Appendix

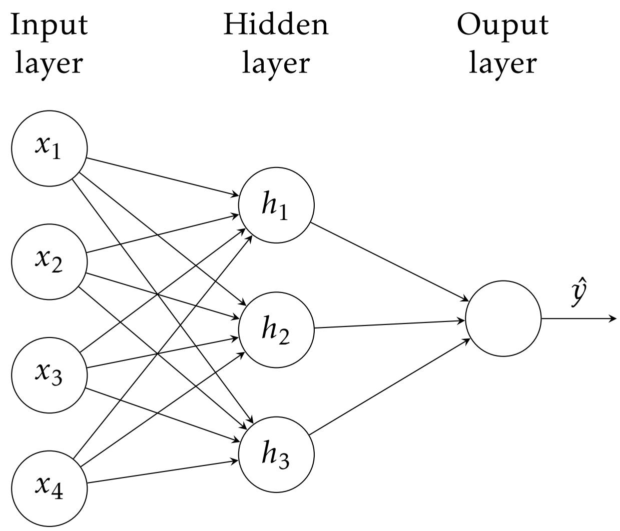

Figure 1: Example of a fully-connected feed-forward NN with a single output
This figure shows a fully-connected MLP with an input layer with 4 explanatory variables and no bias term, a single hidden layer with 3 nodes, and an output layer with 1 output.

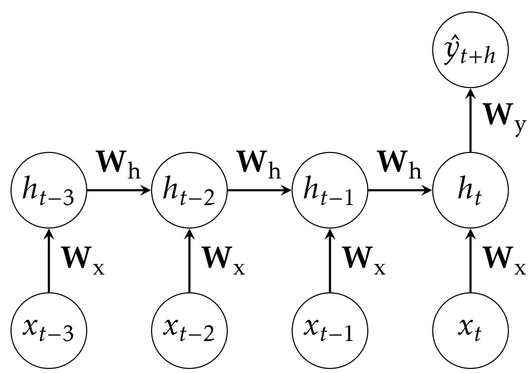

Figure 2: Example of an Elman (1990) RNN.
The figure shows an Elman (1990) RNN with an input sequence fixed to length four and a single output. Each hidden state variable $\mathbf{h}_j$ is a vector of length q, set by the number of hidden units in the layer. The input-to-hidden, hidden-to hidden, and hidden-to-output connections are parameterized by the weight matrices $\mathbf{W}_x$, $\mathbf{W}_h$, and $\mathbf{W}_y$ respectively.

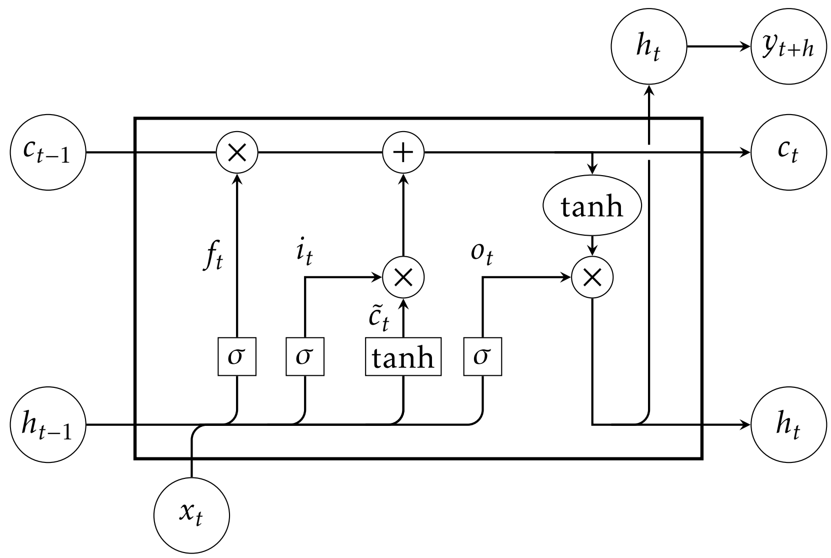

Figure 3: LSTM memory block
This figure illustrates the information flow through an LSTM cell. Outputs depend on the previous cell state, hidden state, and the input variable at the time-step, c_{t-1}, h_{t-1}, and x_t. Information flow is controlled via the input gate i_t, the forget gate f_t, and the output gate o_t. Operations in rectangles require estimated weight matrices. $\otimes$ and $\oplus$ are pointwise multiplications and additions respectively, σ is the sigmoid function.

