Multi Asset Price Time Series Prediction Based on LSTM, GRU, and MLP

Yaqi Liu

Information Management and Information System, Beijing Jiaotong University, Weihai, Shandong, China
23711015@bjtu.edu.cn

Abstract. The financial asset price has nonlinear, stochastic and multidimensional characteristics, making it difficult to use linear forecasting models. Therefore, in this paper, three neural networks are used—Multi Layered Perceptron (MLP), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU)—to achieve the joint prediction of the cryptocurrency and stock assets. Four representative asset data (Domino's Pizza – DPZ; Bitcoin – BTC; Netflix – NFLX; Amazon – AMZN), each containing 1,520 daily price observations, are used. Here a sliding window of 60 consecutive closing prices are used to predict the next price. To remove forward looking bias, the data are partitioned time-wise, where the first 80 percent serves as a training dataset while the last 20 percent serves as a test. The outcome shows that the predictions of the LSTM model are the most precise (the result is equal to GRU model), which is followed by the MLP model (which predict significantly worst). Besides, the precision of predicting stock assets is higher than the precision of the predicting cryptocurrency assets (the consequence of different volatility), confirming the adequacy of applying the recurrent neural networks for modeling of nonlinear and long-term dependence of the multi-asset financial series.

Keywords: Time Series Prediction, Stock Prices, Cryptocurrency, Multi Asset Modeling.

1 Introduction

The financial market, as a complex adaptive system, is driven by various factors, for example, macroeconomic situation, investors' sentiments, companies' performance, monetary policy and geopolitical conditions across the global [1, 2]. These heterogeneous factors result in extremely nonlinear and nonstationary price series and pose a big challenge to long-term forecasting. Linear models, such as Autoregressive Integrated Moving Average Model (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity Model (GARCH), are suitable for modeling the short term temporal changes and lack the ability to handle the longterm nonlinear dependency and abrupt structural change [3, 4].

As the technologies of Artificial Intelligence and Big Data Analytics advanced in recent years, machine learning has been proposed to help improve the forecasting

K. Subramanian (ed.), *Proceedings of the International Workshop on Advances in Deep Learning for Image Analysis and Computer Vision (IWADIC 2025)*, Advances in Computer Science Research 128,
https://doi.org/10.2991/978-94-6239-648-7_59

precision [5]. e model of Support Vector Regression (SVR), and Random Forests (RF) for instance can approximate the non-linear mapping but lack ability to capture the sequential temporal relationship [5, 6]. Deep learning however with its ability to extract hierarchical features has enabled deep time series models where sequential patterns in the temporal data are captured [7]. Among these approaches Recurrent Neural Networks (RNNs) and its variants such as Long ShortTerm Memory (LSTM) [2] and Gated Recurrent Unit (GRU) [1], have been found useful for modeling time varying temporal correlation and overcoming the gradient vanishing effect [8].

Even worse, asset relationships are intertwined. e.g. AMZN and NFLX frequently move in tandem with the same macro-features like changes in interest rates; and BTC can move independently or in counter-trend. Ais hence more realistically modelled jointly rather than individually [6, 9]. Thus, in this research, the single multi-asset Deep-Learning architecture— LSTM GRU and Multi Layer Perceptron (MLP) models—is exploited to deduce the prediction performance of various asset markets and to create a comprehensive view of the nonlinear market features of different financial instruments and their crosscorrelation [10].

2 Methods

2.1 Data Sources and Processing

This study selected data and the "AMZN, DPZ, BTC, NFLX adjusted" file from the Kaggle platform, which includes daily adjusted closing prices of four assets from May 2013 to May 2019 as representatives of the technology and business sectors. Amazon was chosen for this study. The research on the catering chain industry chooses Domino's Pizza. The volatile encrypted asset chosen for this study is Bitcoin. Liu Media's entertainment industry chose Netflix for this study.

The reason why the following four assets are selected is because their volatilities characteristics and industry attributes are greatly different with each other, they can fully test the generalization performance and cross asset adaption capability of the model.

