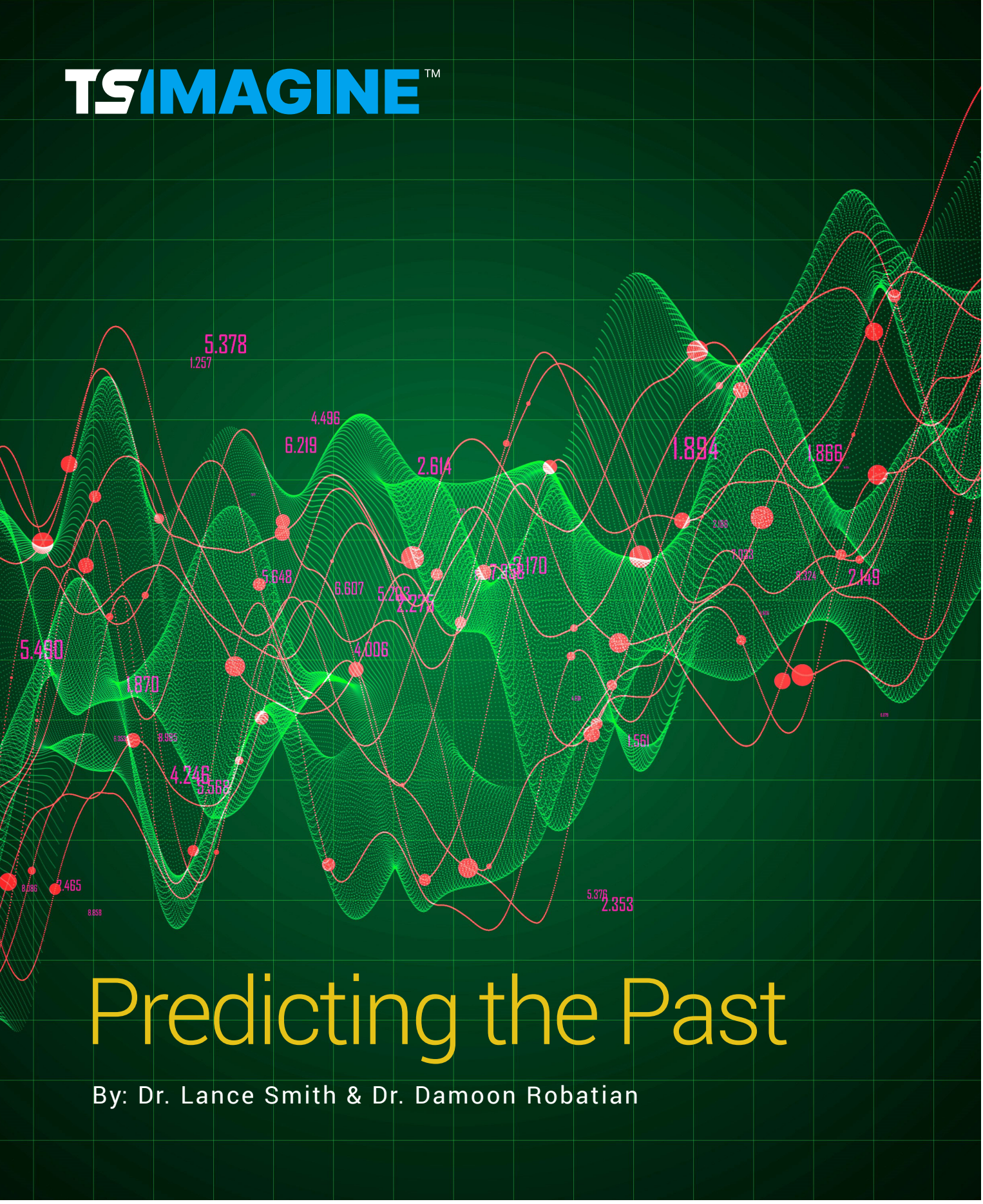
TSIMAGINE™



Predicting the Past

By: Dr. Lance Smith & Dr. Damoon Robatian

August 22, 2022

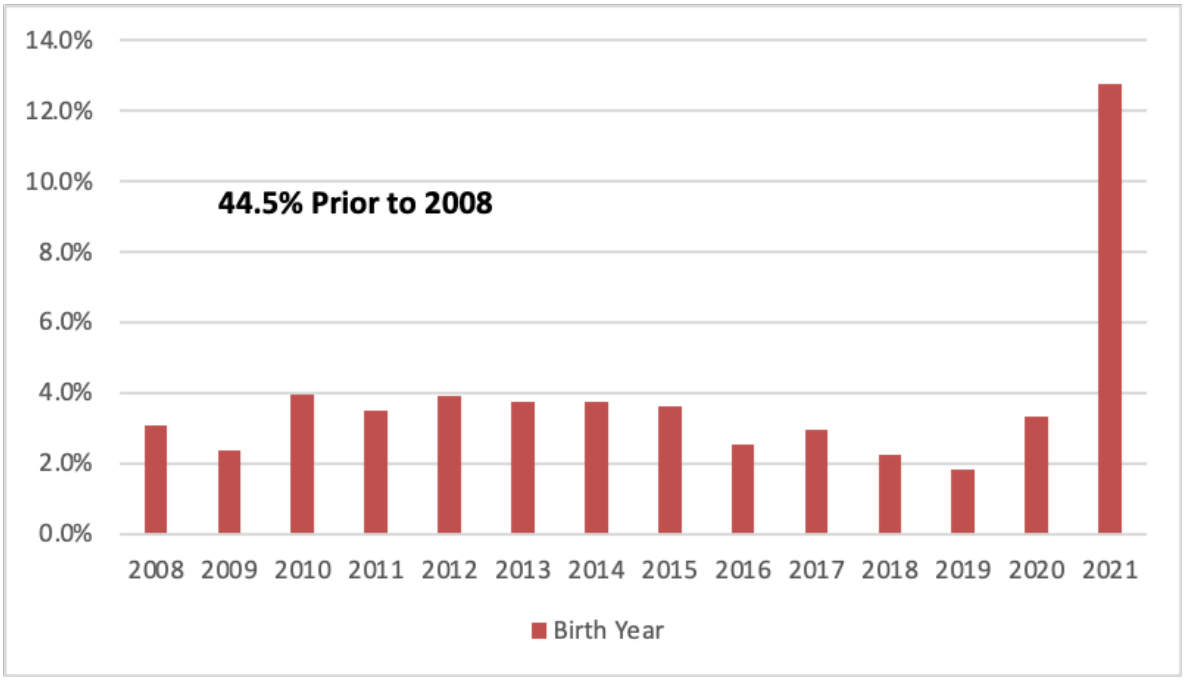
Predicting the Past

By: Dr. Lance Smith & Dr. Damoon Robatian

1 Introduction

Historical VaR (HVaR) has become a standard measurement of risk. Many firms now require a full twelve years of prices (plus data from further back such as the Great Recession of 2008–2009). However, this requirement introduces a conundrum: what do we do when a company has not been around for a full twelve years? In fact, fully half of today's US equities are less than twelve years old. The table below displays the distribution by birth year, going back to 2008.

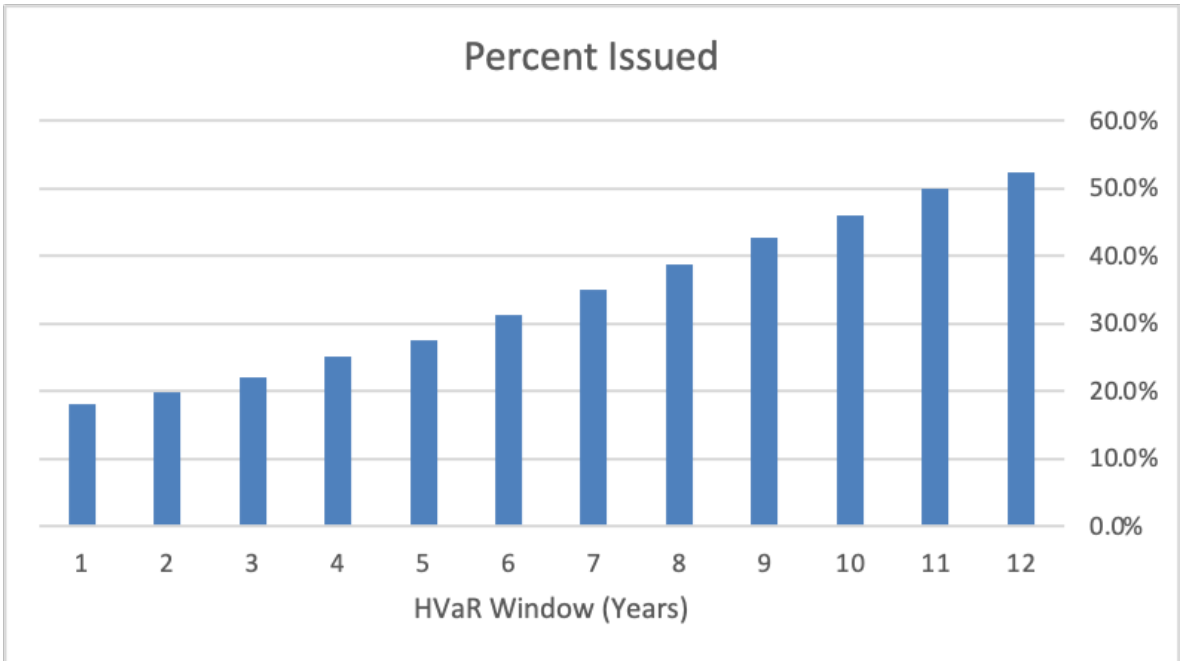
FIGURE 1



1 Introduction *(continued)*

We can also recast this information in terms of the HVaR window: For a given lookback period of HVaR, what percent of the US equities were first issued during that window?

FIGURE 2



The point is, that no matter what the HVaR window, there is likely a need to create synthetic returns. How this is actually done can affect the outcome of your risk calculations. The usual solution is to fallback to a benchmark index using a current beta to substitute for the nonexistent data. In the US that would typically be the S&P 500 Index. However, this introduces some unintended consequences: the returns that are being fabricated will be perfectly correlated with each other, and more importantly the synthetic returns will not adequately capture the tail risk.