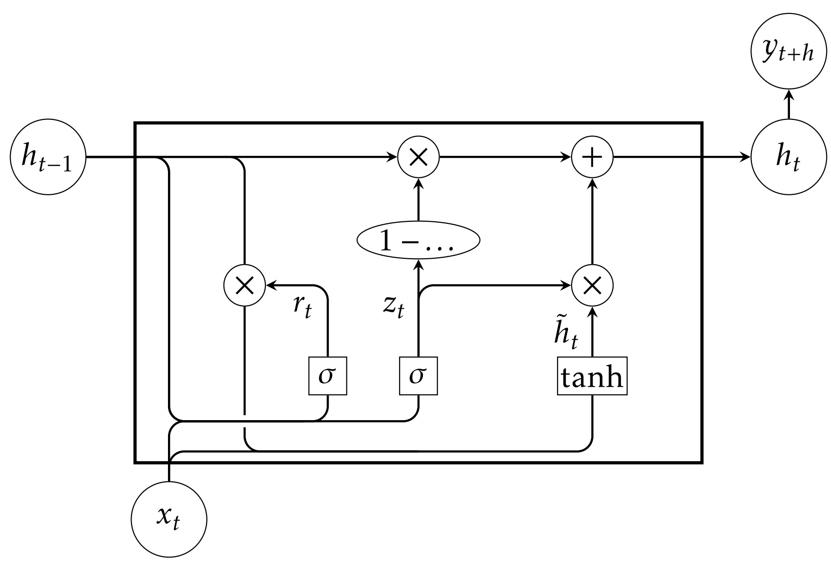

Figure 4: GRU memory block
This figure illustrates the information flow through a GRU memory block. It's output depends on the hidden state vector of the previous time step h_{t-1} and the current vector of input variables x_t. The reset gate r_t controls the degree to which h_{t-1} gets incorporated in the candidate update of the hidden state $\tilde{h}_t$. The update gate z_t controls the degree to which $\tilde{h}_t$ gets added to h_{t+1}. Operations in rectangles require estimated weight matrices. $\otimes$ and $\oplus$ are pointwise multiplications and additions respectively, σ is the sigmoid function.

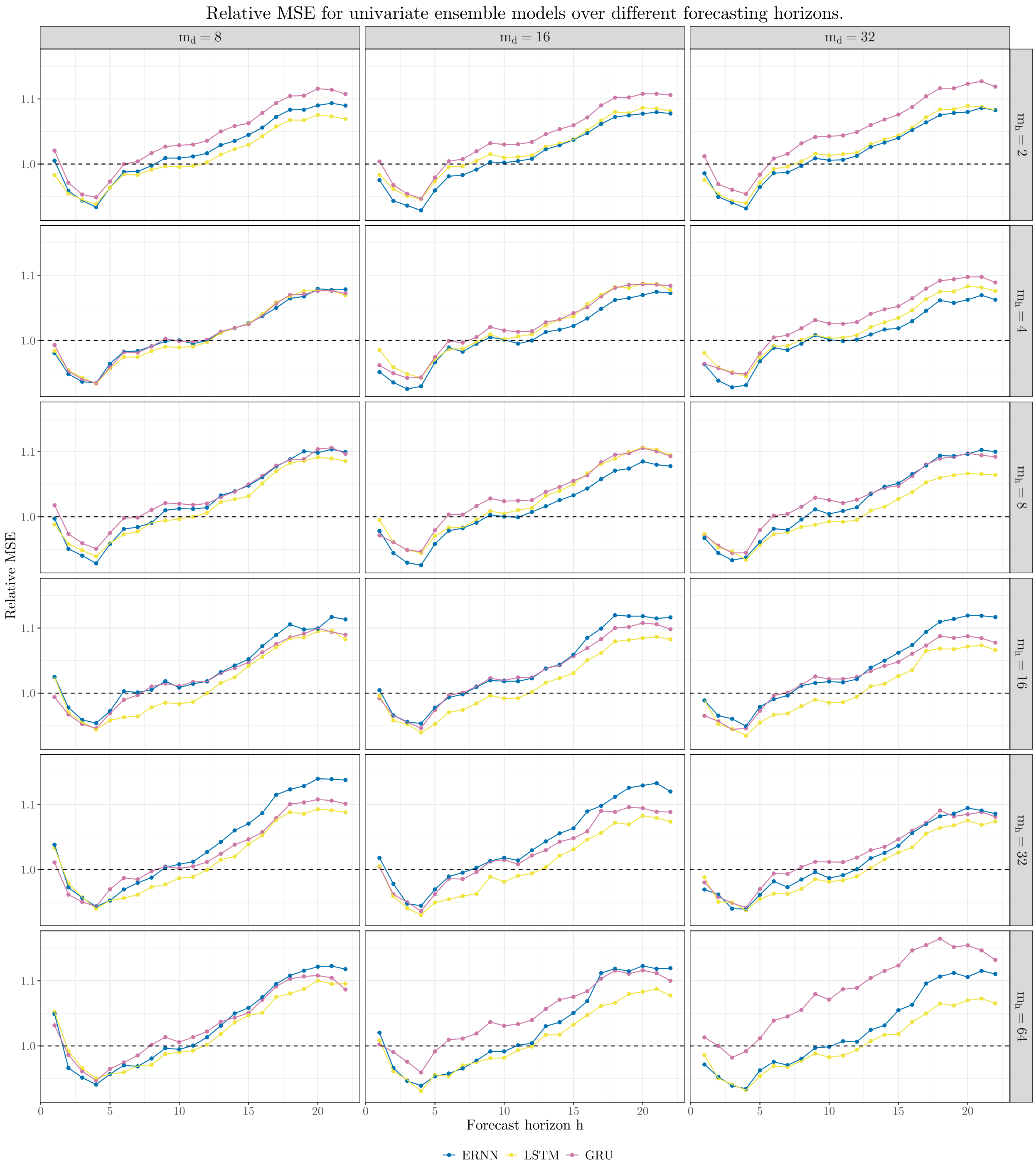

Figure 5: Relative out-of-sample MSE for univariate RV forecasts for an ensemble of 10 models for the different RNN model configurations over all forecast horizons.

