Analysis of Volatility Surfaces with Computer Vision

Alec Janowski, Matt Johnson, Annette Jing

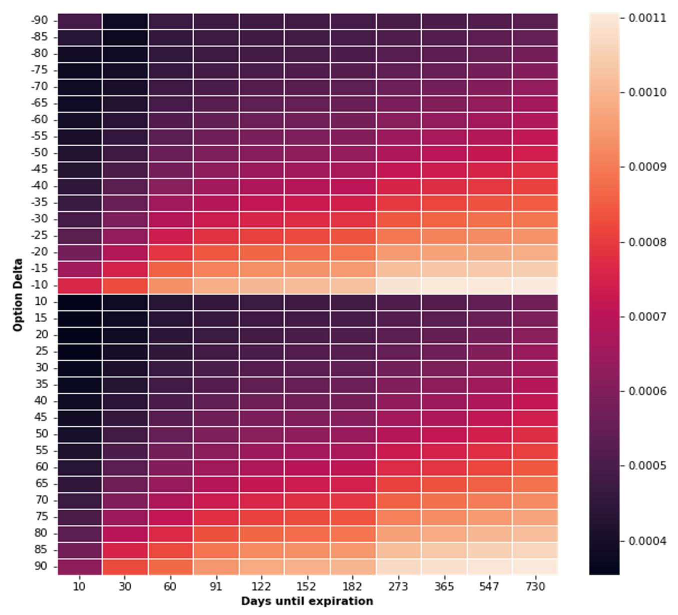
Introduction & Dataset

Introduction

- Volatility forecasting and model fitting important problem in finance.
- Volatility models used by firms to manage risk, comply with regulations, and make portfolio decisions among others.

Dataset

- Daily volatility surfaces pulled from OptionMetrics.
- Access to wide variety of surfaces, but chose S&P500, AAPL, and other US equities for simplicity.
- VIX historical data used as proxy for market volatility.
- Surfaces consist of 3 dimensions: *implied volatility level* as function of *time to expiry* and *strike price* quoted in grid format.



Methods

AR(1) Model (Lagged Linear Regression)

- Simple model that predicts the current target based on past targets.
- We fit the AR(1) model $v_t = \alpha + \beta v_{t-\ell} + \varepsilon_t$ using OLS, where v_t is the VIX index at time t.

GARCH(1, 1) Model

- Estimates the VIX index using S&P 500 returns r_t.
- $r_t = \sigma_t \epsilon_t$, where $\epsilon_t \sim std(v)$ iid, $\sigma_t^2 = \omega + \alpha\epsilon_{t-\ell} + \beta\sigma_{t-\ell}^2$.

2D CNN

- CONV(32)-MAXPOOL-CONV(64)-MAXPOOL-CONV(128)-MAXPOOL-FLATTEN-DENSE(256)-DENSE(128)-DENSE(64).

3D CNN

- Same architecture as 2D CNN, except filters are extended to 3 dimensions.
- Conducted saliency map analysis for each cross-section (delta-tenor, delta-lag, lag-tenor), where saliency for regression is interpreted as the absolute value of the loss gradient with respect to an input surface.

CONV-LSTM

- CONVLSTM(64)-CONVLSTM(128)-CONVLSTM(256)-DENSE(64)
- Nearly identical to LSTM layers in architecture, but convolutions are used for input-to-state and state-to-state transitions.
- Common approach to video frame predictions, which is ideal for our case because surfaces can be interpreted as videos over a number of days.