1 Introduction *(continued)*

Proxying returns using a single factor

Let's write the return, Y, in the usual way as $Y = \beta X + \varepsilon$, with X and ε uncorrelated. Then (ignoring the mean),

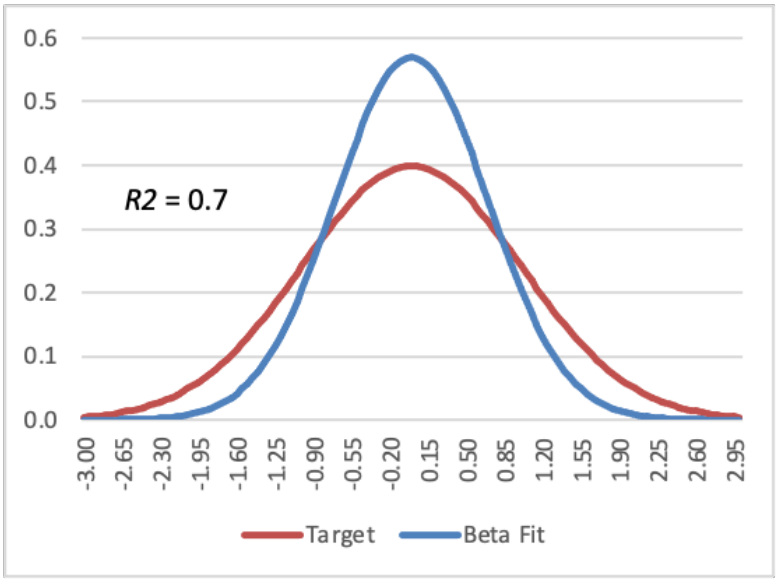
$$\mathbb{E}(Y^2) = \mathrm{Var}(Y) = \beta^2 \,\mathrm{Var}(X) + \mathrm{Var}(\varepsilon).$$

So that in particular the standard deviation is

$$\begin{aligned}\sigma(Y) &= \sqrt{\mathrm{Var}(Y)} \\ &= \sqrt{\beta^2 \,\mathrm{Var}(X) + \mathrm{Var}(\varepsilon)} \\ &> \beta\sigma(X).\end{aligned}$$

In fact, in terms of R^2, we have that $\beta\sigma(X) = R^2\sigma(Y)$. In practice, ordinary regression will yield an R^2 between 0.2 and 0.7, so that in any case the standard deviation of the proxy returns are at least 30% less than those of the actual returns. Now suppose for the moment that the returns are all normally distributed. Then as the following graph makes clear, the proxy return distribution simply cannot properly capture the tail behavior. But that is where HVaR lives. And so if this approach cannot even work for the case of normal returns, there is no hope for it in the real world.

FIGURE 3



2 The Idea

A better solution is to determine, if you can, a multi-factor representation of the missing returns. This approach goes like this:

- Determine a set of factors that collectively best approximate the returns of the target stock
  - The factors themselves need to have the full twelve years+ of history
- Regress the factor returns against the target returns over whatever period is available to determine the weights (a.k.a. multivariate betas). We imposed a cap at the past five years.
- When computing HVaR, fall back to the factor returns only when there are no target returns available, although one can choose to overwrite the target returns as well.

Note: The collection of factors can be tailored to each target stock. There is no requirement to use the same factors across the entire portfolio. This freedom enables us to achieve the best fit possible.

3 Implementation

We begin with a comprehensive list of factors — namely all US equities that have existed for the 12+ years of interest, numbering approximately 11,000. After making adjustments for multicollinearity, this number reduces to about 500. Then for a youthful equity, we choose the five that are most highly correlated and then perform a LASSO regression [Tibshirani, 1996] against these factors utilizing the available time series data to further reduce, if possible, the number of factors. The correlation screening together with the LASSO are called *sure independence screening* (SIS), which is known to improve high-dimensional regression analysis [Fan and Lv, 2008]. Significance of the regularization term (the tuning parameter) is adjusted based on a 5-fold cross-validated mean squared error (MSE). For the details of the variable selection, see section 6.3. However this algorithm still retains the same deficiency of the example above, in that R^2 will still be significantly less than 1, and hence the tails will not be adequately captured.

4 Chasing the Tails

We can more closely fit the tails if we emphasize the larger returns by reweighting. The details of this reweighting, and the corresponding modifications to the regression scheme are explained in the Appendix.

5 How good is the fit?

We first designed a test specifically addressing the tails. For each equity we can compute two numbers — the left 1% and the right 1% tails. That is, using the historical returns (details in Appendix), we compute $F_Y^{-1}(0.01)$ and $F_Y^{-1}(0.99)$. We then compute the same for the simulated returns, and then create a scatter plot so that we can compare the two. In this visualization, equities for which the two match perfectly would lie upon the line $y = x$. The next two plots show the results across all US equities that have been issued after Jan 1, 2008 — a universe of about 11,000. We only show the results for the left tail; the right tail is similar.

Unsurprisingly, almost every single point lies above the line $y = x$. (Figure 4). This is a reflection of the lower volatility of the simulated returns, as discussed previously. Now compare the five factor weighted fit (Figure 5): Fully 96% of the equities fall within the red bands, centered along $y = x$.

