Quantitative Finance

ISSN: 1469-7688 (Print) 1469-7696 (Online) Journal homepage: https://www.tandfonline.com/loi/rquf20

Enhancing the momentum strategy through deep regression

Saejoon Kim

Quantitative Finance, 2019
https://doi.org/10.1080/14697688.2018.1563707

Enhancing the momentum strategy through deep regression

SAEJOON KIM*

Department of Computer Science and Engineering, Sogang University, 35 Baekbeom-ro, Mapo-gu, Seoul 04107, Korea

(*Received 4 June 2018; accepted 19 December 2018; published online 25 January 2019*)

Momentum is a pervasive and persistent phenomenon in financial economics that has been found to generate abnormal returns not explainable by the traditional asset pricing models. This paper investigates some variations of the existing momentum strategies to increase profit and gain other desirable properties such as low kurtosis, small negative skewness and small maximum drawdown. We investigate these by using regression that is based on the latest techniques from deep learning such as stacked autoencoders and denoising autoencoders. Empirical results indicate that our regression-based variations can generate increased returns, and improved higher-order moments and maximum drawdown characteristics. Furthermore, our results reveal such improved performance can only be attained through the use of the latest deep learning technologies.

Keywords: Momentum strategy; Deep learning; Deep neural networks; Autoencoders; Denoising autoencoders

JEL Classification: C45, G11

1. Introduction

Momentum strategy as an investment tool has existed for a long time (Hurst *et al.* 2017). Years of extensive research by academics and practitioners indicate that trending of asset prices is pervasive across diverse asset classes and periods of varying economic conditions. Momentum strategies have generated returns that cannot be explained by the traditional asset pricing models (Fama and French 1996), and for this reason, momentum has been called the ‘premier anomaly’ in financial economics.

Jegadeesh and Titman (1993) have found that stocks that have performed well in the past can outperform, in the near future, stocks that have performed poorly in the past by as much as 1.49% per month. Consequently, they have shown that buying past winners and selling past losers simultaneously can generate favorable excess returns. Nearly a decade later, in Jegadeesh and Titman (2001), they document that their original results in Jegadeesh and Titman (1993) remained equally valid in the following years after their first publication. It was noted that this persistent profit-generating behavior is a characteristic unique to the momentum strategy that is not found in other anomalies such as the small-sized stocks or value stocks. In Jegadeesh and Titman (2001), an equal-weighted (EW) portfolio for each decile in the stock returns ranking of the look-back period was constructed, and the top decile was bought and the bottom decile was sold to implement the momentum strategy. This momentum strategy is commonly known as the WML (winner-minus-loser) strategy and is self-financing by construction. It has been demonstrated to produce profits of at least 1% per month in the holding period of one year.

In Asness *et al.* (2013), along with the finding of a co-movement between momentum and value strategies, a consistent momentum premium across various global markets and assets classes including stocks, currencies, bonds and commodity futures was observed. In Fama and French (2012), it was discovered that the momentum phenomenon is valid not only across international markets but also to small capitalization stocks. A related momentum strategy that only buys the top decile and does not short stocks has also been demonstrated to produce good excess returns (Jegadeesh and Titman 1993, 2001). We shall call this strategy the WO (winner only) strategy in this paper.

The described types of momentum strategy are more broadly called the *cross-sectional* momentum strategy in which the strategy buys *relative* winners and sells *relative* losers in comparison to other assets in the investable universe, or just buys the relative winners without further consideration about the absolute returns they generate. This comprises the majority of the literature on momentum strategy. Another type