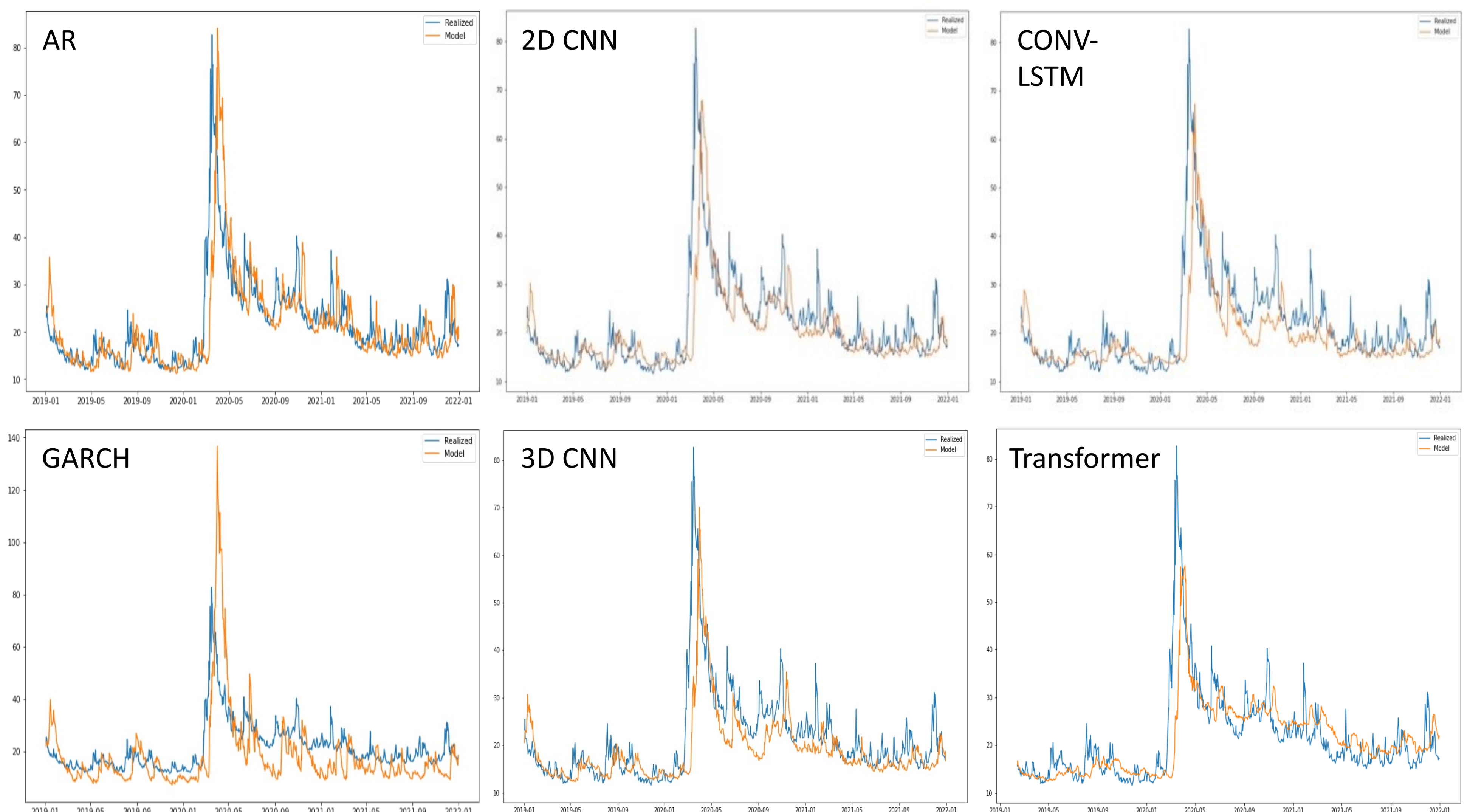
Transformer

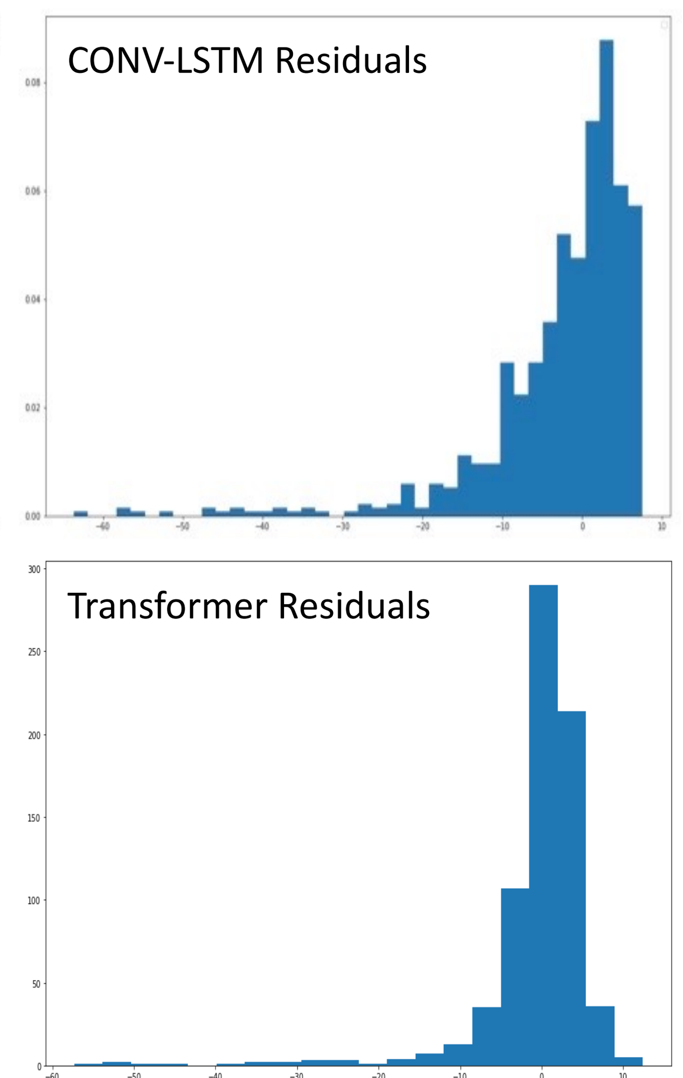
- Our implementation modifies the standard Transformer by removing last softmax layer, and by setting final linear layer to have dimension 1.
- Vol surfaces are first flattened, and then we add positional encoding. Input sequence is therefore 20 of these "embedded" vectors. Embedding dimension has size = number of tenors * number of deltas.
- Model is trained solely using the last temporal output of the transformer. 1-day and 5-day forecasts trained for 500 epochs, 10-day trained for 600 epochs using an Adam optimizer and a step scheduler which decreases the learning rate by 1% after each epoch.

