

# **Forecasting volatility on financial markets**

By : Lazar Kračunović ; Miloš Obrenović

# Uvod

**-U poslednjih nekoliko godina, razvoj novih procenitelja volatilnosti koji se oslanjaju na visoko-frekventne podatke doveo je do novih načina preciznije procene volatilnosti.**

**Ultimativni cilj ovog rada je da prezentuje i uporedi jednostavan autoregresivni model (HAR) sa modelom zasnovanim na neuronskim mrežama (LSTM). Želimo da analiziramo različite horizonte predviđanja i procenimo njihovu tačnost**

**Za periode predviđanja koristićemo dnevne, nedeljne i mesečne intervale. Ultimativni cilj istraživanja je da se ispita da li neuronske mreže mogu postići slične ili bolje rezultate u poređenju s tradicionalnim ekonometrijskim modelima.**



# Volatilnost

Volatilnost je inherentno neuočljiv koncept koji zahteva određene apstrakcije i metode za merenje i analizu.

Ona predstavlja drugi moment (disperziju) povraćaja, odnosno koliko se moguća zarada ili gubitak mogu razlikovati od proseka.

Važno je napomenuti da veća volatilnost obično nosi veći rizik, ali isto tako može dovesti do potencijalno većih prinosa ili gubitaka.

U kontekstu predviđanja, dok tvrdimo da je predviđanje cena kao pronalaženje igle u plastu sena, predviđanje volatilnosti bi se moglo uporediti s procenom veličine tog plasta sena.

Ovaj pristup naglašava složenost i važnost analize volatilnosti na finansijskim tržištima, kao i potrebu za efikasnim modelima i metodama kako bismo bolje razumeli i upravljali rizikom u investicionim odlukama.

# Heterogeni autoregresivni modeli

Heterogeni autoregresivni modeli se zasnivaju na hipotezi heterogenog tržišta u trgovanju. Ova hipoteza proizlazi iz činjenice da je volatilitnost pozitivno korelisana sa prisustvom na tržištu i obimom trgovanja.

Hipoteza heterogenog tržišta sugerise da učesnici na tržištu imaju različite strategije i vremenske horizonte kada reaguju na informacije.

Heterogeni autoregresivni modeli uključuju ove različite vremenske horizonte i strategije reagovanja u svoje specifikacije, što omogućava bolje modeliranje volatilitnosti na tržištima sa različitim tipovima učesnika i njihovim preferencama.

# HAR Model

Najobičniji model pretpostavlja da se volatilitnost može aproksimirati sa samo 3 komponente različitih vremenskih intervala: dnevna, nedeljna i mesečna.

Konačna jednačina koju ćemo koristiti je:

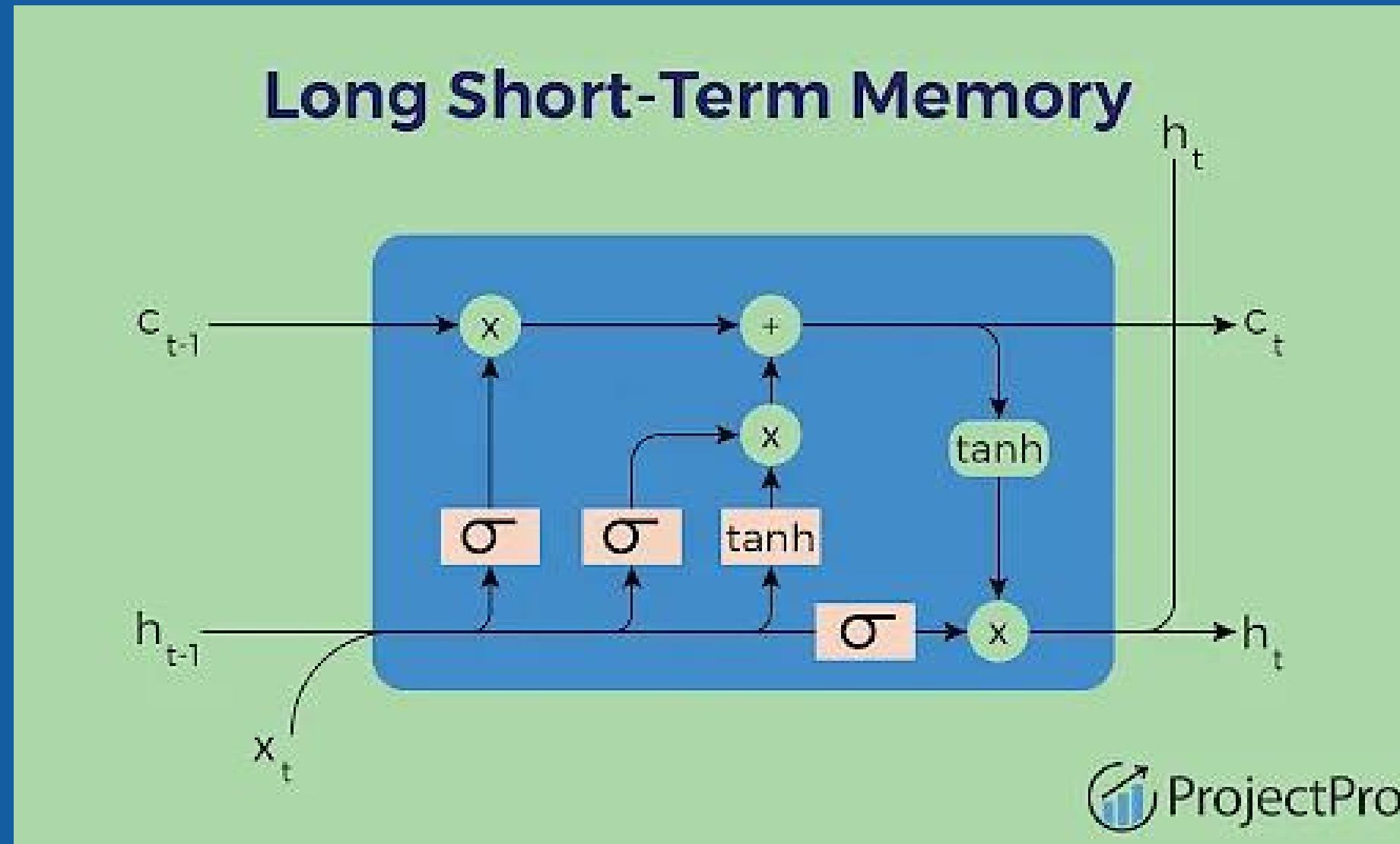
$$\hat{y}_{h,t+h} = y_{t+h} = \mu + \phi_d y_t + \phi_w \frac{1}{4} \sum_{i=1}^4 y_{t-i} + \phi_m \frac{1}{17} \sum_{i=5}^{21} y_{t-i} + \epsilon$$

dok zamenom semi-varijansama dobijamo:

$$\hat{y}_{h,t+h} = \frac{1}{h} \sum_{i=1}^h y_{t+i} = \mu + \phi_d^- y_t^- + \phi_d^+ y_t^+ + \phi_w \frac{1}{4} \sum_{i=1}^4 y_{t-i} + \phi_m \frac{1}{17} \sum_{i=5}^{21} y_{t-i} + \epsilon$$

# Long Short-Term Memory (LSTM) modeli

1. Forget Gate
2. Input Gate
3. Output Gate
- 4.: LSTM Output



# Podaci

01

S&P500 Trust Exchange Traded Fund (ETF), sa uzorcima u intervalima od 5 minuta.

02

Filtrirani tako da obuhvate samo vremenski period od 09:30 do 16:00

03

Uklanjanje ekstremnih vrednosti iz vremenskih serija. Reskaliranje i transformacija podataka.

04

Podela podataka na trening, validacioni i test skup.

Skup	Pocetak	Kraj	Duzina	Duzina (procenti)
Training	01.01.2006	12.31.2012	1,704	$\approx 60\%$
Validation	01.01.2013	12.31.2015	734	$\approx 20\%$
Testing	01.01.2016	12.31.2018	742	$\approx 20\%$



# KRITERIJUM EVALUACIJE

- standardne RMSE (Root Mean Squared Error)
- MAE (Mean Absolute Error)

# Rezultati

Dnevni forecast:

Nedeljni forecast:

	Train			Test		
	RMSE	MAE	R2	RMSE	MAE	R2
$HAR^{RV}$	0.9655	3.2178	0.7243	0.3135	1.7443	0.6334
$HAR_{log}^{RV}$	1.0133	3.1667	0.6963	0.3217	1.6141	0.6142
$HAR^{SV}$	0.9596	3.1411	0.7277	0.3084	1.7099	0.6453
$HAR_{log}^{RV}$	1.0175	3.1470	0.6938	0.3187	1.5989	0.6213
$LSTM_{20}^{RV}$	0.9771	3.2359	0.7176	0.3311	1.6509	0.5912
$LSTM_{30}^{RV}$	1.2983	5.1502	0.5015	0.4195	2.0558	0.3438
$LSTM_{20}^{SV}$	1.1456	3.4010	0.6119	0.3293	1.7497	0.5955
$LSTM_{30}^{SV}$	1.1163	3.7944	0.6315	0.3319	1.9228	0.5893