*Corresponding author. Email: saejoon@sogang.ac.kr

of momentum strategy that considers only the *absolute* returns for the buy or sell trade is called the *time-series* momentum strategy (Moskowitz *et al.* 2012, Hurst *et al.* 2017). In this momentum strategy, the return of each asset in the look-back period determines whether to buy or sell the asset for the holding period. If the return in the look-back period was positive or negative, then the buy or sell trade is executed, respectively. Therefore, in the time-series momentum strategy, the portfolio is constructed based on each asset's own absolute performance while in the cross-sectional strategy, it is constructed based on the relative performances of the assets. In Moskowitz *et al.* (2012), extensive examination of the time-series momentum strategy was conducted and persistent positive excess return performance was observed from 1985, the start year of the examination in that paper. In Hurst *et al.* (2017), equal-weighted time-series momentum strategies for 67 markets across four major asset classes for over a century dating back to 1880 were investigated and the results showed consistent and robust performance across various markets and asset classes under consideration. In particular, time-series momentum persisted through various economic turmoils including the Great Depression and the more recent financial crisis. It is believed that trends in prices are recurrent phenomena and they are mainly due to human behavioral biases such as anchoring and herding (Hurst *et al.* 2017).

The aforementioned studies all indicate that both the cross-sectional and the time-series momentum strategies perform extremely well under most economic conditions. However, it was observed in Barroso and Santa-Clara (2015) and Daniel and Moskowitz (2013) that the WML momentum strategy can occasionally experience devastating 'market crashes' that wipe out returns accumulated over multiple number of years prior to the crash. Specifically, it was reported in Daniel and Moskowitz (2013) that during July and August of 1932, the return of the winner decile portfolio was 32% while that of the loser portfolio was 232%, and during March and May of 2009, the return of the winner decile portfolio was 8% while that of the loser decile portfolio was 163%. Clearly, the WML portfolio underperforms significantly during these two periods, and Grundy and Martin (2001) have elucidated this behavior by noting that past winners tend to have low beta while past losers tend to have high beta after bear markets which results in overall negative beta. In other words, these market crash periods of the WML strategy correspond to rebound periods after prolonged decline of large magnitude in the market. By decomposing the momentum volatility, it was recently shown in Barroso and Santa-Clara (2015) that volatility-scaling, and not time-varying beta, is the key to risk management as hedging against the market beta alone does not remove a large portion of the risk as was also observed in Daniel and Moskowitz (2013). Positive experimental results supporting the analysis were provided in Barroso and Santa-Clara (2015), and their volatility-scaling technique which targets constant volatility using realized variance of the past six months has resulted in stable excess return performance across various economic conditions. It has achieved distinct overall performance improvement compared to the plain WML strategy, i.e. WML strategy without volatility-scaling. For a similar reason, volatility-scaling was employed for the time-series momentum strategy in Moskowitz *et al.* (2012) to achieve more stable performance.

In this paper, we take a different approach to performance improvement to the existing momentum strategies. Our approach in a way selects and retains only the positive features from WML and WO strategies, and can be used in conjuction with the volatility-scaling technique. More specifically, our strategy is based on the finding that if the market is rising so that even the bottom decile's return is positive, then shorting the bottom decile can only negatively effect the return of the WML strategy, and, in this case, the WO strategy is preferred to the WML strategy. Similarly, if the market is falling so that even the top decile's return is negative, then while the WML strategy is preferred to the WO strategy in this case, going long the top decile would reduce the return of the WML strategy. This finding leads naturally to our proposed momentum strategy which is the following. Place the buy trade for the top decile if the expected return of its holding period is positive, and no trade otherwise. Similarly, place the sell trade for the bottom decile if the expected return of its holding period is negative, and no trade otherwise. Clearly, the correct prediction of the returns in the holding period is the key determinant in making our proposed strategy a profitable one, and in particular, it can generate, by construction, higher returns than WML and WO portfolios. Our strategy, therefore, efficaciously combines the cross-sectional and the time-series momentum strategies.

It remains to describe how the calculation of the expected return is made upon which the success of the proposed strategy hinges. To this end, we construct regression schemes that are based on the state-of-the-art machine learning algorithms to make predictions of the returns of the top and bottom deciles.