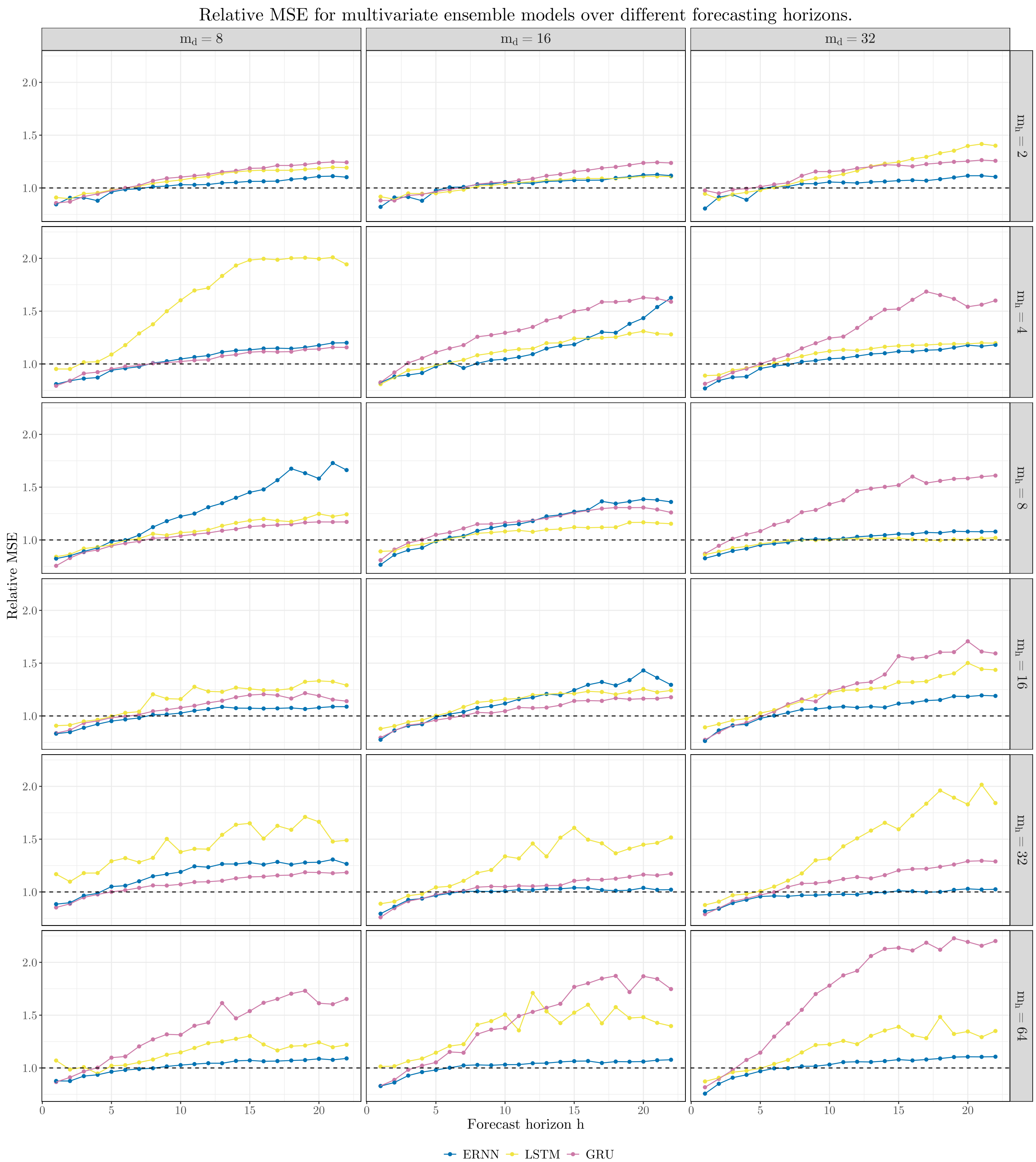

Figure 6: Relative out-of-sample MSE for multivariate RV forecasts for an ensemble of 10 models for the different RNN model configurations over all forecast horizons.