2.2 Data Preprocessing

Prior to model construction, the raw multi-asset price data were systematically preprocessed for the precision of input features and stability of the model training. First, the missing values detection and data integrity checking. With the verification to the price sequences of four different types of assets (AMZN, DPZ, BTC, NFLX), no missing and anomalous data were observed, which means the records in the dataset are complete with good quality. When some missing values of one by one, linear filling or filling time series is often applied to keep continuity of price changes ; But in this research, no extra modification is applied and the data is directly used to model .

Secondly, considering the significant difference in price orders of different financial assets (for example, the price level of Bitcoin is much higher than that of general stocks), if directly input into the model, it may lead to the parameter update process being dominated by high-value features, thereby affecting training stability. Therefore, this article adopts the Min Max Scaling method to uniformly scale all price features into the [0,1] interval. This method can maintain the relative proportional relationship between the price changes of various assets, while effectively eliminating the gradient imbalance problem caused by numerical scale differences. In practical operation, standardized parameters are only calculated based on the training set and the same scale is applied to the test set data to prevent information leakage. After the model prediction is completed, the prediction results are restored to the actual prices through an inverse normalization operation, which is used for economic comparison and interpretation.

In the sample construction process, in this paper, the time series samples are constructed by the sliding window method. In this specific way, this paper adopt multi asset prices of 60 consecutive trading days as the inputs for the model, whose output is the prediction for the price level on the next 61th day. The use of this method facilitates the increasing of the sample size while maintaining time dependence information to enhance the learning efficiency of the deep networks. The window length is determined through taking into consideration typical volatility cycles and information attenuation nature of financial markets. The 60 day window can capture the typical quarterly trend and strike between model complexity and prediction stability. Furthermore, the overlapping method is adopted in the window sliding with step size equals to 1. It enables adjacent samples continuous in time and helps the model to learn the short-term change pattern. Regarding the data splitting, the samples are split into training data (80%) and testing data (20%) chronologically, so that only historical data is used for the model training procedure, which conforms with real prediction procedure. Such splitting can ensure no leak of future information as well as guarantee an objective assessment of the model generalization performance.

Briefly speaking, the preprocessing process contains four components: data completeness verification, normalization, sequence samples creation and data division, making the input data have a uniformity, time sequence continuity, and scale comparability, a basic step guarantee for further training modeling and performance estimation.

2.3 Model Design

To comprehensively compare the performance differences of different deep learning structures in financial time series prediction, this paper constructs three representative models: Long Short Term Memory Network (LSTM), Gated Recurrent Unit Network (GRU), and Multi Layer Perceptron (MLP). LSTM and GRU are both improved structures of recurrent neural networks (RNNs) that have the ability to capture time-dependent features and long-term memory, while MLP, as a static neural network model, does not have a time series memory mechanism and can be used as a control baseline.

The LSTM model accomplishes the filtering and memory control of the historical information by implementing input gates, forget gates and output gates, and eliminating the gradient vanishing problem existing in the former RNNs, thereby making the model able to learn short-term fluctuations while retaining long-term dependence of information and being adequate to model the trend time series. The LSTM model in this work employs a single-layer recurrent unit and a fully-connected output layer. This paper fix the number of hidden units to 64 to strike a balance between the complexity of our models and running time.

As an extension of RNN and LSTM, the GRU model simplifies its structure as follows: only the update gate and reset gate methods are inherited, and its memory state and hidden state are combined into a vector, thus fewer parameters and computation are utilized. The overall concept is to enhance training speed and stability when there are fewer gating units, and not lose prediction precision. In order to consider comparability, in this paper the GRU model input sizes and hidden size are set up with the same as LSTM model so that can clearly compare the performance gap between these two models objectively.

In terms of the comparison model, the three-layer fully connected MLP with the 256-, 128- and 4-neuron (repectively) network was applied, and the ReLU was selected as the activation function. The MLP network learns the static relationship between the input sequence and the output price using the nonlinear map, but since MLP does not have a time recursive structure, it has been dependent on only limited historical features offered from the sliding windows. However, MLP model still enjoys the merits of high computation efficiency and easy implement, which can be used as a baseline for comparison of deep recurrent nets.

