Epoch 33/150
94/96 [============================>.] - ETA: 0s - loss: 1.5448e-04
Epoch 00033: val_loss did not improve from 0.00016
96/96 [==============================] - 1s 10ms/step - loss: 1.5528e-04 -
val_loss: 3.3243e-04
Epoch 34/150
94/96 [============================>.] - ETA: 0s - loss: 1.5249e-04
Epoch 00034: val_loss improved from 0.00016 to 0.00014, saving model to
results/univariate_time_series/rnn.h5
96/96 [==============================] - 1s 10ms/step - loss: 1.5086e-04 -
val_loss: 1.4220e-04
Epoch 35/150
96/96 [==============================] - ETA: 0s - loss: 1.4350e-04
Epoch 00035: val_loss did not improve from 0.00014
96/96 [==============================] - 1s 9ms/step - loss: 1.4350e-04 -
val_loss: 8.7128e-04
Epoch 36/150
91/96 [===========================>..] - ETA: 0s - loss: 1.4087e-04
Epoch 00036: val_loss did not improve from 0.00014
96/96 [==============================] - 1s 9ms/step - loss: 1.4304e-04 -
val_loss: 2.2118e-04
Epoch 37/150
```

```
93/96 [============================>.] - ETA: 0s - loss: 1.4318e-04
Epoch 00037: val_loss did not improve from 0.00014
96/96 [==============================] - 1s 9ms/step - loss: 1.4475e-04 -
val_loss: 5.7758e-04
Epoch 38/150
91/96 [===========================>..] - ETA: 0s - loss: 1.3891e-04
Epoch 00038: val_loss did not improve from 0.00014
96/96 [==============================] - 1s 9ms/step - loss: 1.4370e-04 -
val_loss: 5.5123e-04
Epoch 39/150
96/96 [==============================] - ETA: 0s - loss: 1.3512e-04
Epoch 00039: val_loss did not improve from 0.00014
96/96 [==============================] - 1s 9ms/step - loss: 1.3512e-04 -
val_loss: 2.0821e-04
Epoch 40/150
92/96 [===========================>..] - ETA: 0s - loss: 1.3034e-04
Epoch 00040: val_loss did not improve from 0.00014
96/96 [==============================] - 1s 10ms/step - loss: 1.3073e-04 -
val_loss: 6.1821e-04
Epoch 41/150
95/96 [============================>.] - ETA: 0s - loss: 1.3208e-04
Epoch 00041: val_loss did not improve from 0.00014
96/96 [==============================] - 1s 10ms/step - loss: 1.3247e-04 -
val_loss: 5.5452e-04
Epoch 42/150
95/96 [============================>.] - ETA: 0s - loss: 1.2555e-04
Epoch 00042: val_loss did not improve from 0.00014
96/96 [==============================] - 1s 10ms/step - loss: 1.2543e-04 -
val_loss: 3.1468e-04
Epoch 43/150
92/96 [===========================>..] - ETA: 0s - loss: 1.2727e-04
Epoch 00043: val_loss did not improve from 0.00014
96/96 [==============================] - 1s 12ms/step - loss: 1.2745e-04 -
val_loss: 2.5483e-04
Epoch 44/150
96/96 [==============================] - ETA: 0s - loss: 1.3116e-04
Epoch 00044: val_loss did not improve from 0.00014
96/96 [==============================] - 1s 11ms/step - loss: 1.3116e-04 -
val_loss: 2.0916e-04
Epoch 45/150
93/96 [============================>.] - ETA: 0s - loss: 1.2340e-04
Epoch 00045: val_loss did not improve from 0.00014
96/96 [==============================] - 1s 12ms/step - loss: 1.2624e-04 -
val_loss: 1.5523e-04
Epoch 46/150
95/96 [============================>.] - ETA: 0s - loss: 1.2741e-04
Epoch 00046: val_loss did not improve from 0.00014
96/96 [==============================] - 1s 16ms/step - loss: 1.2674e-04 -
```