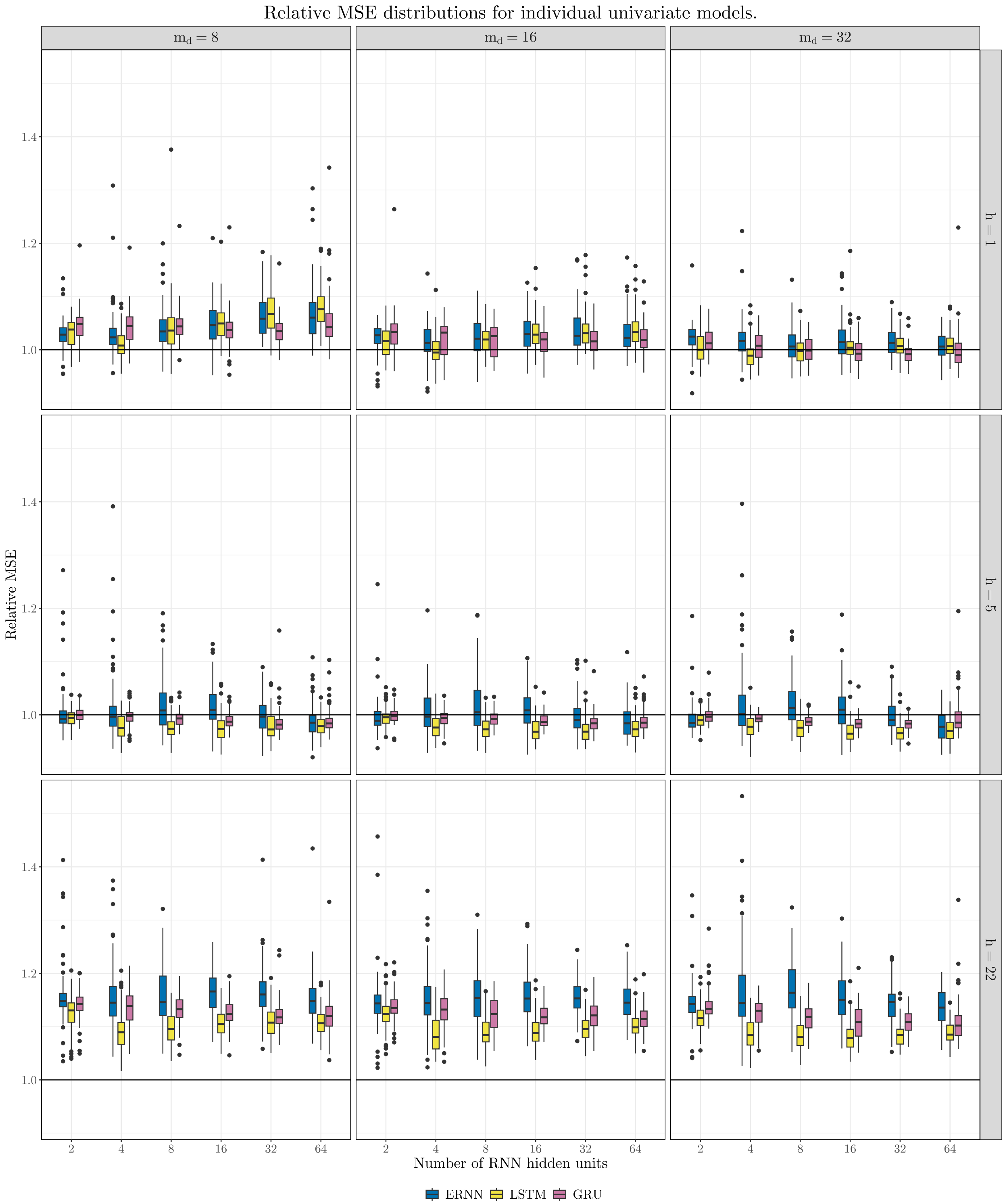

Figure 7: Relative out-of-sample MSE distributions for univariate RV forecasts across the different trained model configuration.

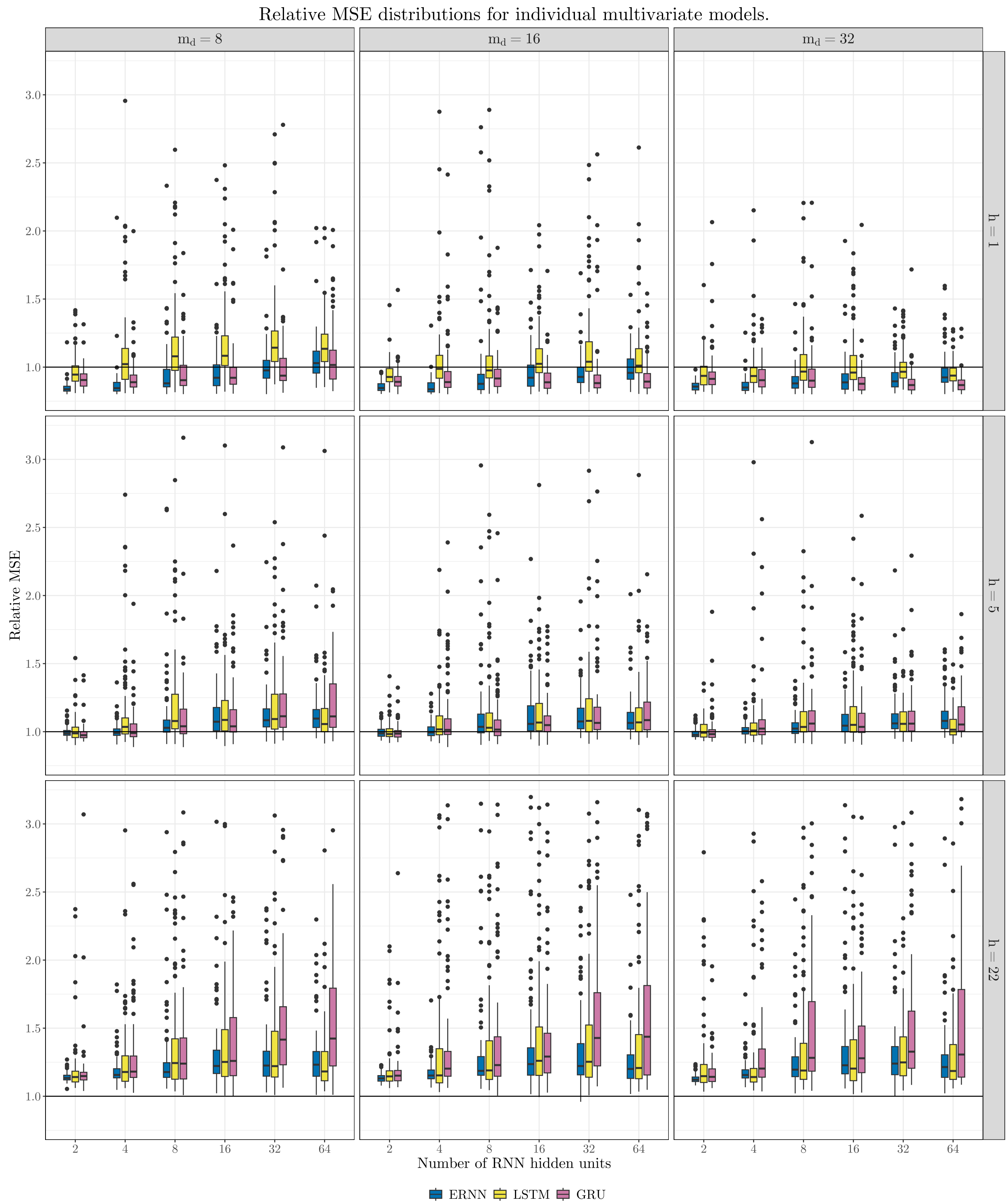

Figure 8: Relative out-of-sample MSE distributions for multivariate RV forecasts across the different trained models.