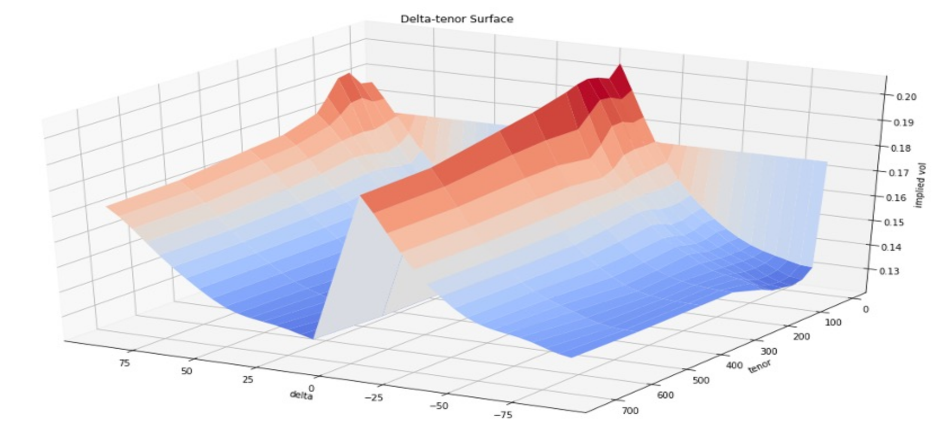
Results

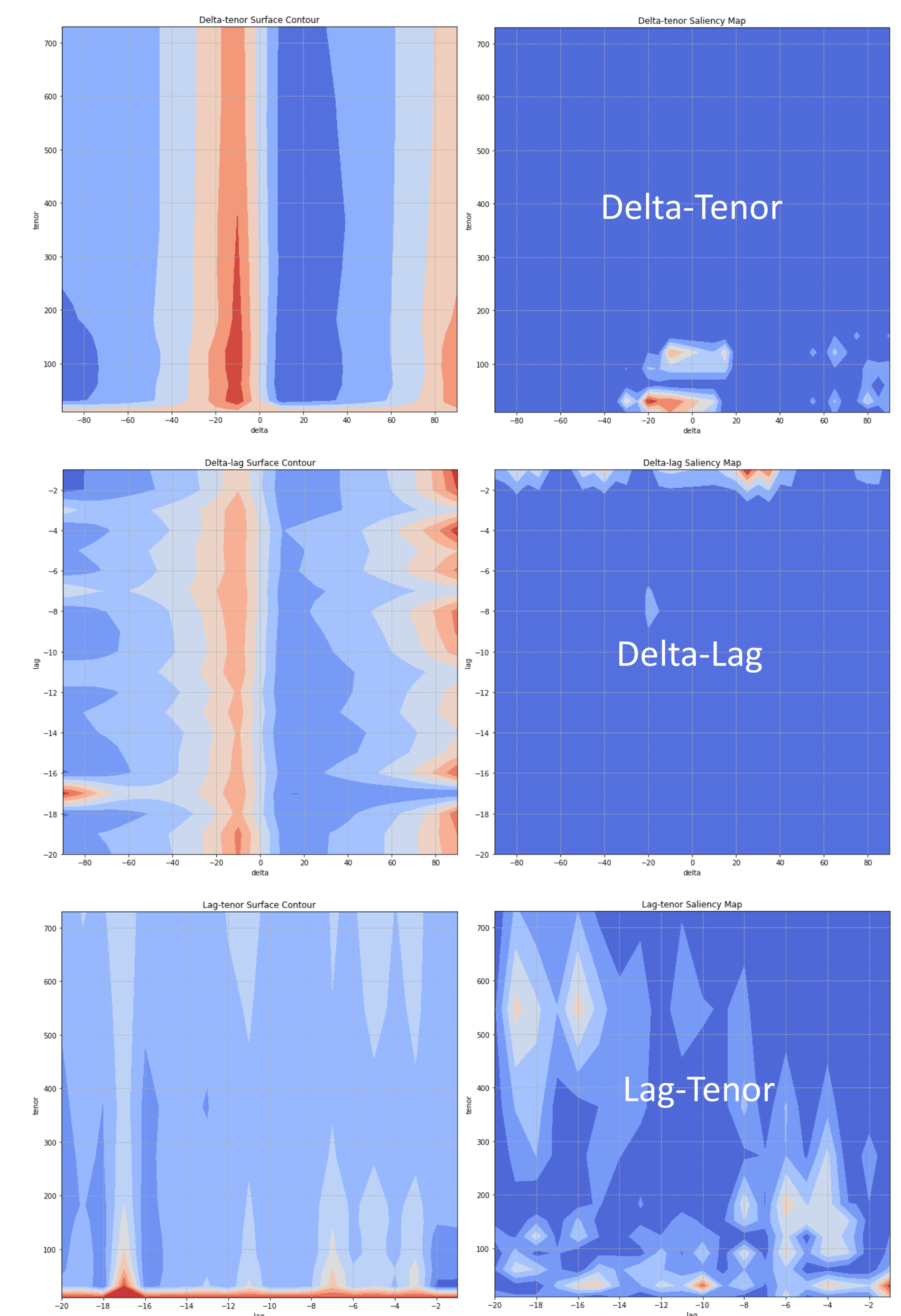
- Despite some models displaying marginally more predictive power than others, none exhibited normal residuals as desired
- Simple 2d CNN approach exhibited surprisingly high performance, low MAE values which points to potential trend capturing
- Post volatility spike in the test data (which is essentially impossible to predict due to Covid impact), 2d CNN and Transformer methods capture trends with reasonable performance
- 3d CNN underperforms even the 2d CNN, which likely means that training convergence is suboptimal and additional hyperparameter tuning is necessary for this method

	AR	GARCH	2D CNN	3D CNN	CONV-LSTM	Transformer
R2	0.90 0.69 0.43	0.08 -0.21 -0.58	0.82 0.62 0.50	0.77 0.67 0.43	0.89 0.63 0.44	0.84 0.67 0.50
MSE	9.07 27.44 50.04	80.64 105.83 138.57	9.81 27.48 45.30	21.38 29.75 51.51	10.10 33.80 50.70	14.67 30.64 46.78
MAE	1.83 3.08 4.05	6.25 6.81 7.53	1.86 2.87 3.60	2.75 3.04 4.10	8.51 8.31 7.82	2.70 3.56 3.67
AE STD	2.39 4.24 5.80	6.45 7.71 9.05	2.52 4.38 5.67	3.72 4.53 5.89	10.06 9.94 9.37	2.71 4.40 5.77









Conclusion & Future Work

Conclusion

- Simple regression approaches are still powerful for short horizons (e.g. 1-day).
- CONV-LSTM and Transformer methods hold better predictive power for longer horizons: 5, 10 days and potentially longer.
- Saliency analysis of CNN approaches demonstrate that models focus on short time-to-expiry options with strikes near ATM (at-the-money) levels.
  These are generally more reactive to changing market conditions with higher delta, so this is excellent to see.
- Transformer methods in particular exhibit a residual distribution closer to gaussian with lower kurtosis in the lower tail, meaning large spikes in volatility are better accounted for out-of-sample.

Future Work

- Would like to explore combining surfaces from multiple equities into one larger surface in the future.
- If possible, would like to obtain access to tick-level data rather than daily.
- Would like to explore different loss methods for training that encourage predictive outputs rather than only numerical accuracy (e.g. reward methods that predict large spikes more).

Primary References

[1] X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Proceedings of the 28th Conference on Neural Information Processing Systems, page 802–810, 2015.

[2] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. Advances in neural information processing systems, 30, 2017.

[3] S.D. Vrontos, J. Galakis, and I.D. Vrontos. Implied volatility directional forecasting: a machine learning approach. Quantitative Finance, 21(10):1687–1706, 2021.

[4] M. Yu. Predicting the volatility index returns using machine learning. PhD thesis, University of Toronto (Canada), 2017.

Additional references included in paper for brevity