```
val_loss: 1.4887e-04
Epoch 47/150
92/96 [===========================>..] - ETA: 0s - loss: 1.2104e-04
Epoch 00047: val_loss did not improve from 0.00014
96/96 [==============================] - 2s 16ms/step - loss: 1.2028e-04 -
val_loss: 3.1503e-04
Epoch 48/150
93/96 [============================>.] - ETA: 0s - loss: 1.2040e-04
Epoch 00048: val_loss did not improve from 0.00014
96/96 [==============================] - 1s 10ms/step - loss: 1.2169e-04 -
val_loss: 3.1043e-04
Epoch 49/150
96/96 [==============================] - ETA: 0s - loss: 1.2513e-04
Epoch 00049: val_loss did not improve from 0.00014
96/96 [==============================] - 1s 10ms/step - loss: 1.2513e-04 -
val_loss: 3.6431e-04
Epoch 50/150
93/96 [============================>.] - ETA: 0s - loss: 1.1566e-04
Epoch 00050: val_loss did not improve from 0.00014
96/96 [==============================] - 1s 10ms/step - loss: 1.1806e-04 -
val_loss: 2.0745e-04
Epoch 51/150
96/96 [==============================] - ETA: 0s - loss: 1.1787e-04
Epoch 00051: val_loss did not improve from 0.00014
96/96 [==============================] - 1s 8ms/step - loss: 1.1787e-04 -
val_loss: 3.8254e-04
Epoch 52/150
94/96 [============================>.] - ETA: 0s - loss: 1.1391e-04
Epoch 00052: val_loss improved from 0.00014 to 0.00013, saving model to
results/univariate_time_series/rnn.h5
96/96 [==============================] - 1s 9ms/step - loss: 1.1292e-04 -
val_loss: 1.3098e-04
Epoch 53/150
95/96 [============================>.] - ETA: 0s - loss: 1.1152e-04
Epoch 00053: val_loss did not improve from 0.00013
96/96 [==============================] - 1s 8ms/step - loss: 1.1118e-04 -
val_loss: 3.5197e-04
Epoch 54/150
90/96 [===========================>..] - ETA: 0s - loss: 1.1793e-04
Epoch 00054: val_loss did not improve from 0.00013
96/96 [==============================] - 1s 9ms/step - loss: 1.1724e-04 -
val_loss: 1.9148e-04
Epoch 55/150
94/96 [============================>.] - ETA: 0s - loss: 1.1052e-04
Epoch 00055: val_loss did not improve from 0.00013
96/96 [==============================] - 1s 9ms/step - loss: 1.1202e-04 -
val_loss: 2.7836e-04
Epoch 56/150
```

```
92/96 [===========================>..] - ETA: 0s - loss: 1.0805e-04
Epoch 00056: val_loss did not improve from 0.00013
96/96 [==============================] - 1s 9ms/step - loss: 1.0802e-04 -
val_loss: 1.3425e-04
Epoch 57/150
94/96 [============================>.] - ETA: 0s - loss: 1.1062e-04
Epoch 00057: val_loss improved from 0.00013 to 0.00012, saving model to
results/univariate_time_series/rnn.h5
96/96 [==============================] - 1s 9ms/step - loss: 1.0968e-04 -
val_loss: 1.1734e-04
Epoch 58/150
90/96 [===========================>..] - ETA: 0s - loss: 1.0471e-04
Epoch 00058: val_loss did not improve from 0.00012
96/96 [==============================] - 1s 9ms/step - loss: 1.0535e-04 -
val_loss: 2.0888e-04
Epoch 59/150
93/96 [============================>.] - ETA: 0s - loss: 1.0651e-04
Epoch 00059: val_loss did not improve from 0.00012
96/96 [==============================] - 1s 9ms/step - loss: 1.0444e-04 -
val_loss: 1.2834e-04
Epoch 60/150
96/96 [==============================] - ETA: 0s - loss: 1.0868e-04
Epoch 00060: val_loss did not improve from 0.00012
96/96 [==============================] - 1s 8ms/step - loss: 1.0868e-04 -
val_loss: 1.7766e-04
Epoch 61/150
95/96 [============================>.] - ETA: 0s - loss: 1.0337e-04
Epoch 00061: val_loss did not improve from 0.00012
96/96 [==============================] - 1s 9ms/step - loss: 1.0307e-04 -
val_loss: 2.2622e-04
Epoch 62/150
91/96 [===========================>..] - ETA: 0s - loss: 1.0358e-04
Epoch 00062: val_loss did not improve from 0.00012
96/96 [==============================] - 1s 9ms/step - loss: 1.0332e-04 -
val_loss: 1.1764e-04
Epoch 63/150
95/96 [============================>.] - ETA: 0s - loss: 1.0141e-04
Epoch 00063: val_loss did not improve from 0.00012
96/96 [==============================] - 1s 9ms/step - loss: 1.0113e-04 -
val_loss: 1.7721e-04
Epoch 64/150
95/96 [============================>.] - ETA: 0s - loss: 1.0574e-04
Epoch 00064: val_loss did not improve from 0.00012
96/96 [==============================] - 1s 9ms/step - loss: 1.0589e-04 -
val_loss: 2.7786e-04
Epoch 65/150
91/96 [===========================>..] - ETA: 0s - loss: 9.9946e-05
Epoch 00065: val_loss did not improve from 0.00012
```