FIGURE 4

S&P Beta Fit
Lower 1% Tail

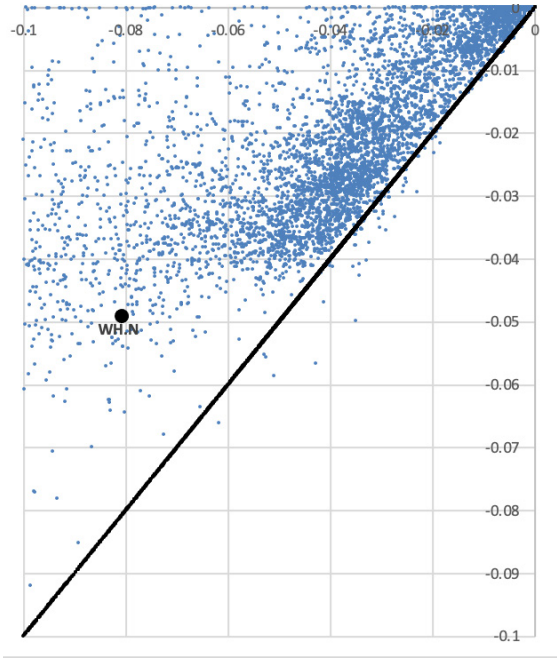
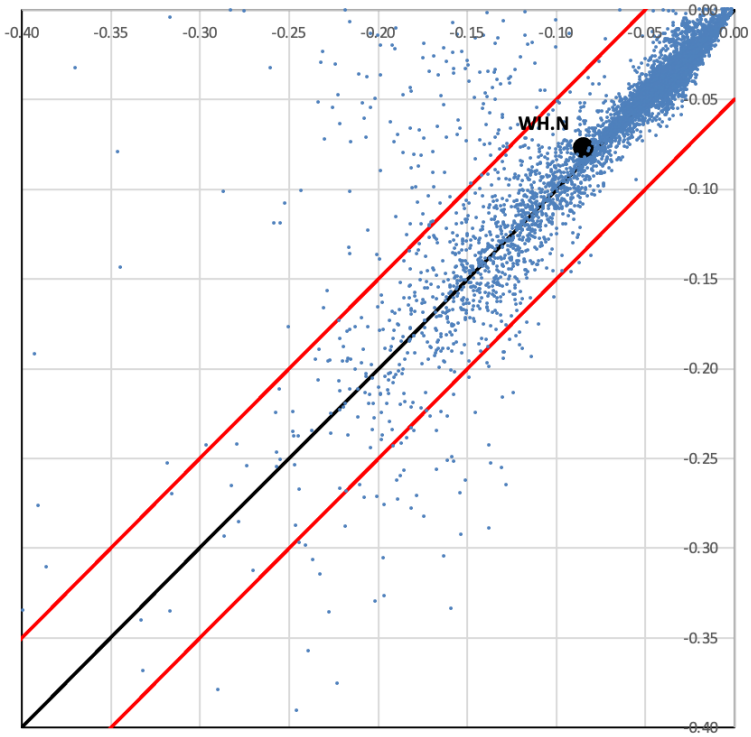


FIGURE 5

Weighted 5 Factor Fit
Lower 1% Tail



5 How good is the fit? *(continued)*

5.1 Example: WH.N (Wyndham Resorts)

FIGURE 6

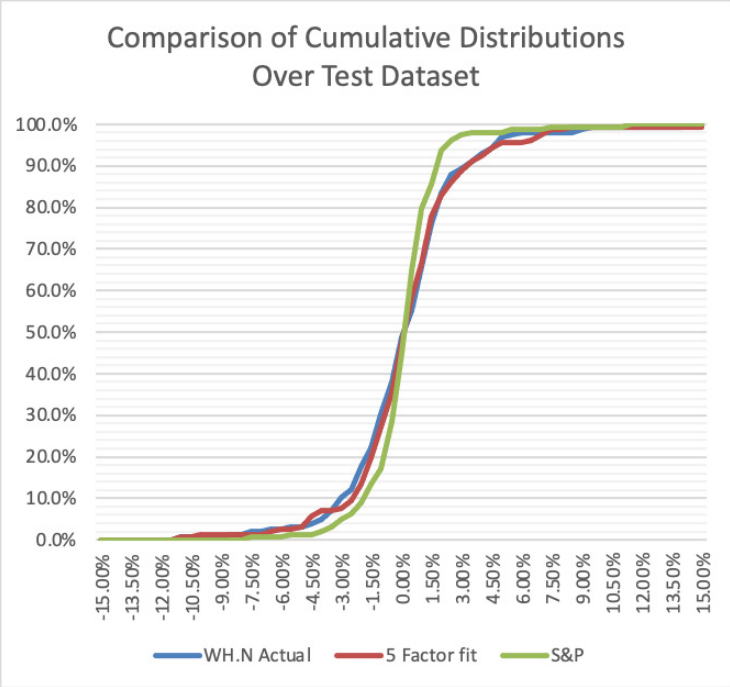
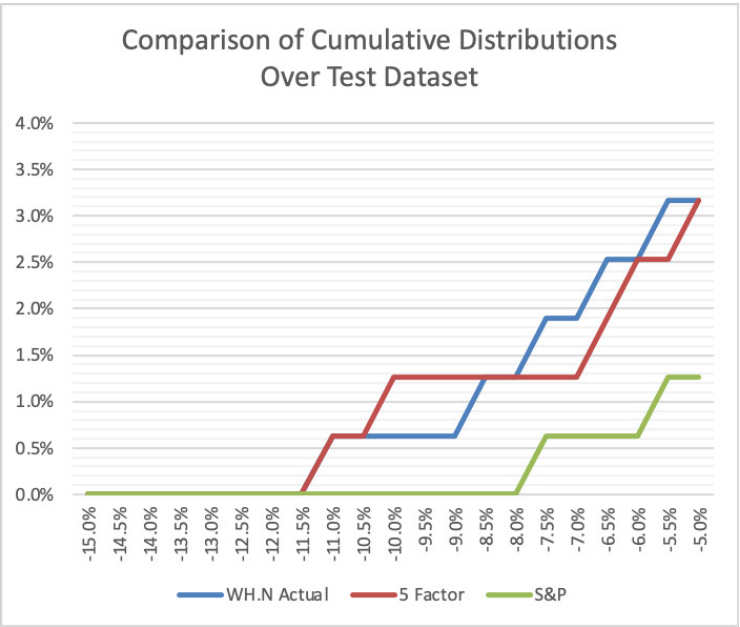


Figure 6 compares the three distributions for WH.N returns using the test data.

