Forecasting Agricultural Commodity Prices Using CNN-GRU Networks with Likelihood Loss Functions

Yang Song

ysong26@ncsu.edu

North Carolina State University

Motivation

[2, 4] present the existence of autocorrelation and non-linear relationship among commodity price series. This article uses a CNN-GRU network model with a likelihood loss to improve the forecasting accuracy of agricultural commodities prices.

Model

Neural networks extract linear combinations of inputs as features and model the output as a nonlinear function of these features[3]. To better extract and store the features, the proposed neural network has the following three types of layers:

- **Convolutional layers**
- **Gated Recurrent Unit (GRU) layers**
- **Fully Connected (Dense) layers**

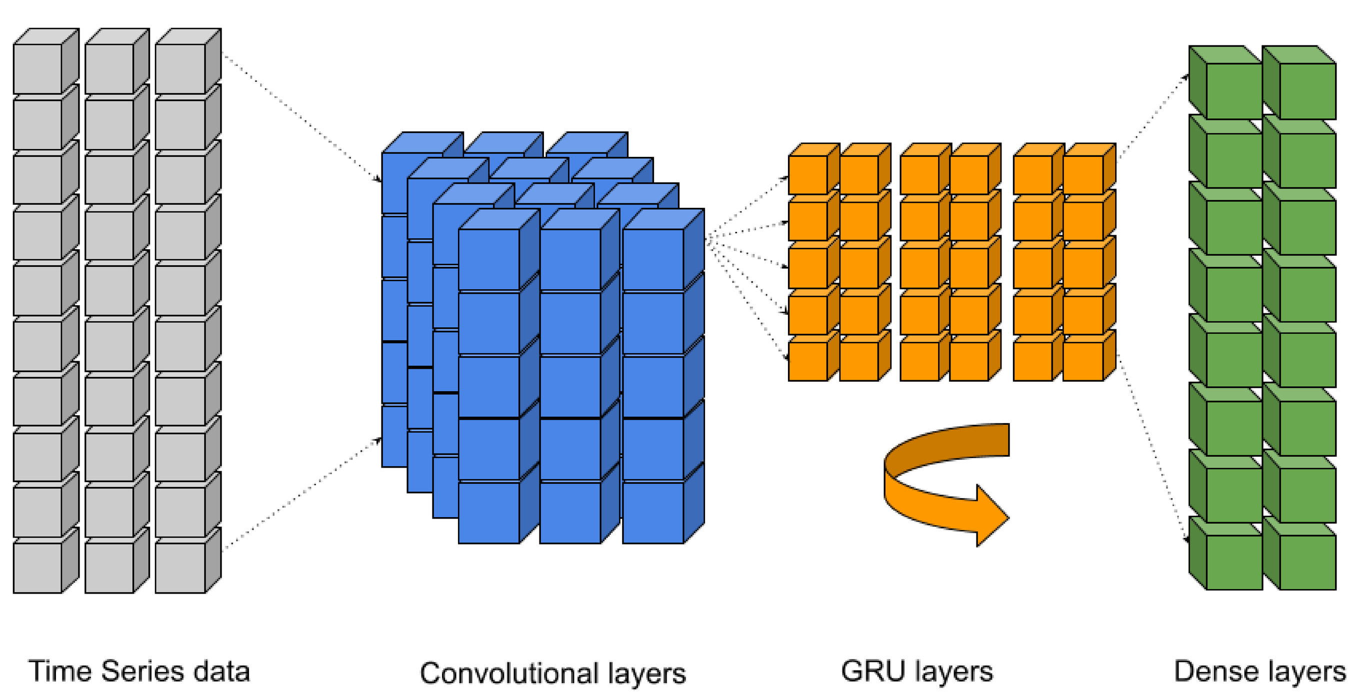

Fig. 1: CNN-GRU model

The architecture has three groups of layers, each designed for a specific task. The first group has a one-dimensional convolutional layer and a pooling layer. This group layer is responsible for pattern recognition and data smoothing. The second group has up to four GRU layers [1] whose primary task is to detect periodic series. The last set of layers are fully connected layers and they seize the nonlinear relationships between smoothed patterns and the output. Intuitively, this architecture dispatches multivariate time series into different channels and identifies recurring patterns that can be helpful in forecasting future returns. In addition, in [2], the likelihood loss reduces forecasting errors at the cost of a tolerable increase in epochs.