```
96/96 [==============================] - 1s 9ms/step - loss: 9.8424e-05 -
val_loss: 2.5257e-04
Epoch 66/150
93/96 [============================>.] - ETA: 0s - loss: 1.0225e-04
Epoch 00066: val_loss did not improve from 0.00012
96/96 [==============================] - 1s 9ms/step - loss: 1.0111e-04 -
val_loss: 1.2785e-04
Epoch 67/150
95/96 [============================>.] - ETA: 0s - loss: 1.0120e-04
Epoch 00067: val_loss did not improve from 0.00012
96/96 [==============================] - 1s 9ms/step - loss: 1.0110e-04 -
val_loss: 1.5218e-04
Epoch 68/150
96/96 [==============================] - ETA: 0s - loss: 9.5239e-05
Epoch 00068: val_loss improved from 0.00012 to 0.00011, saving model to
results/univariate_time_series/rnn.h5
96/96 [==============================] - 1s 9ms/step - loss: 9.5239e-05 -
val_loss: 1.0982e-04
Epoch 69/150
95/96 [============================>.] - ETA: 0s - loss: 9.7534e-05
Epoch 00069: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.7244e-05 -
val_loss: 1.3820e-04
Epoch 70/150
94/96 [============================>.] - ETA: 0s - loss: 1.0033e-04
Epoch 00070: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 1.0096e-04 -
val_loss: 2.6083e-04
Epoch 71/150
94/96 [============================>.] - ETA: 0s - loss: 9.8744e-05
Epoch 00071: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.8619e-05 -
val_loss: 1.4675e-04
Epoch 72/150
93/96 [============================>.] - ETA: 0s - loss: 9.8295e-05
Epoch 00072: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.7614e-05 -
val_loss: 1.7677e-04
Epoch 73/150
95/96 [============================>.] - ETA: 0s - loss: 9.5597e-05
Epoch 00073: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.5436e-05 -
val_loss: 2.8398e-04
Epoch 74/150
91/96 [===========================>..] - ETA: 0s - loss: 9.7810e-05
Epoch 00074: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.7079e-05 -
val_loss: 1.4352e-04
```

```
Epoch 75/150
92/96 [===========================>..] - ETA: 0s - loss: 9.5967e-05
Epoch 00075: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.6704e-05 -
val_loss: 1.2011e-04
Epoch 76/150
92/96 [===========================>..] - ETA: 0s - loss: 9.8459e-05
Epoch 00076: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.8009e-05 -
val_loss: 1.7817e-04
Epoch 77/150
96/96 [==============================] - ETA: 0s - loss: 9.1118e-05
Epoch 00077: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.1118e-05 -
val_loss: 1.1157e-04
Epoch 78/150
95/96 [============================>.] - ETA: 0s - loss: 9.5011e-05
Epoch 00078: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.5155e-05 -
val_loss: 1.5538e-04
Epoch 79/150
96/96 [==============================] - ETA: 0s - loss: 9.4418e-05
Epoch 00079: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.4418e-05 -
val_loss: 1.4241e-04
Epoch 80/150
94/96 [============================>.] - ETA: 0s - loss: 9.4359e-05
Epoch 00080: val_loss improved from 0.00011 to 0.00011, saving model to
results/univariate_time_series/rnn.h5
96/96 [==============================] - 1s 9ms/step - loss: 9.4327e-05 -
val_loss: 1.0896e-04
Epoch 81/150
95/96 [============================>.] - ETA: 0s - loss: 9.4806e-05
Epoch 00081: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.4769e-05 -
val_loss: 1.3463e-04
Epoch 82/150
95/96 [============================>.] - ETA: 0s - loss: 9.3654e-05
Epoch 00082: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.4321e-05 -
val_loss: 1.4250e-04
Epoch 83/150
93/96 [============================>.] - ETA: 0s - loss: 9.5975e-05
Epoch 00083: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.6019e-05 -
val_loss: 1.1075e-04
Epoch 84/150
91/96 [===========================>..] - ETA: 0s - loss: 9.5794e-05
```