Machine learning algorithms have been used to solve regression problems in finance in the past. In particular, support vector regression has been demonstrated to be quite successful for financial time series prediction; see, for example, Sapankevych and Sankar (2009), Kim and Heo (2017), Kim (2018), Law and Shawe-Taylor (2017) and references therein. More recently, with the advent of *deep learning* technologies for image classification applications, deep learning-based regression schemes have been devised to make financial time series predictions. Notably, Bao *et al.* (2017) used deep neural networks to make stock price forecasting and Heaton *et al.* (2016) also used deep neural networks to construct a portfolio of stocks for index reconstruction. Deep learning has also been utilized for other domains in finance such as for high frequency trading applications in Dixon *et al.* (2018) and Sirignano and Cont (2018), for options pricing in Berner *et al.* (2018) and for mortgage risk assessments in Sirignano *et al.* (2018). The aforementioned methods for financial time series prediction share the explicit use of *deep features*, which is the defining characteristic of deep learning, for their respective objectives to obtain improved results. In this paper, we apply both shallow and deep machine learning algorithms in our return prediction setting. We use support vector regression as the choice of shallow learning algorithm as it is generally accepted as the best regression scheme among shallow learning technologies (Kim and Heo 2017, Kim 2018). Then, we propose deep neural network architectures that generate more accurate return predictions than ones made by support vector regression. Furthermore, we will empirically show that our proposed momentum strategies based on deep learning generate significantly improved returns in various forms, higher-order moments and maximum drawdown compared to both the WML and WO strategies.

Contribution of this paper is three-fold. The first contribution is the construction of the new momentum strategy that is based on selective long/short determined through the prediction of returns in the holding period as described above. In the second contribution, which is our main one, we propose and apply some of the latest architectures in deep machine learning for return prediction in our momentum strategies. To the best of our knowledge, this is the first work that these architectures are established for use in any type of financial applications domain. More importantly, it will be demonstrated that our proposed architectures will lead to momentum strategies that possess the best properties among the strategies we consider in this paper. In particular, it will be shown that our proposed deep neural network architecture is required to produce results superior to those by WML and WO strategies as consistent superior results were not attainable by other strategies considered in this paper. In addition to these architecures, we also apply state-of-the-art deep learning techniques for portfolio's return prediction which constitutes our last contribution. This is also the first work known to the author in which these deep learning techniques in any form are utilized in the domain of momentum strategy. Hyperparameters of deep neural networks' architectures will carefully be chosen to obtain improved performance results. We also document extensive performance results of various momentum strategies considered in this paper for the period that spans over 60 years from 1953 to 2017 using data taken from Kenneth French's data library (French 2018). This data library provides daily and monthly returns of the momentum portfolio from 1927 to present, and since the proposed strategies in this paper require training data, the performance results shown here are from the entire period provided in French's data library at the time of this writing.

The outline of this paper is as follows. In the next section, we review the mentioned momentum strategies and their risk-managed versions obtained by volatility-scaling. In Section 3, we present the main contribution of this paper, namely, the selective long/short momentum strategies facilitated by prediction algorithms from shallow and deep machine learning. In Section 4, extensive performance results of the momentum strategies discussed in this paper are compared and contrasted, and we end the paper with conclusions in the following section.

2. Existing momentum strategies

The most prevalently considered form of momentum strategy is the winner minus loser (WML) strategy and the long-only or the winner only (WO) strategy. In this paper, we will consider the implementations of these strategies in which, in the former, the top decile of stocks is bought and the lowest decile of stocks is sold, while in the latter, only the top decile of stocks is bought. We will also consider the implementation in which the look-back period, interchangeably called the formation period, for the portfolio is the recent 12-month skipping the most recent 1-month, and the portfolio is rebalanced monthly. The returns of these portfolios using monthly returns data are listed at Kenneth French's website (French 2018) which we used exclusively in this paper.

Large gains of the WML strategy contain fat left tail in the return curve that is evidenced by high kurtosis and large negative skewness pointing to large maximum drawdown (Barroso and Santa-Clara 2015). In particular, the WML momentum strategy generated -91.59% during a two-month period in 1932 and -73.42% during a three-month period in 2009 both periods corresponding to market rebounds after large decline. For this matter, improvements of the plain WML strategy have followed. To hedge such market crashes of the WML strategy, past 6-month variance was used as the estimate of the variance of the holding period of one month in Barroso and Santa-Clara (2015). This estimate was then used to scale the portfolio size inversely proportional to the estimate of the variance which had the effect of the loss generated from the market rebound being sufficiently diminished. We call this hedged momentum strategy by *volatility-scaled* momentum strategy in this paper and it can be applied to both the WML and WO strategies. Specifically, to describe the strategy, let us first introduce the following notation.