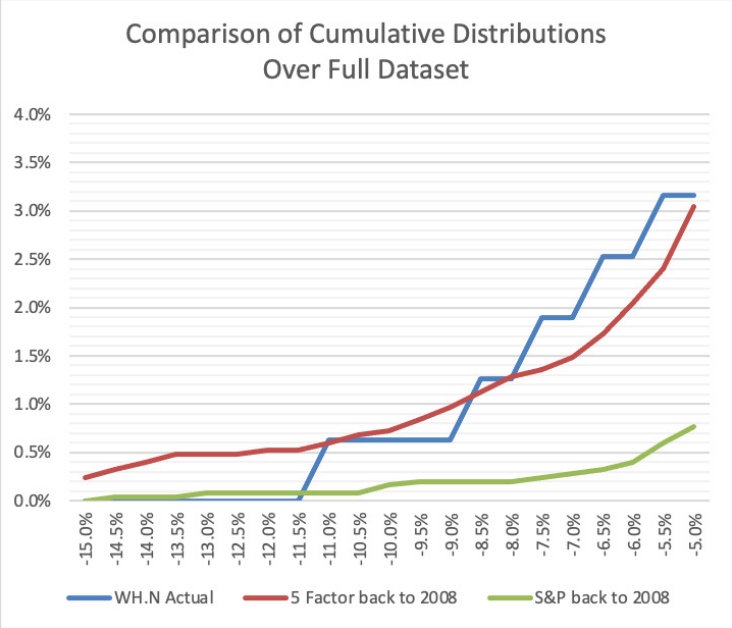
FIGURE 7



As WH.N began trading about four years ago, and the test data consists of 20% randomly chosen dates, there are only 158 data points, hence the wobbliness in the Figure 6 graph. Let's zero in on the left tail.

5 How good is the fit? *(continued)*

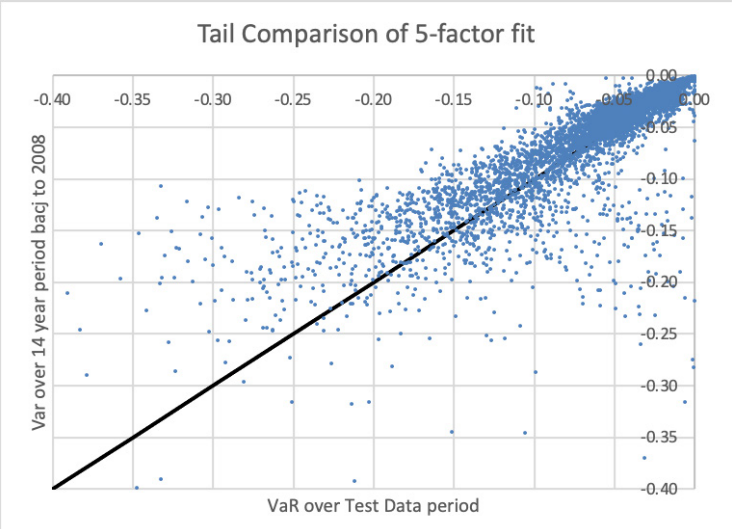
FIGURE 8



Of course, the data is sparse, but suppose we now enrich the two simulated distributions by using the entire 14 years of history going back to 2008/01/01.

We see that the simulated distribution for WH.N going back to 2008 continues to closely track the 1% threshold for WH.N. Is this a fluke? The scatter plot below suggests not.

FIGURE 9



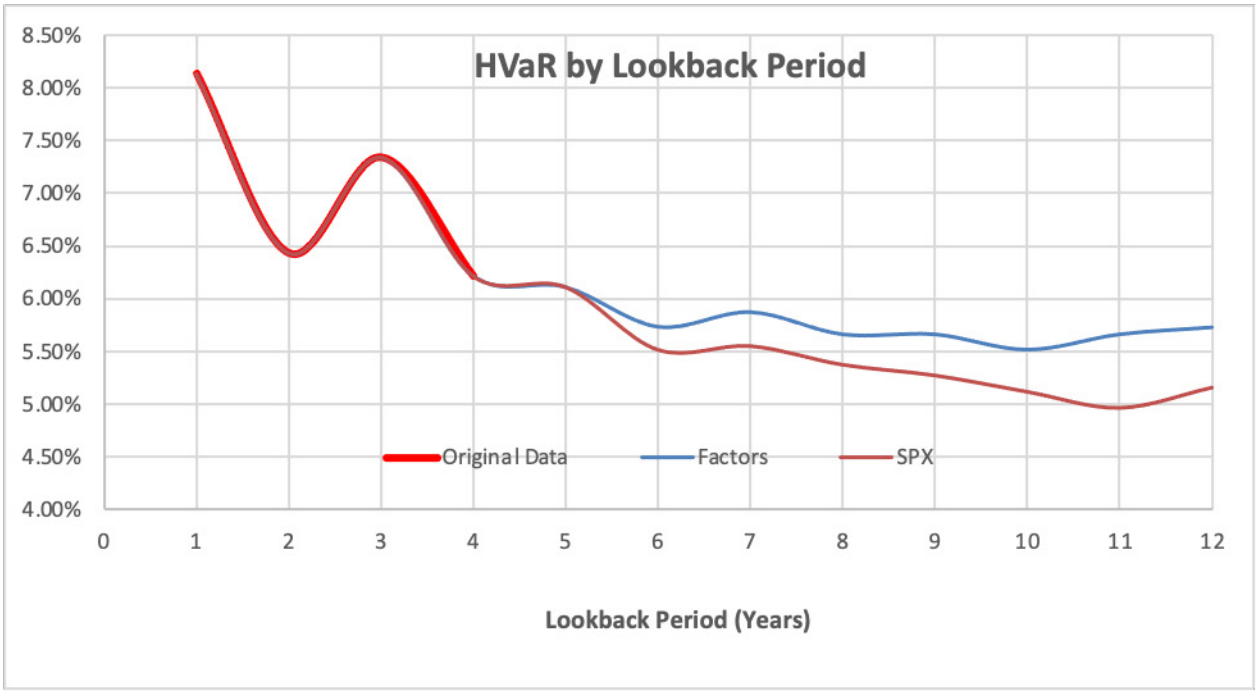
Some explanation is in order. This scatter plot compares the lower tail of the 5-factor sims for each equity using two datasets: test data vs data back to 2008. This shows that in general there is not significant tail dissipation when going to the longer time period. That is, the longer dated synthetic distributions continue to maintain their "fat tails".