```
Epoch 00084: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.6212e-05 -
val_loss: 2.0834e-04
Epoch 85/150
91/96 [===========================>..] - ETA: 0s - loss: 8.9862e-05
Epoch 00085: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.7901e-05 -
val_loss: 1.2320e-04
Epoch 86/150
95/96 [============================>.] - ETA: 0s - loss: 9.0540e-05
Epoch 00086: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 8ms/step - loss: 9.0288e-05 -
val_loss: 1.1778e-04
Epoch 87/150
91/96 [===========================>..] - ETA: 0s - loss: 9.1800e-05
Epoch 00087: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.0853e-05 -
val_loss: 1.8410e-04
Epoch 88/150
94/96 [============================>.] - ETA: 0s - loss: 9.0630e-05
Epoch 00088: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.1169e-05 -
val_loss: 1.5173e-04
Epoch 89/150
95/96 [============================>.] - ETA: 0s - loss: 9.0615e-05
Epoch 00089: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.0969e-05 -
val_loss: 2.6647e-04
Epoch 90/150
91/96 [===========================>..] - ETA: 0s - loss: 8.4236e-05
Epoch 00090: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 10ms/step - loss: 8.6761e-05 -
val_loss: 2.2579e-04
Epoch 91/150
95/96 [============================>.] - ETA: 0s - loss: 9.0738e-05
Epoch 00091: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.0392e-05 -
val_loss: 1.4102e-04
Epoch 92/150
91/96 [===========================>..] - ETA: 0s - loss: 8.7762e-05
Epoch 00092: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.1175e-05 -
val_loss: 1.5866e-04
Epoch 93/150
92/96 [===========================>..] - ETA: 0s - loss: 8.8039e-05
Epoch 00093: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.9345e-05 -
val_loss: 2.2237e-04
```

```
Epoch 94/150
92/96 [===========================>..] - ETA: 0s - loss: 8.7299e-05
Epoch 00094: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.8101e-05 -
val_loss: 2.0188e-04
Epoch 95/150
91/96 [===========================>..] - ETA: 0s - loss: 8.7247e-05
Epoch 00095: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.7141e-05 -
val_loss: 1.1806e-04
Epoch 96/150
95/96 [============================>.] - ETA: 0s - loss: 9.1653e-05
Epoch 00096: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.1519e-05 -
val_loss: 1.1541e-04
Epoch 97/150
91/96 [===========================>..] - ETA: 0s - loss: 8.8514e-05
Epoch 00097: val_loss improved from 0.00011 to 0.00011, saving model to
results/univariate_time_series/rnn.h5
96/96 [==============================] - 1s 9ms/step - loss: 8.6380e-05 -
val_loss: 1.0814e-04
Epoch 98/150
96/96 [==============================] - ETA: 0s - loss: 8.6094e-05
Epoch 00098: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.6094e-05 -
val_loss: 1.1347e-04
Epoch 99/150
90/96 [===========================>..] - ETA: 0s - loss: 8.7294e-05
Epoch 00099: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.9347e-05 -
val_loss: 1.3592e-04
Epoch 100/150
96/96 [==============================] - ETA: 0s - loss: 8.6722e-05
Epoch 00100: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.6722e-05 -
val_loss: 7.0361e-04
Epoch 101/150
92/96 [===========================>..] - ETA: 0s - loss: 9.0485e-05
Epoch 00101: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.9286e-05 -
val_loss: 1.1654e-04
Epoch 102/150
94/96 [============================>.] - ETA: 0s - loss: 8.7473e-05
Epoch 00102: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.6557e-05 -
val_loss: 1.1325e-04
Epoch 103/150
94/96 [============================>.] - ETA: 0s - loss: 8.9702e-05
```