$\{r_t\}_{t=1}^{T}$, monthly returns of plain momentum strategy

$\{r_t(d)\}_{d=1}^{D}$, daily returns in month t of plain momentum strategy

The daily returns of the momentum strategies WML and WO are also listed in French's website. Then, the forecast of the variance of the holding period at month t, $\hat{\sigma}_t^2$, using the variance of the previous six months is calculated as follows

$$\hat{\sigma}_t^2 = \frac{21}{126}\sum_{\tau=1}^{6}\sum_{d=1}^{21} r_{t-\tau}^2(d), \tag{1}$$

and the return for the plain momentum strategy is accordingly scaled as

$$r_t^* = \frac{\sigma_{\text{target}}}{\hat{\sigma}_t} \cdot r_t, \tag{2}$$

where r_t and r_t^* are the returns of the plain momentum and volatility-scaled momentum strategies, respectively, in month t, and σ_{target} is the target variance which is a constant. The ratio of standard deviations in equation (2) is the *weight* of the volatility-scaled momentum strategy, and σ_{target} was set to the value corresponding to the annualized volatility of 12%. In summary, we consider four types of existing momentum strategies in this paper: plain WML, plain WO, volatility-scaled WML, and volatility-scaled WO strategy.

3. Selective long/short momentum strategies

In this section, we describe our proposed selective long/short momentum strategies which is not purely WML or WO. In the

WO strategy, the investor *always* buys the top 10% of stocks in the investable universe regardless of the trend of the market or that of the top 10%. This strategy usually generates good excess returns, however, is exposed to large losses when the market experiences a precipitous decline so that even the top 10% generates negative returns. To resolve this, that is, to form a strategy that can generate profit independently of the market condition, in the WML strategy, the investor buys the top 10%-performing stocks and sells the bottom 10%-performing stocks simultaneously forming a market-neutral strategy. Empirical results have shown that the WML strategy generates profit consistently and persistently throughout various conditions of the market. Furthermore, if the WML strategy is traded in conjunction with volatility-scaling, it was shown in Barroso and Santa-Clara (2015) and Daniel and Moskowitz (2013) that the volatility-scaled WML strategy can generate returns that are significantly higher than the pure WML strategy.

Both the WO and the WML strategies have strengths and weaknesses that are complementary. Our proposed selective long/short momentum strategy makes efficacious use of the strengths of the two strategies through predictions of the top and bottom deciles' returns, and therefore extends the modification of WML with respect to WO by *selectively* buying the top decile and *selectively* selling the bottom decile. In other words, in our proposed momentum strategy, the investor buys the top decile only when some condition is satisfied, and sells the bottom decile only when another condition is satisfied. Specifically, in our proposed momentum strategy, for the top decile,

$$\begin{array}{ll} \text{Buy} & \text{if } \hat{r} > 0, \\ \text{No trade} & \text{otherwise,} \end{array}$$

where $\hat{r}$ is the predicted return of the top decile in holding period. Similarly, for the bottom decile,

$$\begin{array}{ll} \text{Sell} & \text{if } \hat{r} < 0, \\ \text{No trade} & \text{otherwise,} \end{array}$$

where $\hat{r}$ is the predicted return of the bottom decile in the holding period. As our strategy selectively buys the winners and sells the losers, we call it the *selective winner minus loser* (SWML) strategy. If the *sign* of the predicted return is the same as that of the actual return of the holding period, then, unlike WML or WO strategy, our SWML strategy never generates negative return in either side of the trade! In fact, when no trade is executed for both long and short sides, one can invest in, say, the Treasury bonds to earn risk-free interest rate. Therefore, the success of our SWML strategy is determined by the correct prediction of the 'sign' of the return in the holding period.