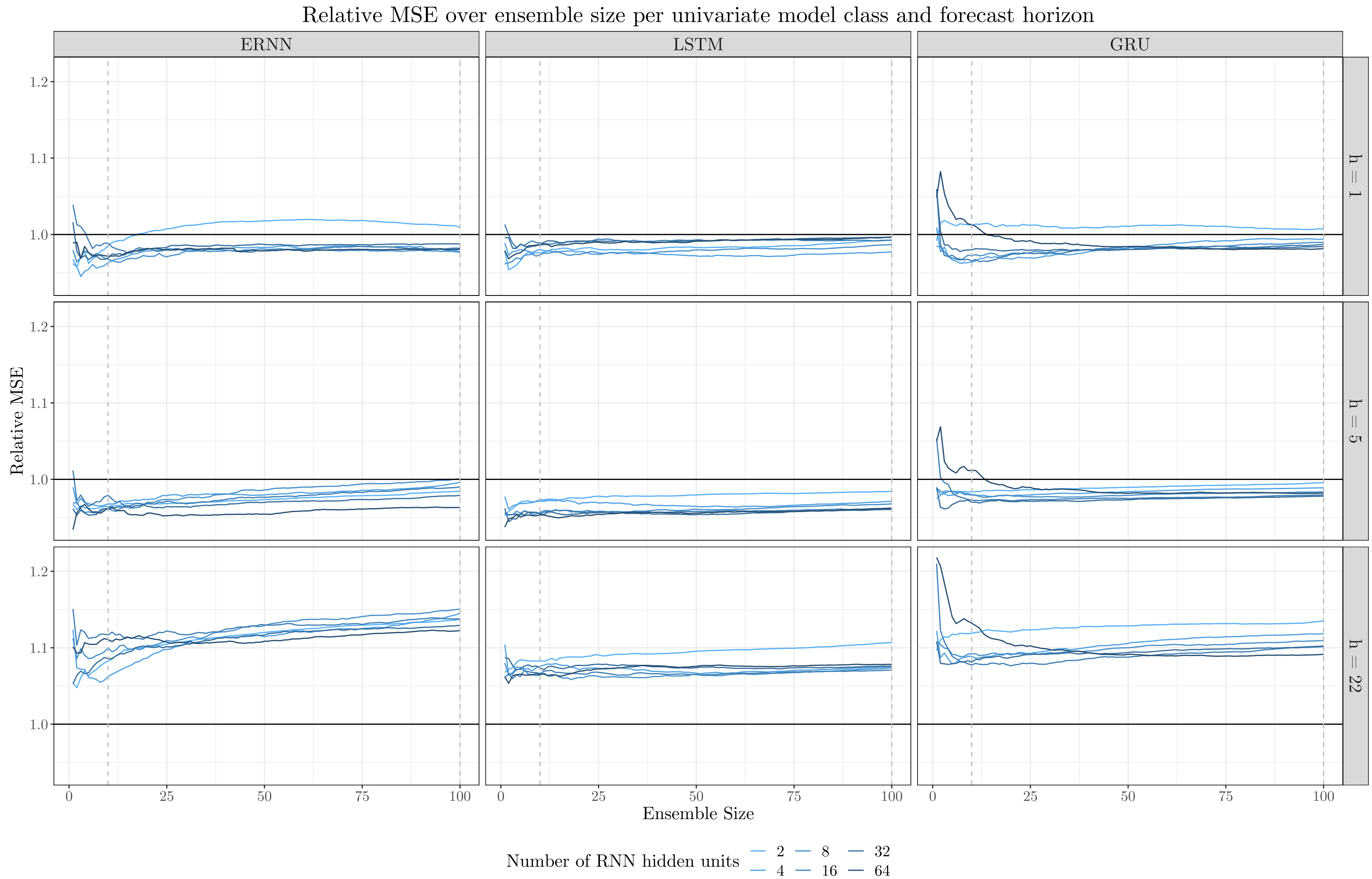

Figure 9: Relative out-of-sample MSE for univariate RV forecasts for an increasing ensemble size and different numbers of hidden RNN units. The number of hidden units in the feed-forward layer is fixed to $m_d = 32$. The dashed grey lines correspond to the ensemble sizes of 10 and 100, respectively.