```
Epoch 00103: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.9611e-05 -
val_loss: 2.9964e-04
Epoch 104/150
91/96 [===========================>..] - ETA: 0s - loss: 8.8178e-05
Epoch 00104: val_loss improved from 0.00011 to 0.00011, saving model to
results/univariate_time_series/rnn.h5
96/96 [==============================] - 1s 9ms/step - loss: 8.8027e-05 -
val_loss: 1.0701e-04
Epoch 105/150
90/96 [===========================>..] - ETA: 0s - loss: 8.4287e-05
Epoch 00105: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.5441e-05 -
val_loss: 2.6547e-04
Epoch 106/150
93/96 [============================>.] - ETA: 0s - loss: 8.5374e-05
Epoch 00106: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.6718e-05 -
val_loss: 1.4452e-04
Epoch 107/150
92/96 [===========================>..] - ETA: 0s - loss: 8.6347e-05
Epoch 00107: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.8267e-05 -
val_loss: 1.1976e-04
Epoch 108/150
95/96 [============================>.] - ETA: 0s - loss: 8.7606e-05
Epoch 00108: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.7117e-05 -
val_loss: 1.0868e-04
Epoch 109/150
96/96 [==============================] - ETA: 0s - loss: 9.0933e-05
Epoch 00109: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.0933e-05 -
val_loss: 1.1853e-04
Epoch 110/150
90/96 [===========================>..] - ETA: 0s - loss: 8.8712e-05
Epoch 00110: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.6446e-05 -
val_loss: 1.0818e-04
Epoch 111/150
95/96 [============================>.] - ETA: 0s - loss: 8.7685e-05
Epoch 00111: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.8124e-05 -
val_loss: 2.3456e-04
Epoch 112/150
91/96 [===========================>..] - ETA: 0s - loss: 8.6183e-05
Epoch 00112: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.6186e-05 -
```

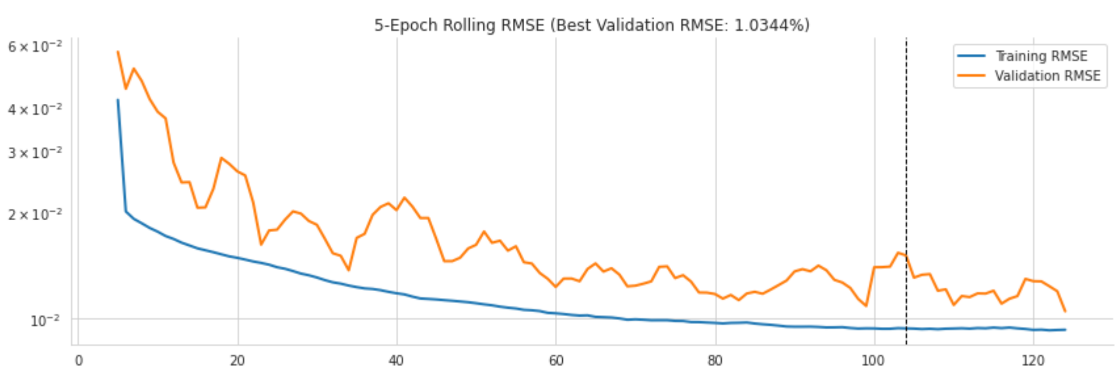
```
val_loss: 1.1133e-04
Epoch 113/150
96/96 [==============================] - ETA: 0s - loss: 9.0423e-05
Epoch 00113: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.0423e-05 -
val_loss: 1.3993e-04
Epoch 114/150
94/96 [============================>.] - ETA: 0s - loss: 8.8848e-05
Epoch 00114: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.9644e-05 -
val_loss: 1.1688e-04
Epoch 115/150
95/96 [============================>.] - ETA: 0s - loss: 9.0703e-05
Epoch 00115: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 9.0592e-05 -
val_loss: 1.3260e-04
Epoch 116/150
92/96 [===========================>..] - ETA: 0s - loss: 8.5621e-05
Epoch 00116: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.4973e-05 -
val_loss: 1.0909e-04
Epoch 117/150
91/96 [===========================>..] - ETA: 0s - loss: 8.9659e-05
Epoch 00117: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.9528e-05 -
val_loss: 1.5286e-04
Epoch 118/150
96/96 [==============================] - ETA: 0s - loss: 8.5693e-05
Epoch 00118: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.5693e-05 -
val_loss: 1.6449e-04
Epoch 119/150
93/96 [============================>.] - ETA: 0s - loss: 8.7340e-05
Epoch 00119: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.6209e-05 -
val_loss: 3.1345e-04
Epoch 120/150
92/96 [===========================>..] - ETA: 0s - loss: 8.4168e-05
Epoch 00120: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.5270e-05 -
val_loss: 1.1120e-04
Epoch 121/150
93/96 [============================>.] - ETA: 0s - loss: 8.6140e-05
Epoch 00121: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.6331e-05 -
val_loss: 1.0745e-04
Epoch 122/150
91/96 [===========================>..] - ETA: 0s - loss: 8.7073e-05
```