In the following subsections, we will describe how the returns of the holding periods of the top and the bottom deciles are predicted using various techniques from machine learning. Specifically, the techniques will include support vector regression (SVR), deep neural network (DNN) using stacked autoencoder (SAE), and DNN using stacked denoising autoencoder (SdAE). Momentum strategy that is based on the last of these techniques, namely, DNN using SdAE, will be our main proposed strategy. But first, let us start with the prediction technique that has had the most prevalent use in financial economics, namely, the look-back period-based prediction. In all of our experiments for the various strategies, the returns under consideration are monthly returns, unless specified otherwise, and the look-back or the formation period of 12 months, and the holding period of one month were adopted.

3.1. *Look-back period-based selective long/short*

In this strategy, the return of the look-back period is used as the estimate for the sign of the return in the holding period. On the day before the start of month t, the sign of the following

$$\prod_{i=1}^{12}(1 + r_{t-i}) - 1$$

is calculated as the prediction of the sign of r_t, the monthly return of month t. Using the value obtained in the look-back period to predict the corresponding value of the holding period not only provides one with a simple and intuitive approach to prediction, it also has been found to be an effective prediction method for, say, the volatility of asset prices (Moskowitz *et al.* 2012, Barroso and Santa-Clara 2015).

3.2. *SVR-based selective long/short*

Support vector regression (SVR) has been used extensively in the past to make predictions of diverse set of time-series variables (Sapankevych and Sankar 2009), and in particular, it has also found success in financial applications for, say, predicting the stock prices (Kim and Heo 2017) and predicting the volatility of portfolios (Kim 2018). SVRs are one of the best shallow machine learning techniques.

In our SVR-based SWML strategy, SVR is used to predict the monthly return of month t, r_t, using the monthly returns of the previous 12 months, $(r_{t-1}, r_{t-2}, \ldots r_{t-12})$. Training set for this regression of r_t consists of the most recent 300 input-output pairs in which for each pair the input is

$$(r_{t'-1}, r_{t'-2}, \ldots, r_{t'-12})$$

and output is

$$r_{t'}$$

for $t' = t-1, t-2, \ldots, t-300$.

In our experiments, we used is the Gaussian function as the kernel, and the most recent 10% of training set data was used as the validation set to pick the best triple (C, ϵ, γ) for $C \in \{10^i,\ 5 \cdot 10^i : -2 \le i \le 3\}$, $\epsilon \in \{10^i,\ 5 \cdot 10^i : -5 \le i \le -1\}$, $\gamma \in \{10^i,\ 5 \cdot 10^i : -5 \le i \le -1\}$. Here, C is the regularization constant, ϵ is the tube size, and γ is the parameter of the Gaussian kernel. The accruracy measure used in the validation set was the mean square error (MSE) measure. Thus, if r_t represents the actual monthly return of month t and $\hat{r}_t$ represents the predicted monthly return of month t by SVR, the triple that minimized $\frac{1}{30}\sum_{i=1}^{30}(r_{t-i} - \hat{r}_{t-i})^2$ was chosen as the parameter for the prediction of r_t.

3.3. *DNN using SAE-based selective long/short*

In deep neural networks (DNNs), the input goes though multiple processing stages to generate *deep features* that are purportedly better representations of the input than the features generated by *shallow learning* techniques that include support vector regression. Many realizations of DNNs exist and they include the convolutional neural networks, recurrent neural networks and restricted Boltzmann machines to name a few. In this paper, we consider the realization in which stacked autoencoder (SAE) is used to hierarchically extract deep features as it suits our purpose of return regression better than other ones. DNN using SAE was first conceived in Bengio *et al.* (2006), and has been tested for stock price forecasting in Bao *et al.* (2017), for constructing portfolios that replicate stock indices in Heaton *et al.* (2016) and for classifying hyperspectral data in Chen *et al.* (2014) all with great successes.