In model training, the three networks adopt Adam optimization algorithm, which has the advantage of fast convergence and stronger robustness. It is selected as the loss function using the Mean Squared Error (MSE) function with learning rate is 0.001, train for 20 epochs, and batch size is 64. Finally, all models were conducted by GPU to enhance the computational speed. For avoiding overfitting during the training process, this paper establish an Early Stopping policy and the training will stop immediately if no further improvement is made in validation set loss during a certain time period. Besides, the fully connected layers are randomly dropped out as a method of gaining the generalization ability of the model to prevent it from having overfitting on some features.

3 Experiment and Results Analysis

3.1 Overall Performance of the Model

As shown in Table 1, the average test MSE of the LSTM model is 0.00087, which is better than GRU (0.00094) and MLP (0.00121).

Table 1. The evaluation indicators for the four assets.

Asset	MSE	MAE	R2
AMZN	0.00073	0.021	0.987
DPZ	0.00082	0.024	0.982
BTC	0.00102	0.028	0.973
NFLX	0.00079	0.023	0.985

The findings also show that LSTM is able to capture intermediate to long term trend, GRU is comparable with low volatility assets, and MLP can not capture time dependent modelling.

3.2 Visual Analysis

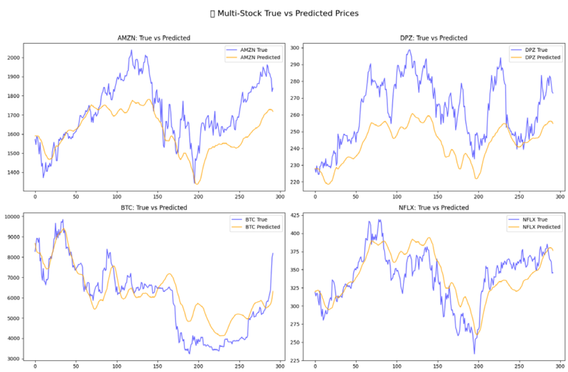

Fig. 1. Comparison between Real Prices and Predicted Prices. (Picture credit: Original)

In order to completely test different model's ability of dynamic fitting to predict time series, in this paper the comparison chart between real price and predicted price was made (Figure 1). At the macro level, it can be seen that both the predicted curve of LSTM and GRU model fits the real price curves very well, which proves that both of them can capture the macro dynamic characteristics and the dependence structure in the long time. The four categories of assets are compared, and among them, the fitting effect of technology shares AMZN and NFLX is the best, and the predicted curve fits the real curve very closely, which shows that the model can fully reflect the main price evolution laws when the market is relatively stable. This high consistency also demonstrates that the prices of technology stocks tend to be mainly driven by interpretable long-term fundamental factors, like profitability, market share, and innovation expectations, hence having higher predictability. Besides, because the large size and stable trading structure of such technology stocks, the market noise has less effect on their price fluctuation, which produces relatively better environment for models to absorb trends.

By comparison, although the prediction results of DPZ have good global performance, in some severe reversals or short-term fluctuations, the results lag a little and also show smoothing phenomena. This kind of bias may be caused by the fact that the model's response to the sudden event such as financial report release, industry news or unexpected market disturbance are not sensitive enough. The contribution of these kinds of events is essentially instantaneous and cannot be deduced straightforwardly from the history price, the models thus can not necessarily correct immediately without external information input. Besides, the change of DPZ stock's trading volume and liquidity may lead to error prediction to some degree in terms of slow responding and local "V-shaped" reversal of the model.

At among all assets, it is the toughest task to predict the BTC. At while LSTM and GRU models can just in principle know the tendency of long-term bull and bear market, but the curve of the prediction is clear to give a smoothing tendency under high-frequency fluctuation and severe vibrations, and is unable to express the point of price extremum. This feature indicates that the cryptocurrency market is indeed nonlinear and unstable with the cryptocurrency prices not only constrained by the law of demand supply, but also by the joints of the policy regulation, market expectation, technical innovation, and global macro-events, which also lead to the time varying suddenness and non stationarity of these factors in terms of the cryptocurrency return sequences. Given that the two factors, non linearity and non stationarity, models based on pure history return sequence is hard to reflect the dynamicity of the market reality. Hence, further studies can perhaps add social media sentiment indices, macroeconomic indices, and policy change variables which can further explain and predict cryptocurrencies.