6 Some sample HVaR calculations

The results of the regression for each equity were then loaded into its risk environment, making it available to be invoked to supply missing returns whenever needed for an HVaR calculation.

6.1 Example: A Long portfolio of equities

FIGURE 10



This graph shows that when the data is falling back, the weighted factor HVaR does a much better job at maintaining the tails that contribute to HVaR. (For the initial four year look-back period in all three cases the actual data is being used, so that the graphs coincide in that region).

Appendix

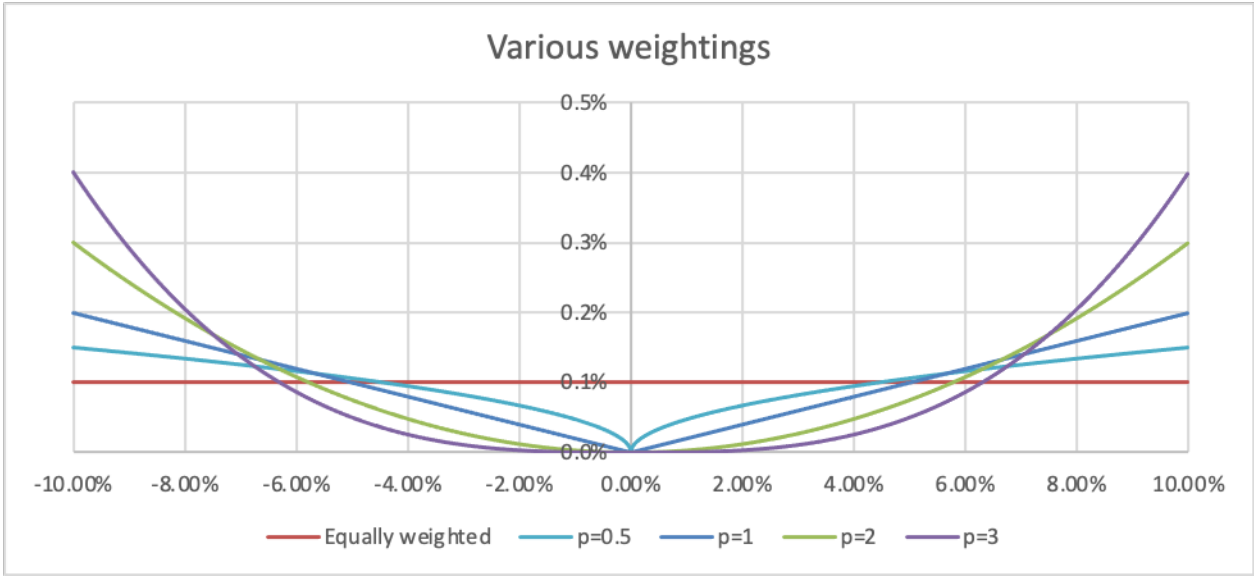
6.2 Capturing the Tails

The usual regression approximates the returns by regressing against a collection of factors; the LASSO adds on attempts to screen out the insignificant ones. Ultimately we arrive at an expression:

$$Y = \sum_{j=1}^{d} \beta_j X_j + \varepsilon, \tag{1}$$

with $\mathbb{E}(Y^2) = \mathbb{E}\left(\sum_j \beta_j X_j\right)^2 + \mathbb{E}(\varepsilon^2)$. Now it is clear that $\mathbb{E}(\varepsilon^2)$is likely to be significant,

FIGURE 11



which introduces a problem: if the goal is to better model the tail probabilities, how can we do this if the variance of the simulated returns (i.e. the factor portion) is significantly less than the actual variance of the data?

In our approach, we reweight the returns so that tail events have a higher weighting and will therefore be better modeled. This will necessarily give rise to a higher residual error, and so we will not be as close to the target for moderate returns. That is OK, because our ultimate goal is to estimate HVaR, which lives in the tails.

6.2.1 Weights

Suppose that $Y_1, Y_2, \ldots, Y_N$ are the factor returns. Let $p \geq 0$ be a fixed real. For every $i \in \{1, 2, \ldots, N\}$, define

$$w_{p_i} = \frac{|Y_i|^p}{\|\mathbf{Y}\|_p^p} \tag{2}$$

where $\|.\|_p$ denotes the p-norm.The weight w_{p_i} magnifies the effect of the larger returns. The case $p = 0$ is the base, equally weighted case. Figure 11 illustrates various weighting schemes for a sample of 1,000 data points.

6.2.2 Scale Invariance

We note that the weights are independent of scale. That is, if the returns are all rescaled (by the standard deviation, for example), the weights remain the same. Put another way, the choice of an optimal value of p is independent of the standard deviation, and is more influenced by the relative shape (fat tails) of the distribution.