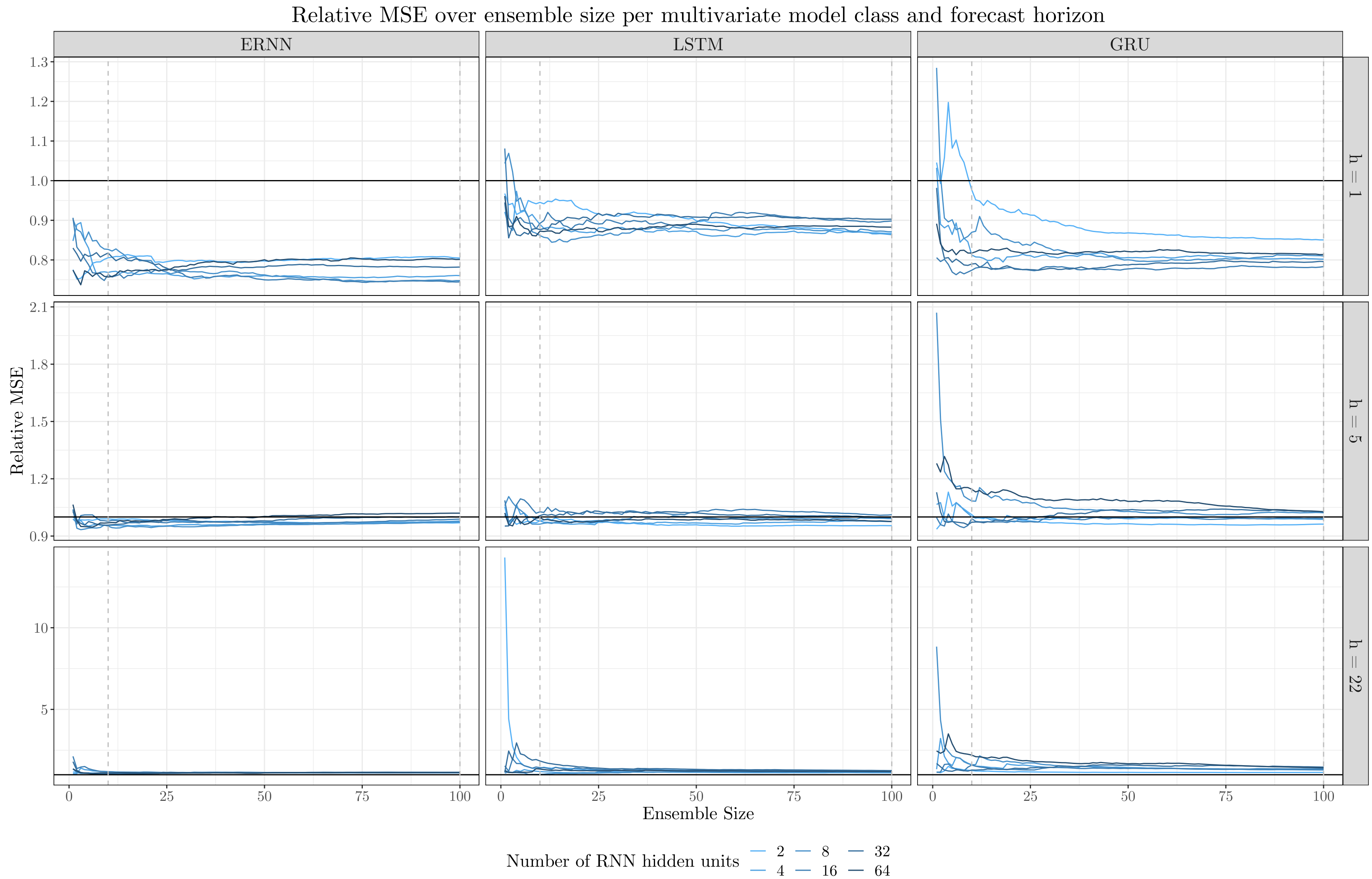

Figure 10: Relative out-of-sample MSE for multivariate RV forecasts for an increasing ensemble size and different numbers of hidden RNN units. The number of hidden units in the feed-forward layer is fixed to $m_d = 32$. The dashed grey lines correspond to the ensemble sizes of 10 and 100, respectively.

Table 1: Relative out-of-sample MSE of univariate sequential RNN ensembles of size 10.

	ERNN			*LSTM*			*GRU*		
m_d	$h=1$	$h=5$	$h=22$	$h=1$	$h=5$	$h=22$	$h=1$	$h=5$	$h=22$
$m_h = 2$									
8	1.005	0.964	1.090	0.983	0.964	1.069	1.021	0.973	1.108
16	0.975	0.959	1.078	0.984	0.973	1.081	1.004	0.979	1.106
32	0.986	0.964	1.083	0.976	0.972	1.083	1.012	0.984	1.119
$m_h = 4$									
8	0.981	0.964	1.078	0.984	0.957	1.069	0.993	0.959	1.072
16	0.951*	0.966	1.073	0.985	0.971	1.077	0.962	0.974	1.084
32	0.963	0.968	1.062	0.981	0.973	1.076	0.964	0.980	1.089
$m_h = 8$									
8	0.997	0.958	1.100	0.988	0.959	1.085	1.018	0.975	1.097
16	0.978	0.958	1.078	0.995	0.971	1.095	0.971	0.979	1.093
32	0.967	0.961	1.100	0.974	0.957	1.064	0.972	0.979	1.092
$m_h = 16$									
8	1.025	0.972	1.113	1.024	0.958	1.083	0.994	0.969	1.090
16	1.005	0.978	1.116	0.996	0.952	1.083	0.992	0.974	1.098
32	0.989	0.979	1.117	0.989	0.955	1.066	0.965	0.973	1.078
$m_h = 32$									
8	1.038	0.952	1.137	1.033	0.952	1.088	1.011	0.970	1.101
16	1.018	0.970	1.120	1.005	0.949	1.074	1.004	0.962	1.088
32	0.969	0.961	1.086	0.988	0.955	1.074	0.980	0.970	1.081
$m_h = 64$									
8	1.049	0.957	1.118	1.052	0.957	1.095	1.032	0.965	1.087
16	1.020	0.953	1.119	1.008	0.956	1.077	1.002	0.992	1.100
32	0.971	0.962	1.111	0.986	0.954	1.065	1.013	1.012	1.132

Note:
This table presents the out-of-sample MSE of different univariate model specifications for the RV forecasts relative to the HAR benchmark for different horizons h. The individual RNN forecasts of RV are based on the average prediction across 10 estimated models each, with the selected models producing the lowest in-sample validation error out of 100 candidate models. m_h and m_d denote the number of hidden units in the RNN cell and the feed-forward layer, respectively. *, **, and *** denote if the Diebold-Mariano test's null hypothesis of equal forecast accuracy is rejected at the 10%, 5%, and 1% level, respectively.