Negative Log Likelihood as the Loss Function

One of the contributions of this paper is using a negative likelihood function [3] to improve the performance of a CNN-GRU neural network. It is known that a Mean Squared Error loss function is equivalent to a negative likelihood loss if we assume homoskedasticity,
or $\sigma_t^2 = \sigma^2$, for all t.

$$
\begin{aligned}
L(w|y_1, y_2, ..., y_n) &= \prod_{i=1}^{n} P(y_i|w, x) \\
&= \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2\sigma^2}(y_i - \hat{y}_i)^2) \\
\log L &= -\frac{1}{2}\sum_{i=1}^{n}(\log 2\pi + \log \sigma^2 + \frac{(y_i - \hat{y}_i)^2}{\sigma^2}) \\
&= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + constant
\end{aligned}
$$

Where $\hat{y}_t = f(x_t)$ is the output of neural network. If we estimate σ_t^2 at each time step [5, 6], it can be shown that the forecasting error is lower with slightly more computation.

$$
\begin{aligned}
Loss(y, \hat{y}) &= -\log L \\
&= -\frac{1}{2}\sum_{t=1}^{n}(\log 2\pi + \log \sigma_t^2 + \frac{(y_t - f(x_t))^2}{\sigma_t^2}) \\
&= \frac{1}{2}\sum_{t=1}^{n}(\log \sigma_t^2 + \frac{(y_t - f(x_t))^2}{\sigma_t^2}) + C
\end{aligned}
$$

Model Training

Before making forecasts on the test set, a random grid search is performed on the parameter space by the training set to select an optimal structure of the CNN-GRU network. The model incorporates the incoming sample with historical model states, estimates the σ_t^2, fits the model in the current time step, and outputs the forecast value.

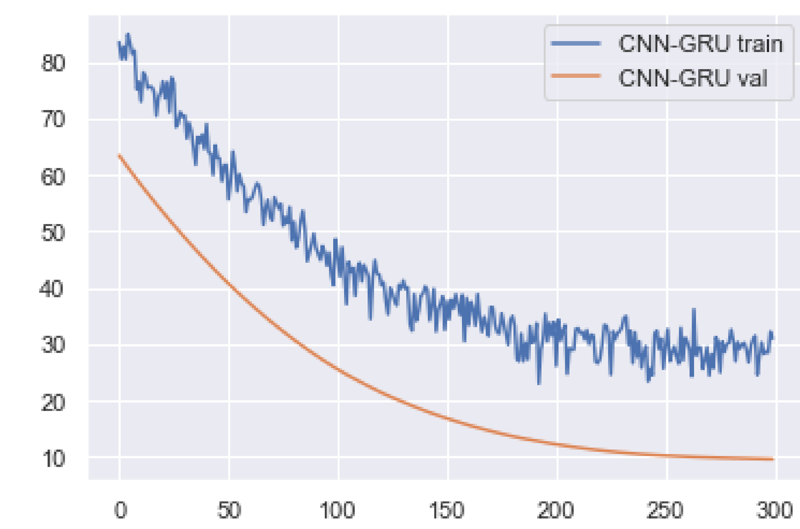

Fig. 2: Training with the likelihood loss

- More epochs in training($\approx$ 1.8 times of MSE loss)
- Higher accuracy ($\approx 5\%$ lower RMSE)
- Oscillation in loss requires regularization (L1/L2/Dropout)

Methodology

This article evaluates model performance in an out-of-sample rolling window fashion. The training set has 90% of the data, and the remaining 10% is the test set. The experiment compares the likelihood-based CNN-GRU model's forecasting accuracy against VAR, ARIMA, MLP, and MSE-based CNN-GRU.

