## Related Work

The most renowned model for linear univariate time series forecasting is the autoregressive integrated moving average (ARIMA) [1], which encompasses other autoregressive time series models, including autoregression (AR), moving average (MA), and autoregressive moving average (ARMA). Additionally, linear support vector regression (SVR) [2] treats forecasting problem as typical regression problem with time-varying parameters. However, these models are rarely used in high dimensional multivariate time series forecasting due to their high computational cost.

To forecast MTS data, vector autoregression (VAR), which is a generalization of AR-based models, is proposed [3,4,5]. VAR models extend AR models to the multivariate setting but ignores the dependencies between output variables. Nevertheless, neither AR-based models nor VAR-based models are capable of capturing non-linearity. For that reason, substantial effort has

been made for non-linear models for time series forecasting based on kernel methods [], ensembles [] and Gaussian Processes (GP) [].

GP is a non-parametric method for modeling distributions over a continuous domain of functions. This contrasts with models defined by a parameterized class of functions such as VARs and SVRs. It can be applied to multivariate time series forecasting task as suggested in [28], However, the power of GP comes with the price of high computation complexity. A straightforward implementation of Gaussian Process for multivariate time-series forecasting has cubic complexity over the number of observations. What's more, GP and VAR apply predetermined non-linearity and may fail to recognize different forms of non-linearity for different MTS.

Recently, deep neural networks have received great amount of attention due to their adaptable abilities in capturing non-linear interdependencies. Two variants of RNN, namely long short-term memory (LSTM) [] and gated recurrent unit (GRU) [], have shown promising results in several NLP tasks and have also be employed on MTS forecasting. Previous work in this area starts from using naive RNN [], to hybrid models that combined ARIMA and Multilayer Perceptron [], and to the latest Dynamic Boltzmann Machine with RNN []. Although these models can be applied to MTS, they mainly target univariate or bivariate time series.