Table 2: Relative out-of-sample MSE of multivariate sequential RNN ensembles of size 10.

	ERNN			*LSTM*			*GRU*		
m_d	$h = 1$	$h = 5$	$h = 22$	$h = 1$	$h = 5$	$h = 22$	$h = 1$	$h = 5$	$h = 22$
$m_h = 2$									
8	0.845*	0.962	1.102	0.910*	0.982	1.192	0.858*	0.975	1.242
16	0.821*	0.977	1.117	0.918*	0.949*	1.109	0.881*	0.967	1.237
32	0.805*	0.990	1.106	0.945	0.979	1.400	0.977	1.012	1.257
$m_h = 4$									
8	0.809*	0.941	1.201	0.952	1.089	1.943	0.793*	0.952	1.157
16	0.824*	0.977	1.626	0.809*	0.986	1.280	0.826*	1.111	1.589
32	0.769*	0.957*	1.181	0.890*	0.979	1.197	0.812*	1.001	1.600
$m_h = 8$									
8	0.823*	0.986	1.662	0.841*	0.953	1.243	0.755*	0.945	1.171
16	0.765*	0.988	1.360	0.892*	0.985	1.154	0.809*	1.050	1.260
32	0.827*	0.952*	1.079	0.861*	0.962	1.020	0.870	1.084	1.610
$m_h = 16$									
8	0.832*	0.951	1.088	0.908	0.996	1.291	0.837*	0.984	1.140
16	0.775*	0.988	1.295	0.879*	1.003	1.242	0.794*	0.961	1.177
32	0.763*	0.978	1.190	0.892*	1.027	1.436	0.774*	0.991	1.592
$m_h = 32$									
8	0.885*	1.051	1.266	1.169	1.291	1.489	0.853*	1.001	1.184
16	0.794*	0.967	1.021	0.889	1.044	1.516	0.760*	0.974	1.172
32	0.818*	0.957	1.025	0.876*	1.007	1.842	0.790*	0.972	1.289
$m_h = 64$									
8	0.877*	0.964	1.090	1.070	1.021	1.220	0.867*	1.097	1.653
16	0.828*	0.981	1.078	1.014	1.145	1.397	0.830*	1.052	1.747
32	0.757*	0.969	1.107	0.873*	0.996	1.350	0.817*	1.145	2.202

Note:
This table presents the out-of-sample MSE of different multivariate model specifications for the RV forecasts relative to the HAR benchmark for different horizons h. The individual RNN forecasts of RV are based on the average prediction across 10 estimated models each, with the selected models producing the lowest in-sample validation error out of 100 candidate models. m_h and m_d denote the number of hidden units in the RNN cell and the feed-forward layer, respectively. *, **, and *** denote if the Diebold-Mariano test's null hypothesis of equal forecast accuracy is rejected at the 10%, 5%, and 1% level, respectively.

References

Andersen, Torben G., Tim Bollerslev and Francis X. Diebold (2007), 'Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility', *Review of Economics and Statistics* **89**(4), 701–720.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold and Heiko Ebens (2001), 'The distribution of realized stock return volatility', *Journal of Financial Economics* **61**(1), 43–76.

Andersen, Torben G, Tim Bollerslev, Francis X Diebold and Paul Labys (2001), 'The Distribution of Realized Exchange Rate Volatility', *Journal of the American Statistical Association* **96**(453), 42–55.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold and Paul Labys (2003), 'Modeling and Forecasting Realized Volatility', *Econometrica* **71**(2), 579–625.

Barndorff-Nielsen, Ole E. and Neil Shephard (2002), 'Econometric analysis of realized volatility and its use in estimating stochastic volatility models', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(2), 253–280.

Barndorff-Nielsen, Ole E, P Reinhard Hansen, Asger Lunde and Neil Shephard (2009), 'Realized kernels in practice: Trades and quotes'.

Bekaert, Geert and Marie Hoerova (2014), 'The VIX, the variance premium and stock market volatility', *Journal of Econometrics* **183**(2), 181–192.

Bollerslev, Tim (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**(3), 307–327.

Bollerslev, Tim, George Tauchen and Hao Zhou (2009), 'Expected Stock Returns and Variance Risk Premia', *The Review of Financial Studies* **22**(11), 4463–4492.

Bucci, Andrea (2020*a*), 'Cholesky-ANN models for predicting multivariate realized volatility', *Journal of Forecasting* **39**(6), 865–876.

Bucci, Andrea (2020*b*), 'Realized Volatility Forecasting with Neural Networks', *Journal of Financial Econometrics* **18**(3), 502–531.

Chakraborty, Shayak, Jayanta Banik, Shubham Addhya and Debraj Chatterjee (2020), 'Study of Dependency on number of LSTM units for Character based Text Generation models', *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)* **00**, 1–5.

Cho, Kyunghyun, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk and Yoshua Bengio (2014), 'Learning phrase representations using rnn encoder-decoder for statistical machine translation', *arXiv preprint arXiv:1406.1078* .