6.3 Variable (Factor) Selection

Our initial collection of factors consisted of every US equity that has existed since 2008, numbering about 11, 000. We apply two types of selection criteria to reduce this number, in the following order:

1. *Variance inflation factor (VIF),*
2. *Weighted sure independence screening (WSIS)* [Fan and Lv, 2008], which itself comprises two components:
   a) Screening via weighted marginal correlations, and
   b) Variable selection using a regularization method.

Each of the above criteria are explained in the coming sections.

6.3.1 Variance Inflation Factor

Due to the strong multicollinearity among the factors, passing them directly to LASSO causes trouble in selecting significant factors. For this reason, we first reduce the number of the factors to about 500 applying the so-called variance inflation factor. It quantifies the severity of multicollinearity in an ordinary least squares (OLS) regression analysis. VIF provides a score illustrating how much the variance of a regression coefficient estimate is "inflated" as the result of collinearity among the covariates. Consider the linear model introduced by Equation (1). VIF for every independent variable X_j can be calculated in two steps as follows:

1. Regress X_j on the rest of the explanatory variables (excluding the response) in an OLS model, i.e.,

$$\begin{aligned} X_j = \alpha_0 + \alpha_1 X_1 + \cdots + \alpha_{j-1} X_{j-1} \\ + \alpha_{j+1} X_{j+1} + \cdots + \alpha_d X_d + \epsilon, \end{aligned} \tag{3}$$

   with ϵ being the error term.

2. Calculate the VIF factor for $\hat{\beta}_j$:

$$\mathrm{VIF}(\hat{\beta}_j) = \frac{1}{1 - R_j^2}, \tag{4}$$

   where R^2 is the R-squared of model (3).

A high VIF is interpreted as significant multicollinearity. In practice, 5 or 10 might be used as a threshold, i.e., if VIF($\hat{\beta}_j$) exceeds the threshold, X_j is regarded as highly collinear with the rest of covariates.

Given the number of factors we have, the computation of VIF is time consuming. Hence, we perform the reduction itself in two steps:

1. First, all factors are divided into batches of approximately 500 equities, and in each batch we keep only the stocks with VIF less than 10.
2. Second, all the remaining factors from all batches undergo another round of reduction altogether. However, this time any factor with VIF more than 5 is removed.

Finishing the two steps above, we end up with 538 factors.

6.3.2 Weighted Sure Independence Screening

Next, we explain the two components of the weighted sure independence screening.

(a) Screening

The main objective of feature screening in high-dimensional data is to achieve a massive decrease in the number of the independent variables. In general, the screening step of SIS ranks all features using the marginal correlation coefficient $\hat{\rho}(X_j, Y)$ and keeps only the highest correlated covariates. We apply the same idea with a slight modification. This modification is aligned with our objective of detecting the distribution tails better. That is, instead of the conventional Pearson correlation coefficient, we utilize a weighted correlation coefficient to measure the correlation between the response and each explanatory variable. Let $\mathbf{U} = (U_1, \ldots, U_N)$ and $\mathbf{V} = (V_1, \ldots, V_N)$ be two vectors. For a specific value of p we will let $\langle \mathbf{U}, \mathbf{V} \rangle_p$ denote the weighted inner product of $\mathbf{U}$ and $\mathbf{V}$ using the p-weightings, i.e.,

$$\langle \mathbf{U}, \mathbf{V} \rangle_p = \sum_{i=1}^{N} w_{p_i} U_i V_i,$$

where w_{p_i} is defined by Equation (2). We define the weighted correlation of $\mathbf{U}$ and $\mathbf{V}$ as

$$\rho_p(\mathbf{U}, \mathbf{V}) = \frac{\langle \mathbf{U}, \mathbf{V} \rangle_p}{\langle \mathbf{U}, \mathbf{U} \rangle_p^{\frac{1}{2}} \langle \mathbf{V}, \mathbf{V} \rangle_p^{\frac{1}{2}}}. \tag{5}$$

To complete the screening phase, it is necessary to fix the value of p. In fact, we treat p as a tuning parameter and pick the optimal p through a 5-fold cross-validated LASSO regression, for each individual response. The details of the tuning procedure are explained in sub-section 6.4. In the meantime, let us assume that the optimal value of p is known. Therefore, one can apply the weighted correlation, Equation (5), to rank the covariates according to the magnitude of their correlation with the response. Eventually, we retain only the five factors (out of 538) corresponding to the first five highest correlations.

(b) Selection by Regularization

After the large-scale reduction of the covariate numbers in the screening phase of SIS, any regularization method, such as LASSO, SCAD, or Elastic Net, can be used for variable selection. Here, LASSO is the method of our choice [Santosa and Symes, 1986, Tibshirani, 1996]. Before we break down the LASSO-

based selection, recall that we assumed that the optimal value of p is known. Now, given a response, the LASSO regression is run on the five covariates obtained in the screening step to further reduce the number of factors by shrinking some of the regression coefficients to zero. However, the LASSO is also weighted because of the very same reason as for the correlation. More accurately, instead of the residual sum of squares (RSS) in the LASSO cost function, we use the weighted RSS. That is, the objective function to be minimized has the following form:

$$\sum_{i=1}^{N} w_{p_i} \left(Y_i - \sum_{j=1}^{d} \beta_j X_{i_j} \right)^2 + \lambda \sum_{j=1}^{d} |\beta_j|, \tag{6}$$

where w_{p_i} is the weight vector defined by Equation (2). Note that, in the regular LASSO, $w_{p_i} = 1/\mathrm{N}$, for all i.