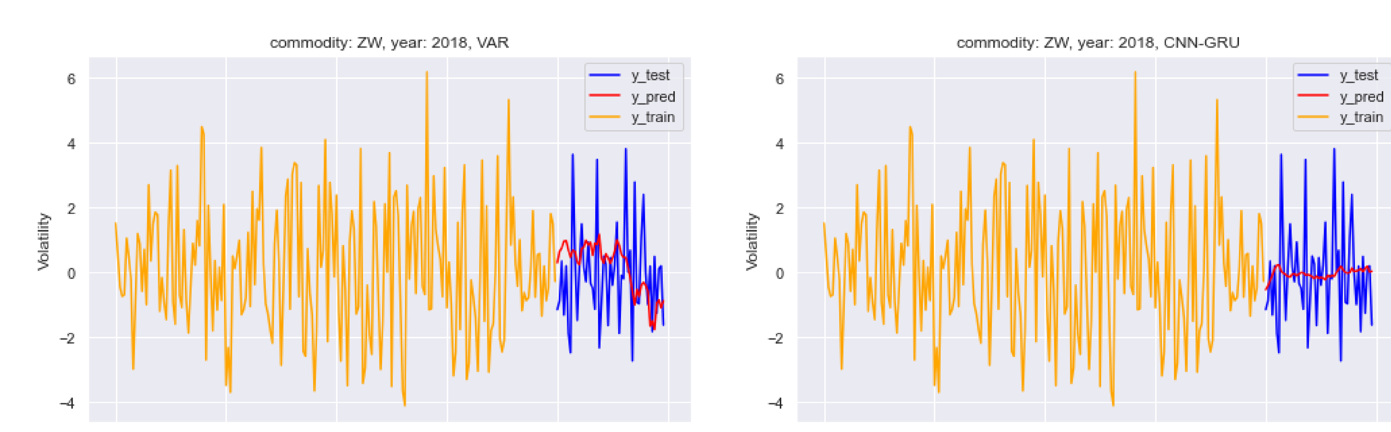

Fig. 3: Wheat(ZW) returns forecast comparison between VAR and CNN-GRU

Results

This study utilizes agricultural commodity futures prices with auxiliary variables from 2014 to 2020. Commodities include: **Corn, Wheat, Cotton, and Soybean**. The following table shows the **RMSE** of each model on commodities.

Commodity/Model	CNN-GRU (NLL)	CNN-GRU (MSE)	MLP	VAR
Corn	0.8602	1.2019	1.2388	0.9266
Wheat	1.5619	1.5446	1.8764	1.6581
Cotton	1.4853	1.5326	1.8965	1.6401
Soybean	1.1282	1.1561	1.3646	1.2333

Concluding Remarks

- A likelihood loss function can improve forecasting accuracy for a CNN-GRU model
- To handle the oscillation in training, the model must use regularization techniques at the cost of more training time
- The empirical result suggests that the proposed CNN-GRU model outperforms VAR, ARIMA, MLP, and MSE-based CNN-GRU in forecasting corn, wheat, soybean, and cotton returns from 2014 to 2019 in out-of-sample rolling window tests

References

[1] Kyunghyun Cho et al. "On the properties of neural machine translation: Encoder-decoder approaches". In: *arXiv preprint arXiv:1409.1259* (2014).

[2] Angus Deaton and Guy Laroque. "On the behaviour of commodity prices". In: *The review of economic studies* 59.1 (1992), pp. 1–23.

[3] Trevor Hastie et al. *The elements of statistical learning: data mining, inference, and prediction.* Vol. 2. Springer, 2009.

[4] Ioannis E Livieris, Emmanuel Pintelas, and Panagiotis Pintelas. "A CNN–LSTM model for gold price time-series forecasting". In: *Neural computing and applications* 32.23 (2020), pp. 17351–17360.

[5] David A Nix and Andreas S Weigend. "Estimating the mean and variance of the target probability distribution". In: *Proceedings of 1994 ieee international conference on neural networks (ICNN'94).* Vol. 1. IEEE. 1994, pp. 55–60.

[6] Ser-Huang Poon and Clive WJ Granger. "Forecasting volatility in financial markets: A review". In: *Journal of economic literature* 41.2 (2003), pp. 478–539.