DNN using SAE consists of a cascade of layers of autoencoders followed by the last (or the highest level) layer which is normally chosen for particular supervised learning application. For the classification application, for example, the logistic regression layer is normally used for the last layer. Each autoencoder in SAE can be formulated as the following for an input $x \in \mathbb{R}^d$:

$$y = f(W_y x + b_y), \tag{3}$$

$$z = f(W_z y + b_z). \tag{4}$$

Here, $y \in \mathbb{R}^{d'}$ represents the (learned) latent representation, and $z \in \mathbb{R}^d$ is the reconstructed version of x. f is some activation function, typically nonlinear, whose examples include the sigmoid function and the rectified linear function. The goal of the autoencoder is to minimize the reconstruction error of z which can be measured by the squared error loss $L_s(x, z) = ||x - z||^2$. This is achieved through estimating the weights W_y, W_z, b_y, b_z where W_y is a $d' \times d$ weight matrix, b_y is a length-d' weight vector, W_z is a $d \times d'$ weight matrix and b_z is a length-d weight vector. Estimation of the weights in equations (3) and (4) constitutes the pretraining of an autoencoder. An example of an autoencoder is illustrated in figure 1. In a nutshell, the functionality of an autoencoder is mapping an input x to its reconstructed version z through the latent representation y. If the mapping is defined so that z is a good approximation to x, then this suggests that y serves as a good features set that well describes the input. To avoid the trivial cases of y and z in which the autoencoder simply performs the identity mapping ($d = d'$ case) or the principal component analysis ($d' < d$ case), nonlinear $f(x)$ is conventionally used, and in particular, we used the sigmoid function $f(x) = 1/(1 + e^{-x})$ as is normally done in the literature (Chen *et al.* 2014).

In a stacked autoencoder, from the autoenoder in the first (or the lowest level) layer, its latent representation y is used as the input for the next autoencoder of the next layer. This second autoencoder is then pretrained just as the first layer to produce its latent representation and the reconstructed output. The latent representation of this second layer is then used as the input for the next layer of autoencoder, and this process is repeated until the last autoencoder layer. In the final layer of the overall architecture, a supervised regression is applied, and in particular, the latent representation of the last layer of autoencoders is used as the input to this final layer. In our contruction, the supervised regression layer makes the prediction of the monthly return of the holding period. As the highest level layer is implemented by a supervised machine learning algorithm, the overall DNN provides supervised learning.

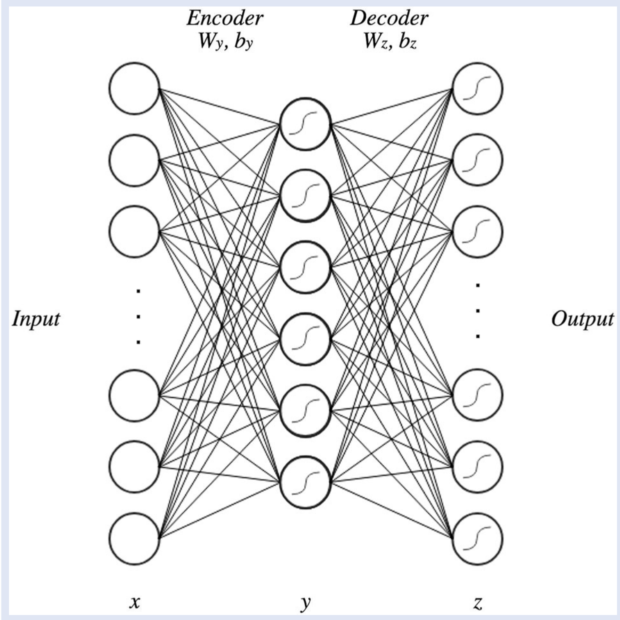

Figure 1. Autoencoder.