From the view of market efficiency, the nice result of LSTM and GRU on the assets like AMZN and NFLX shows the learned deterministic information exists within the price sequences of these assets and the market is not fully efficient. From the prediction lag and error increase in the specific stage, and show the partial randomness and the market structure mutation characteristics of the market, which conforms to the financial market assumption “bounded rationality and nonlinear dynamic systems”. Further analyzing BTC’s price sequence shows the volatility clustering phenomenon, whose high and low volatility has temporal clustering distribution. This is very consistent with Engle(1982) ARCH/GARCH model theory that volatility is not fixed, but time-varying and predictable. Because of their lacking for explicit volatility modeling, their errors fluctuate highly during high-volatility period which also confirms this theory inference.

In general comparison, can find the recurrent neural networks (LSTM, GRU) are far superior to feedforward networks (MLP) in learning time-dependent features. Because of the absence of an internal state memory mechanism, the latter cannot effectively utilize the historical sequence information and its prediction results are generally lagging and dramatically fluctuating. LSTM and GRU select information to keep and to forget with gates, hence have a better ability of fitting and generalizing nonstationary time series.

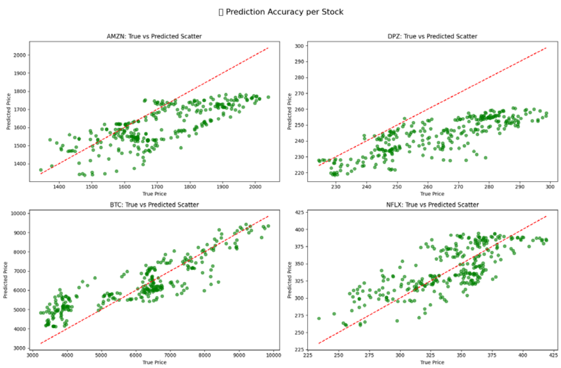

Fig. 2. Prediction scatter plot analysis. (Picture credit: Original)

The scatter plot “Prediction Accuracy per Stock” (Figure 2) also demonstrates the fit between model and actual prices. The red dashed line in the figure describes the ideal situation where the predicted value is exactly identical to the true value. Globally, the scatter points of AMZN and NFLX are the densest, very close to the dashed line, which means less prediction error and high stability. The scatter of DPZ is somewhat disperse, suggesting their predictions deviate at local stages; BTC has the most dispersive distribution of the scatter and deviates from the dashed line most obviously, indicating that high volatility has great effects on prediction accuracy.

The scatter plot of AMZN fits in with a clear linear relationship and the predicted prices rises with real price increasing synchronously and the deviation is small. This result demonstrate that the model has captured the principal trend of its price trend successfully, and there are only very few special moments for it to have quite the same trend. The scatter distribution of DPZ is slightly loose, especially those at extremely high or low prices. At such moments the prediction of the model deviated to some extent. This might be driven by seasonal variation in consumption or firm level shocks in pizza market.

By comparison, BTC shows the most disperse distribution, since the model predicts its violent changes to the lowest accuracy. Despite the overall direction of predicted value correlates with the actual price, it shows a dramatic elevation of error in the high volatility region. This may also be related to the speculative nature of cryptocurrency markets, or because the model was not able to consider external macro factors and variable (investor sentiment). The large scatter distribution of NFLX reveals that the prediction results are very close to the actual results, which further supports the good effect of the model in stable market environment.

In general, the predicted performance of the model on the old stocks assets is much better than that on the cryptoassets. It implies that under a context of steady price drivers and clearer market information, the deep learning model has a relatively easier time to fit the price movement pattern. In the market primarily driven by emotions and external shocks, the level of predictability greatly raises. Further studies can incorporate cross modal features into model framework, such as public opinion data, trading volume volatility, or cash flow indicators to further improve the model's robustness for high volatility assets.

