us to determine if the ARR supplies any information above that provide the volatility time series alone.

For tests without the ARR, we continue to use Equation (18) but omit ARR values, i.e. $\omega(t) = \emptyset$. We evaluate regression performance using the R^2 of forecasted log RV.

Crash Prediction Methodology

We take a binary classification approach to forecasting market crashes, using our benchmark models to predict the onset of a sharp drawdown. First, we define a z-score metric for returns $\gamma(t, \lambda)$ as below:

$$\gamma(t, \lambda) = \frac{r(t) - m(t, \lambda)}{s(t, \lambda)}, \tag{19}$$

where $r(t)$ is the return for the total market index, $m(t, \lambda)$ is the exponentially weighted moving average for returns with a half-life of λ, and $s(t, \lambda)$ its exponentially weighted moving standard deviation.

Next, based on our z-score metric, we define a market crash to be a sharp decline in market returns, i.e.:

$$c(t) = \mathbb{I}(\gamma(t, \lambda) < C), \tag{20}$$

where $c(t) \in 1, 0$ is a crash indicator, $\mathbb{I}(.)$ is an indicator function, and C is the z-score threshold for returns. For our experiments, we set λ to be 10 discrete time steps and $C = -1.5$.

We then model crash probabilities using a similar form to Equation (18):

$$p(c(t) = 1) = g_2(\psi(t), \omega(t)), \tag{21}$$

where $g_2(.)$ is a function mapping inputs to crash probabilities $p(c(t) = 1)$.

Given that crashes are rare by definition – with $c(t) = 1$ for less than 10% of time steps for daily frequencies – we also oversample the minority class to address the class imbalance problem. Classification performance is evaluated using the area under the receiver operating characteristic (AUROC).

Results and Discussion

The results for both volatility forecasting and market crash prediction can be found in Exhibit 7, both including and excluding ARR values. To determine the statistical significance of improvements, we conduct a bootstrap hypothesis test under the null hypothesis that performance results are better when the ARR is included, using a non-parametric bootstrap with 500 samples.

From the volatility forecasting R^2 values in Exhibit 7a, we can see that the ARR consistently improves 5-min and 1-hour forecasts for all non-linear models, and significant improvements are observed for linear models for 5-min sampling frequencies. This indicates that the ARR is informative for short-term forecasts, and can help enhance risk predictions in the near-term. For longer horizons however (i.e. 1-day and 1-week) we observe that the inclusion of the ARR reduces prediction accuracy – potentially indicating the presence of overfitting on the training set when ARRs are introduced.

For crash predictions AUROC results in Exhibit 7b, we note that the ARRs are observed to improve forecasts for all models and sampling frequencies – with statistical significance at the 99% level observed for both linear and GBDT forecasts over shorter horizons. This echoes the volatility forecasting results – indicating that ARRs can be useful to inform short-term risk predictions.

Conclusions

We introduce the Autoencoder Reconstruction Ratio (ARR) in this paper, using it as a real-time measure of asset co-movement. The ARR is based on the normalised reconstruction error of a deep sparse denoising autoencoder applied to a basket of asset returns – which condenses the returns vector onto a lower dimensional set of latent variables. This replaces the PCA modelling approach used by the Absorption Ratio of Kritzman *et al.* (2011), which allows the ARR to better model returns that violate basic PCA assumptions (e.g. non-Gaussian returns). Through experiments on a basket of 11 CRSP US sector indices, we demonstrate that the autoencoder significantly improves the out-of-sample reconstruction performance when compared to PCA, increasing the combined R^2 by more than 35%.

Given the links identified between increased asset co-movements and the fragility of the overall market in previous works (Kritzman *et al.*, 2011; Campbell *et al.*, 2002), we also evaluate the use of the ARR in a systemic risk application. First, we conduct an empirical analysis of the relationship between risk metrics of the combined market (using the CRSP US Total Market Index as a proxy) and the ARR computed from its sub sectors of the market. Based on an analysis of the KDE plots of risk metrics vs. ARRs, we show that low values of the ARR coincide with high volatility and high drawdown periods in line with previous findings. Next, we evaluate the information content of the ARR by testing how much the ARR improves risk predictions for a various benchmark models. We find that the ARR is informative for both volatility and market crash predictions over short horizons, and significantly increases 5-min and 1-hour forecasting performance across most model benchmarks.