

CS229 Project Proposal

Study of Calibrating Time Series Models by Recurrent Neural Network

Jencir Lee jli14

Introduction

The Econometrics has established some standard time series models: Autoregressive Integrated Moving Average (ARIMA) model [Hamil1994], Regime-Switching model [Hamil1994], Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model [Engle1982] [Boll1986], to name a few, for modelling macroeconomy, financial markets, or general time series. The calibration of these models is generally tailored to the structure of each model, and for most of them, it usually involved conducting the calibration by stages; at each stage, a specific theory would establish the consistency and convergency guarantees of the estimator for the parameters being estimated.

A more unified approach is via Bayesian learning. We would impose priors on the parameters in the model and do the calibration usually by the iterative Monte Carlo Markov Chain algorithm. We would lose theoretical guarantees as it relied on the prior and the sampling procedure but this represents a step towards a unified approach.

Another route would be to use a Recurrent Neural Network and once we fix the hyperparameters, apply it to the calibration of the time series models. This is a black-box approach: we won't have the convergence guarantee in priori, we don't even specify a specific form of evolution equation for the time series sample.

In this study with limited scope, we propose to apply the Long Short-Term Memory (LSTM) model [Hoch1997] to the calibration of a few typical time series models. We'd pre-fix the parameters of the time series model, simulate trajectories from it, so that we would be able to understand the prediction performance in relation to the ground truth. We'd also study if tuning the hyper-parameters or adding penalty term would impact the prediction performance.

The second fold of this study's objective, would be to see if the internal units (Hidden Cell, Memory Cell) of the LSTM network actually captured statistical information about the latent states of the time series models, for complex ones that do involve latent variables. If the LSTM not only gives reasonable prediction performance but also could learn to certain extent the latent state of the time series models, that'd give a big boost to the confidence when we apply this approach to a new time series sample.

Outline of the Study Procedure

In summary we'd follow through these steps:

1. (Optional) Given an empirical time series, estimate the parameters for the ARIMA model, Regime-switching model, and GARCH model, using the standard or Bayesian approach.
2. We fix the parameters of these time series models, and simulate n trajectories from each model. We split these trajectories into training set and test set, fit LSTM on the training set, document the calibration performance on the test set, as well as in relation to the ground-truth values.
3. We take the simulated trajectories and fitted LSTM in Step 2, and try to measure if any combination of the internal units of the fitted LSTM is "closest" statistically to the latent states of the simulated time series.

In order to do it, we'd first minimise the Wasserstein Distance between two distributions P, Q , P for the distribution of a certain combination of the internal units of LSTM, Q for the distribution of the latent states of time series. The Wasserstein Distance $W(P, Q)$ is shown to be equivalent to [Vill2009]

$$W(P, Q) = \sup_f |E_{X \sim P}(f(X)) - E_{Y \sim Q}(f(Y))|,$$

where f is a Lipschitz function in $\mathbb{R} \rightarrow \mathbb{R}$.

The training would be adversarial: one neural network would attempt "combination" of the internal units of LSTM to produce the uni-variate distribution P and minimise the Wasserstein Distance, another network would try out f to maximise the same distance.

Once the above training is done, we'd have decided the "combination" of the internal units of the LSTM that is "closest" in distribution to the latent state of the simulated time series. We finally compute the Spearman rank-based correlation or the p-value of

the Kolmogorov-Smirnov test statistic between the distribution of "combination" of internal units, and that of the latent state of the time series, as an indication of how well the trained LSTM has learned about those latent states.

References

- [Hamil1994] James D. Hamilton, *Time Series Analysis*, 1994.
- [Engle1982] Robert F. Engle, *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*, *Econometrica*, 50 (4), 1982.
- [Boll1986] Tim Bollerslev, *Generalized Autoregressive Conditional Heteroscedasticity*, *Econometrics*, 31 (3), 1986.
- [Hoch1997] Sepp Hochreiter, Jürgen Schmidhuber, *Long short-term memory*, *Neural Computation*, 9 (8), 1997.
- [Vill2009] Cédric Villani, *Optimal Transport: Old and New*, *Grundlehren der mathematischen Wissenschaften*, 2009.