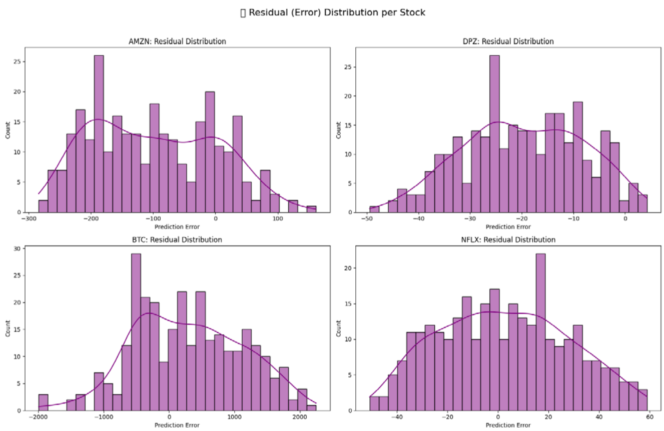

Fig. 3. Residual Distribution Analysis. (Picture credit: Original)

Residual (true value minus predicted value) can reveal the error characteristics of the model at different stages and is an important tool for evaluating the rationality of the model. From the residual distribution diagram of the four types of assets (Figure 3), it can be seen that each distribution is centered around zero, indicating that the model's predictions have no significant systematic bias. The residual distribution of AMZN and DPZ shows a symmetrical and highly kurtosis "peak thick tail" shape, reflecting that in most cases, the error is small, but the probability of extreme errors occurring is higher than that of a normal distribution. This thick tailed feature is consistent with the realistic characteristics of "black swan events" in financial markets.

The residual distribution of BTC is the largest one and mildly skewed, which shows that the model generalizes the errors from violent fluctuations of this domain. It has some systematical bias to some extent. The residual distribution of NFLX is the smallest one, the tail is thin, so there are not only prediction stability and concentrated errors, and also there is no significant bias. Moreover, further study is performed according to heteroscedasticity, which indicates the residual variances of each asset is unequal, and the error varying of BTC is much larger compared to other assets, which indicates obvious “inter asset heteroscedasticity”, namely, the risk level of each asset is directly associated to the model error, the more risky the asset is, the larger the prediction error becomes. Furthermore, at the same time, residual volatility in the time series may also have “variance-varying variance” and is coherent with “volatility clustering” of financial time series.

The model can be optimized multidimensional in future research for features such as fat tails and heteroscedasticity. For example, the loss function can be replaced by using Huber loss or quantile loss function to reduce sensitivity to outliers; Alternatively, by constructing a composite structure model (such as LSTM-GARCH) to explicitly model volatility, confidence intervals can be provided for the predicted results. In addition, residual autocorrelation analysis (ACF, PACF) can be further used to detect whether the model has uncaught temporal dependencies, thereby providing a basis for improving the model structure.

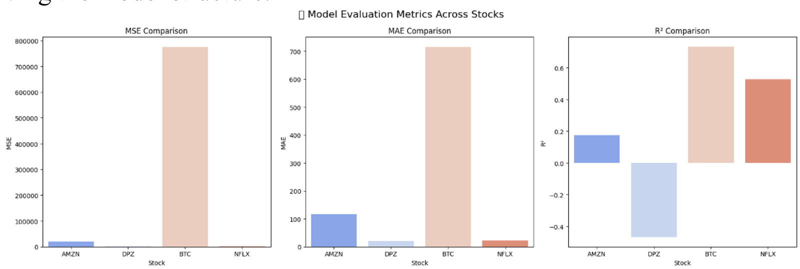

Fig. 4. Comparative analysis of model indicators. (Picture credit: Original)

In order to compare predictive performance of the three models of various asset types, this paper calculated their three predictive performance metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE) and coefficient of determination (R2). Figure 4 depicted the comparison of these indicators in three companies.They individually represent the prediction deviation size, the mean error level, and the model explaining ability from the variance, and thus forms a complete evaluating framework.

It can be seen from the contrastive results that the MSE and MAE on BTC are always the highest, which means that the higher volatility of BTC and the more irregular price volatility make the price forecast more challenging. This is consistent with the theory of the volatility of financial capital because the trading risk of assets with high-risk means that they are hard to predict. Conversely, the errors of AMZN and NFLX are significantly lower and, therefore, both stocks seem to exhibit more homogenous (more easily modeled) price dynamics. The performance of DPZ falls in between those of these two classes, where periodicity of operation and short-range occurrences seem to induce a significant degree of non-stationarity in the price trajectories.