The Shoehorn Approach

We can trick the standard approach into solving the weighted problem by simply multiplying the target returns and the factor returns as well, by $\sqrt{w_{p_i}}$. Thus existing Python tools can be easily utilized by simply massaging the inputs.

6.4 Tuning p

First, consider a set of candidate values for p, namely, {0, 0.5, 1, 1.5, 2, 2.5, 3, 5}. Since capturing the tails are of primary importance, the optimal p value should result in regression coefficients $\hat{\beta}_j$ that best retrieve the extreme quantiles like the 1% and 99% quantiles. Now, fix a stock's daily returns as the response. Randomly, split the data into training and testing sets. We used 80% of the data for training and the remaining 20% for testing. For every candidate value of p repeat the following steps:

1. Find the 5 factors performing the WSIS screening on the training set.
2. Estimate the regression coefficients $\hat{\beta}_j$ through LASSO variable selection. At this point, we apply 5-fold cross validation for training the LASSO and tuning its tuning parameter λ. Obviously, this step is also completed by using the training set only.
3. Having estimated $\hat{\beta}_j$s, estimate the returns $\hat{Y}_i$ and, consequently, the quantiles of interest $\hat{\mathbf{q}} = (\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_m)$ Note that, in our implementation, we focused on the 1% and 99% percentiles, i.e., $\hat{\mathbf{q}} = (\hat{q}_{.01}, \hat{q}_{.99})$.

Let $\mathbf{q} = (q_1, q_2, \ldots, q_m)$ denote the corresponding vector of quantiles calculated using the observed

responses Y_i. In our case, $\mathbf{q} = (q_{.01}, q_{.99})$ Now, the optimal p is the one that minimizes

$$\sum_{k=1}^{m} \xi_k (\hat{q}_k - q_k)^2, \tag{7}$$

where $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_m)$ is an arbitrary vector with

$$\sum_{k=1}^{m} \xi_k = 1.$$

In fact, $\boldsymbol{\xi}$ adjusts the relative importance of each quantile. We assigned identical weights to the target quantiles, i.e., $\boldsymbol{\xi} = (0.5, 0.5)$.

References

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267288, 1996. ISSN 00359246. URL http://www.jstor.org/stable/2346178.

Jianqing Fan and Jinchi Lv. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5):849911, 2008. doi: 10.1111/j.1467-9868.2008.00674.x. URL https://onlinelibrary.wiley.com/doi/ 10.1111/j.1467-9868.2008.00674.x.

Fadil Santosa and William W. Symes. Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, 7(4):13071330, 1986. doi: 10.1137/0907087. URL https://doi.org/10.1137/0907087.

The Authors

Dr. Lance Smith

Chief Strategy Officer

As a co-founder of TS Imagine, Dr. Smith sets strategic direction, guides development of TS Imagine's award-winning analytics, and lends his expertise to sales efforts. Dr. Smith brings a rare combination of business and academic experience to his role as CEO, with thirty-plus years of direct involvement in trading and risk management practices, mathematical modeling, and system development.

Dr. Smith has a PhD in mathematics, and is a former Assistant Professor of Mathematics at Columbia University, as well as a former Courant Instructor at the Courant Institute. He began his finance career at Salomon Brothers in 1986, where he served as the chief quant for the equity derivatives business. He then headed Sakura Global Capital's Equity Derivatives Trading desk, prior to founding TS Imagine in 1993.

Dr. Damoon Robatian

Senior Data Scientist

Damoon Robatian holds a PhD in mathematics and works as a senior data scientist at TS Imagine. He joined the company in January 2022.

While for several years, his academic research was aimed at the theory of topological dynamics (a branch of pure mathematics), in 2015, he also became interested in data science. Since then, Damoon has been focusing on research in statistical learning theory and the mathematical foundations of function estimation via neural networks.

Damoon conducted extensive research on learning from size-biased, incomplete data during his second PhD. From the practical viewpoint, analysis of unrepresentative data is essential because such data are frequently encountered in multiple areas, including finance, actuarial science, and reliability engineering.

As a statistical learning expert, before joining TS Imagine, Damoon collaborated with companies such as Air Canada and Huawei Noah's Ark Lab to help them solve data problems in flight scheduling and telecommunications. He has been a statistical learning lecturer and course designer at the McGill School of Continuing studies since 2018.

TS IMAGINE

The leader in trading technologies.

Born from the combination of two leading SaaS platforms, TradingScreen and Imagine Software, TS Imagine provides real-time financial securities trading, portfolio management, and risk solutions for capital markets. The platform is uniquely positioned to streamline complex and tedious workflows across front, middle, and back-office functions.

TS Imagine has over 400 employees in 10 offices worldwide, serving approximately 500 buy-side and sell-side institutions in North and South America, EMEA, and Asia-Pacific, including hedge funds, traditional asset managers, pension funds, mutual funds, and financial institutions.

Headquartered in New York City, TS Imagine has offices in London, Hong Kong, Paris, Singapore, Tokyo Zurich and Montreal.

NEW YORK LONDON HONG KONG PARIS SINGAPORE TOKYO ZURICH MONTREAL

New York	+ 212 359 4100
London	+ 44 0 207 149 3100
Hong Kong	+ 852 3192 7533
Paris	+33 1 5332 2950
Singapore	+65 6499 9800
Tokyo	+81 0 3 6704 4334
Zurich	+ 41 43 210 8630
Montreal	+1 514 313 9450

TS IMAGINE™

www.tsimagine.com