Model	Train			Test		
	RMSE	MAE	R2	RMSE	MAE	R2
$HAR^{RV}$	0.8599	3.0847	0.7330	0.2831	1.8292	0.6129
$HAR_{log}^{RV}$	0.9215	3.0687	0.6934	0.2814	1.5765	0.6176
$HAR^{SV}$	0.8537	3.0448	0.7368	0.2784	1.7945	0.6257
$HAR_{log}^{SV}$	0.9201	3.0620	0.6944	0.2802	1.5780	0.6208
$LSTM_{20}^{RV}$	1.1551	3.6549	0.5183	0.3351	1.8505	0.4577
$LSTM_{30}^{RV}$	0.8190	3.0413	0.7578	0.2923	1.9365	0.5873
$LSTM_{20}^{SV}$	0.8504	2.9026	0.7389	0.2860	1.6566	0.6050
$LSTM_{30}^{SV}$	0.8378	3.0829	0.7465	0.3224	1.7299	0.4982

-Kao što smo očekivali HAR modeli zasnovani na semi-varijansi su bolj od modela zasnovanih na realizovanom volatilitiju, dok to nije uvek tačno za LSTM modele.

-Kako RMSE kažnjava velika odstupanja, dok MAE sve greške posmatra isto, možemo zaključiti da LSTM{RV} obično daje bolja predviđanja nego LSTM{SV}.

-Takođe, možemo primetiti da modeli zasnovani na većem lagu daju bolje rezultate, odakle možemo pretpostaviti da postoji dodatna informacija koja je poželjna u daljoj istoriji.

-Pored stvari koje smo već diskutovali, možemo primetiti da u ovom slučaju modeli sa SV su bolji od RV modela kada posmatramo MAE u okviru LSTM modela.

# Mesečni forecast

Model	Train			Test		
	RMSE	MAE	R2	RMSE	MAE	R2
$HAR^{RV}$	0.9030	3.6703	0.6447	0.2983	2.3624	0.3722
$HAR_{log}^{RV}$	0.9840	3.4763	0.5780	0.2649	1.6223	0.5049
$HAR^{SV}$	0.8974	3.6554	0.6490	0.2963	2.3365	0.3806
$HAR_{log}^{SV}$	0.9836	3.4740	0.5784	0.2652	1.6246	0.5037
$LSTM_{20}^{RV}$	0.9352	3.3539	0.6188	0.2767	1.7476	0.4599
$LSTM_{30}^{RV}$	1.1763	3.7923	0.3970	0.2980	1.9844	0.3735
$LSTM_{20}^{SV}$	0.9283	3.2693	0.6245	0.2705	1.8752	0.4839
$LSTM_{30}^{SV}$	0.9541	3.2824	0.6033	0.2928	1.8847	0.3954

- Primećujemo da je LSTM značajno bolji u odnosu na HAR modele kada je u pitanju mesečno predviđanje.

# ZAKLJUČAK

Ovaj rad je imao za cilj da pokaže kako se neuronska mreža može koristiti umesto standardnih statističkih metoda. Iako su rezultati bili nešto lošiji u poređenju sa klasičnim metodama za jednostavne mreže i mali broj podataka, to je donekle bilo i očekivano. Kao što smo već napomenuli, jasno je da neuronske mreže pokazuju svoje prave sposobnosti tek kada se koriste sa velikim brojem podataka, kao i kada je distribucija podataka relativno normalna. Deformacija distribucije može dovesti do različitih problema koji otežavaju predviđanje.

Takođe, prilikom implementacije korišćena je min-max transformacija koja normalizuje podatke u odnosu na minimum i maksimum u skupu podataka koji moraju biti premereni. Korišćenje samo trening podataka za skaliranje može dovesti do problema, jer je nemoguće predvideti ekstremne vrednosti koje donosi budućnost.

Takođe, jasno je da su modeli poput HAR-a lako interpretirani, pa su prilagođeniji za razumevanje, dok neuronska mreža predstavlja "crnu kutiju" koja nije lako interpretirati, stoga je izazovno istražiti kako se model ponaša.