```
Epoch 00122: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.5809e-05 -
val_loss: 1.0791e-04
Epoch 123/150
92/96 [===========================>..] - ETA: 0s - loss: 8.7627e-05
Epoch 00123: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 10ms/step - loss: 8.7544e-05 -
val_loss: 1.1710e-04
Epoch 124/150
92/96 [===========================>..] - ETA: 0s - loss: 8.9236e-05
Epoch 00124: val_loss did not improve from 0.00011
96/96 [==============================] - 1s 9ms/step - loss: 8.7432e-05 -
val_loss: 1.0717e-04
```

Training stops after 51 epochs; the `early_stopping` callback restores the weights for the best model (after 41 epochs)

1.10 Evaluate model performance

[31]:
```python
fig, ax = plt.subplots(figsize=(12, 4))

loss_history = pd.DataFrame(lstm_training.history).pow(.5)
loss_history.index += 1
best_rmse = loss_history.val_loss.min()

best_epoch = loss_history.val_loss.idxmin()

title = f'5-Epoch Rolling RMSE (Best Validation RMSE: {best_rmse:.4%})'
loss_history.columns=['Training RMSE', 'Validation RMSE']
loss_history.rolling(5).mean().plot(logy=True, lw=2, title=title, ax=ax)

ax.axvline(best_epoch, ls='--', lw=1, c='k')

sns.despine()
fig.tight_layout()
fig.savefig(results_path / 'rnn_sp500_error', dpi=300);
```



[32]:
```
train_rmse_scaled = np.sqrt(rnn.evaluate(X_train, y_train, verbose=0))
test_rmse_scaled = np.sqrt(rnn.evaluate(X_test, y_test, verbose=0))
print(f'Train RMSE: {train_rmse_scaled:.4f} | Test RMSE: {test_rmse_scaled:.
↪4f}')
```

```
Train RMSE: 0.0085 | Test RMSE: 0.0103
```

[33]:
```
train_predict_scaled = rnn.predict(X_train)
test_predict_scaled = rnn.predict(X_test)
```

[34]:
```
train_ic = spearmanr(y_train, train_predict_scaled)[0]
test_ic = spearmanr(y_test, test_predict_scaled)[0]
print(f'Train IC: {train_ic:.4f} | Test IC: {test_ic:.4f}')
```