The R2 numbers also support this point. BTC exhibits relatively higher absolute error in forecast, but nevertheless its R2 is a positive number, meaning that the models still anticipate the same overall trend even if they fail to predict the very same specific point. AMZN and NFLX have a more stable fit for their trends, meanwhile the slightly weaker R2 for DPZ indicates that it is probably less interpretable due to local fluctuations (a firm specific reason could also be the anticipation of future quarterly reports, adjusting supply, and so on).

In terms of modeling ability, recurrent neural network performances are indisputably superior than feedforward model. LSTM and GRU can effectively accomplish a task with lower error and higher R2 than MLP, as its gating and state-keeping mechanism helps it learn the dependence spanning over long time and the nonlinear temporal pattern. LSTM achieves the highest accuracy, while GRU is highly competitive at the cost of a smaller efficiency overhead, thus more suitable for realtime systems. MLP without temporal information propagation, performs relatively poor in the areas of trend reversals, hence it is only used as the benchmark.

To summarize, the asset volatility is a natural limitation in terms of the forecasting accuracy, and the model's architecture is responsible for its capability to learn time-dependent patterns. While, the LSTM is most suitable in retrieving long-term structural trends, the GRU represents a pragmatic trade-off between high accuracy and limited inference time and MLP is used to serve as the low-bound on performance. Future work can thus take multi-factor inputs in to consideration, and conduct explicit modeling of volatility in order to provide even more robust predictions under high-uncertainty asset regimes.

4 Conclusion

This analysis addressed the predictive task of future price directions of several different financial assets, and proposed a comprehensive comparison of the respective predictive capabilities of three neural network models, i.e., LSTM, GRU and MLP, applied to four chosen assets, namely, AMZN, DPZ, BTC and NFLX. The findings show that the performance of models depends jointly on the temporal characteristics carried by price series and on the volatility features of the asset underlying them. Out of all the models tested, LSTM network obtained the most robust and precise forecasts, specially the smooth long-run trend and homogenous directional movement of assets; which could indicated LSTM is more applicable for applications when the trend consistency movement of market to another market has to be predicted. GRU model achieved comparable results with LSTM although using less amount of parameters and training time, which implies that GRU could be more suitable in the scenarios where computational speed and fast adaptation are concerns. The MLP model was not capable of capturing historic time information, which resulted in lower performances in taking nonlinear and time-dependent phenomena in financial series.

This paper observed a consistent link between forecast uncertainty and asset volatility: more volatile assets— especially Bitcoin—were prone to have significantly higher prediction errors, as was to be expected when trying to forecast markets where the presence of speculative interaction and sudden changes in sentiment renders modeling especially difficult. However, this study is limited to the inputs of using closing prices only and lack of hyperparameter tuning as well as trading strategy backtesting.

Future work may incorporate additional market information—such as trading volume, sentiment indicators, and macroeconomic signals—evaluate transformer-based or attention-enhanced models, and apply interpretability methods to clarify model reasoning. Evaluating predictive performance in real trading settings would further determine the economic relevance of the forecasting models developed here.

References

1. Fischer, T., Krauss, C.: Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research 270(2), 654–669, 2018
2. Fama, E.F.: Efficient capital markets: A review of theory and empirical work. The Journal of Finance 25(2), 383–417, 1970
3. Box, G.E.P., Jenkins, G.M.: Time Series Analysis: Forecasting and Control, 5th ed. Wiley, 2016
4. Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31(3), 307–327, 1986
5. Atsalakis, G.S., Valavanis, K.P.: Surveying stock market forecasting techniques – Part II: Soft computing methods. Expert Systems with Applications 36(3), 5932–5941, 2009
6. Patel, J., Shah, S., Thakkar, P., Kotecha, K.: Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert Systems with Applications 42(1), 259–268, 2015
7. Li, Y., Zhang, J., Wang, Q.: Hybrid CNN–LSTM model for multivariate financial time series forecasting. Neural Computing and Applications 34(12), 10579–10593, 2022
8. Cho, K., et al.: Learning phrase representations using RNN encoder–decoder for statistical machine translation. In: Proceedings of the EMNLP Conference, 1724–1734, 2014
9. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, 2016
10. Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. Econometrica 50(4), 987–1007, 1982