Christensen, Kim, Mathias Siggaard and Bezirgen Veliyev (2022), 'A Machine Learning Approach to Volatility Forecasting', *Journal of Financial Econometrics* .

Chung, Junyoung, Caglar Gulcehre, KyungHyun Cho and Yoshua Bengio (2014), 'Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling', *arXiv* .

Corsi, F (2009), 'A Simple Approximate Long-Memory Model of Realized Volatility', *Journal of Financial Econometrics* **7**(2), 174–196.

Di Persio, Luca and Oleksandr Honchar (2017), 'Recurrent neural networks approach to the financial forecast of google assets', *International journal of Mathematics and Computers in simulation* **11**, 7–13.

Diebold, F. X. and R.S. Mariano (1995), ‘Comparing predictive accuracy’, *Journal of Business & Economic Statistics* **13**(3), 253–263.

Diebold, Francis X (2015), ‘Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests’, *Journal of Business & Economic Statistics* **33**(1), 1–1.

Dixon, Matthew F, Igor Halperin and Paul Bilokon (2020), *Machine learning in finance*, Vol. 1170, Springer.

Elman, Jeffrey L. (1990), ‘Finding Structure in Time’, *Cognitive Science* **14**(2), 179–211.

Engle, Robert F (1982), ‘Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation’, *Econometrica* **50**(4), 987.

Fischer, Thomas and Christopher Krauss (2018), ‘Deep learning with long short-term memory networks for financial market predictions’, *European Journal of Operational Research* **270**(2), 654–669.

Fičura, Milan (2017), ‘Forecasting Stock Market Realized Variance with Echo State Neural Networks’, *European Financial and Accounting Journal* **12**(3), 145–156.

Ge, Wenbo, Pooia Lalbakhsh, Leigh Isai, Artem Lenskiy and Hanna Suominen (2022), ‘Neural Network-Based Financial Volatility Forecasting: A Systematic Review’, *ACM Computing Surveys* **55**(1), 1–30.

Goodfellow, Ian, Yoshua Bengio and Aaron Courville (2016), *Deep Learning*, MIT Press.

Granger, Clive WJ and Paul Newbold (1976), ‘Forecasting transformed series’, *Journal of the Royal Statistical Society: Series B (Methodological)* **38**(2), 189–203.

Gu, Shihao, Bryan Kelly and Dacheng Xiu (2020), ‘Empirical Asset Pricing via Machine Learning’, *The Review of Financial Studies* **33**(5), 2223–2273.

Heber, Gerd, Asger Lunde, Neil Shephard and Kevin Sheppard (2009), ‘Oxford-man institute’s realized library’, *Version 0.3, Oxford&Man Institute, University of Oxford* .

Hochreiter, Sepp and Jürgen Schmidhuber (1997), ‘Long short-term memory’, *Neural computation* **9**(8), 1735–1780.

Jung, Gunho and Sun-Yong Choi (2021), ‘Forecasting Foreign Exchange Volatility Using Deep Learning Autoencoder-LSTM Techniques’, *Complexity* **2021**, e6647534.

Kim, Ha Young and Chang Hyun Won (2018), ‘Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models’, *Expert Systems with Applications* **103**, 25–37.

Kingma, Diederik P and Jimmy Ba (2014), ‘Adam: A method for stochastic optimization’, *arXiv preprint arXiv:1412.6980* .

Maas, Andrew L., Awni Y. Hannun and Andrew Y. Ng (2013), Rectifier Nonlinearities Improve Neural Network Acoustic Models, *in* ‘Proc. icml’, Vol. 30.

McAleer, Michael and Marcelo C Medeiros (2008), ‘Realized volatility: A review’, *Econometric reviews* **27**(1-3), 10–45.

Patton, Andrew J. and Kevin Sheppard (2015), ‘Good Volatility, Bad Volatility: Signed Jumps and The Persistence of Volatility’, *Review of Economics and Statistics* **97**(3), 683–697.

Paye, Bradley S. (2012), ‘’Déjà vol’: Predictive regressions for aggregate stock market volatility using macroeconomic variables’, *Journal of Financial Economics* **106**(3), 527–546.

Pong, Shiuyan, Mark B Shackleton, Stephen J Taylor and Xinzhong Xu (2004), 'Forecasting currency volatility: A comparison of implied volatilities and ar (fi) ma models', *Journal of Banking & Finance* **28**(10), 2541–2563.

Rahimikia, Eghbal and Ser-Huang Poon (2020), 'Machine learning for realised volatility forecasting', *Available at SSRN* **3707796**.

Shen, Guizhu, Qingping Tan, Haoyu Zhang, Ping Zeng and Jianjun Xu (2018), 'Deep Learning with Gated Recurrent Unit Networks for Financial Sequence Predictions', *Procedia Computer Science* **131**, 895–903.

Shi, Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong and Wang-chun Woo (2015), 'Convolutional lstm network: A machine learning approach for precipitation nowcasting', *Advances in neural information processing systems* **28**.

Wilms, Ines, Jeroen Rombouts and Christophe Croux (2021), 'Multivariate volatility forecasts for stock market indices', *International Journal of Forecasting* **37**(2), 484–499.

Xiong, Ruoxuan, Eric P Nichols and Yuan Shen (2015), 'Deep Learning Stock Volatility with Google Domestic Trends', *arXiv* .

Zhang, Xuan, Xun Liang, Aakas Zhiyuli, Shusen Zhang, Rui Xu and Bo Wu (2019), At-lstm: An attention-based lstm model for financial time series prediction, *in* 'IOP Conference Series: Materials Science and Engineering', Vol. 569, IOP Publishing, p. 052037.