```
Train IC: 0.9986 | Test IC: 0.9817
```

1.10.1 Rescale predictions

[35]:
```
train_predict = pd.Series(scaler.inverse_transform(train_predict_scaled).
↪squeeze(), index=y_train.index)
test_predict = (pd.Series(scaler.inverse_transform(test_predict_scaled)
                          .squeeze(),
                          index=y_test.index))
```

[36]:
```
y_train_rescaled = scaler.inverse_transform(y_train.to_frame()).squeeze()
y_test_rescaled = scaler.inverse_transform(y_test.to_frame()).squeeze()
```

[37]:
```
train_rmse = np.sqrt(mean_squared_error(train_predict, y_train_rescaled))
test_rmse = np.sqrt(mean_squared_error(test_predict, y_test_rescaled))
f'Train RMSE: {train_rmse:.2f} | Test RMSE: {test_rmse:.2f}'
```

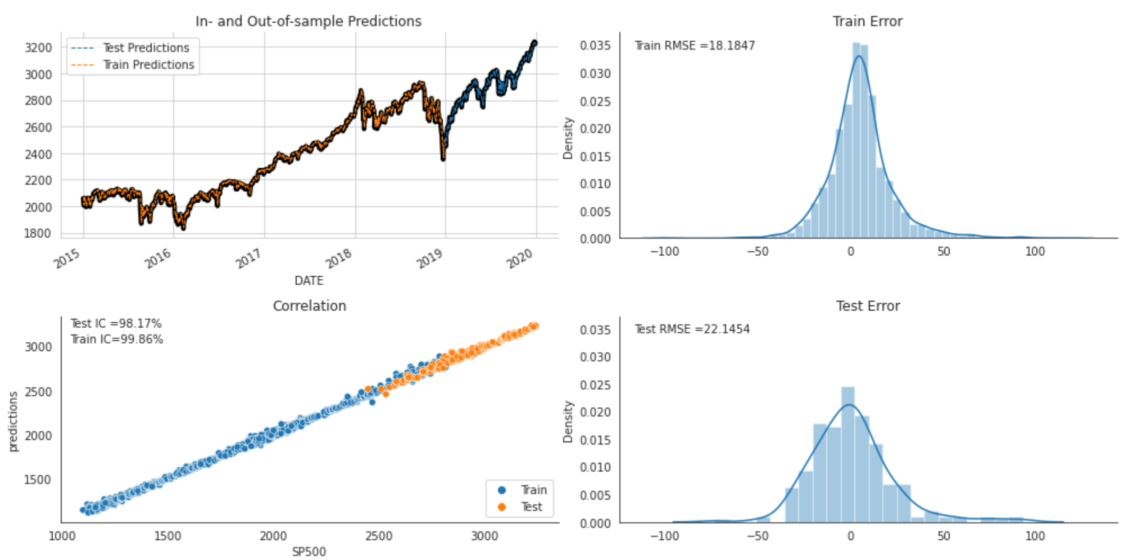
[37]:
```
'Train RMSE: 18.18 | Test RMSE: 22.15'
```

[38]:
```
sp500['Train Predictions'] = train_predict
sp500['Test Predictions'] = test_predict
sp500 = sp500.join(train_predict.to_frame('predictions').assign(data='Train')
                        .append(test_predict.to_frame('predictions').
↪assign(data='Test')))
```

1.10.2 Plot Results

[39]:
```
fig=plt.figure(figsize=(14,7))
ax1 = plt.subplot(221)

sp500.loc['2015':, 'SP500'].plot(lw=4, ax=ax1, c='k')
```

```python
sp500.loc['2015':, ['Test Predictions', 'Train Predictions']].plot(lw=1, ax=ax1, ls='--')
ax1.set_title('In- and Out-of-sample Predictions')


with sns.axes_style("white"):
    ax3 = plt.subplot(223)
    sns.scatterplot(x='SP500', y='predictions', data=sp500, hue='data', ax=ax3)
    ax3.text(x=.02, y=.95, s=f'Test IC ={test_ic:.2%}', transform=ax3.transAxes)
    ax3.text(x=.02, y=.87, s=f'Train IC={train_ic:.2%}', transform=ax3.transAxes)
    ax3.set_title('Correlation')
    ax3.legend(loc='lower right')

    ax2 = plt.subplot(222)
    ax4 = plt.subplot(224, sharex = ax2, sharey=ax2)
    sns.distplot(train_predict.squeeze()- y_train_rescaled, ax=ax2)
    ax2.set_title('Train Error')
    ax2.text(x=.03, y=.92, s=f'Train RMSE ={train_rmse:.4f}', transform=ax2.transAxes)
    sns.distplot(test_predict.squeeze()-y_test_rescaled, ax=ax4)
    ax4.set_title('Test Error')
    ax4.text(x=.03, y=.92, s=f'Test RMSE ={test_rmse:.4f}', transform=ax4.transAxes)

sns.despine()
fig.tight_layout()
fig.savefig(results_path / 'rnn_sp500_regression', dpi=300);
```



```
[ ]:
```