The DNN using SAE that we used in our experiments is illustrated in figure 2. It shows an architecture that has 12 input nodes, 3 hidden layers each having 6 nodes, and 1 output node. The 12 input nodes represent the monthly returns of previous 12 months, i.e. $(r_{t'-1}, r_{t'-2}, \ldots, r_{t'-12})$ and the output node represents the monthly return of the current month $r_{t'}$. Labeling the input to the last regression layer as $y_1, y_2, \ldots, y_6$ with the corresponding weights as $w_1, w_2, \ldots, w_6$ and a bias term as b, the weights and the bias term of the regression layer are fine-tuned to minimize the MSE between the monthly return of month t', $r_{t'}$, and the predicted monthly return of month t', $\hat{r}_{t'}$, where $\hat{r}_{t'} = \sum_{j=1}^{6} w_j y_j + b$. The training and test sets for the DNN-based predictions were chosen to be the same as those for the SVR-based prediction for fair comparison. Once the DNN using SAE is pretrained and fine-tuned using $(r_{t'-1}, r_{t'-2}, \ldots, r_{t'-12})$ as the input and $r_{t'}$ as the output for $t' = t-1, t-2, \ldots, t-300$, $(r_{t-1}, r_{t-2}, \ldots, r_{t-12})$ is fed into the network which outputs the prediction of r_t.

3.4. *DNN using SdAE-based selective long/short*

Denoising autoencoder is an extension of the ordinary autoencoder described in the previous subsection, and has originated in the quest for a better understanding of what constitutes a ‘good’ representation of the input. It is based on the idea that the latent representation that minimizes only the reconstruction error may not be a sufficiently useful one in the sense that if the input was initially noise-corrupted, then the reconstructed version of it would not be a useful one. In

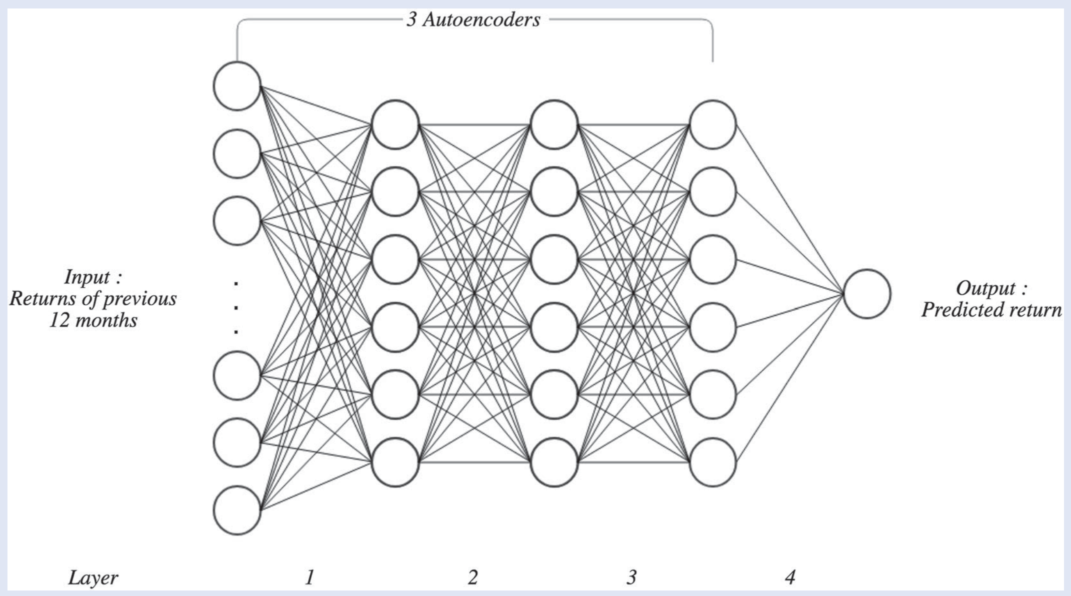

Figure 2. Deep neural network using stacked autoencoder.

other words, the input may contain noise which ideally should be removed when reconstructing the output. In Vincent *et al.* (2010), denoising autoencoder was introduced for this matter that can 'denoise' the inputs thereby providing robustness to noise-corrupted inputs. In fact, denoising autoencoder is almost identical to the ordinary autoencoder except that the original input x is corrupted to x' through some stochastic mapping before being fed as the input into the autoencoder structure shown in figure 1. Everything else remains the same, and the associated equations for the denoising autoencoder is as follows.

$$y = f(W_y x' + b_y), \tag{5}$$

$$z = f(W_z y + b_z), \tag{6}$$

where x' is a noise-corrupted version of the clean x. Comparing equations (5) and (6) with equations (3) and (4), the only difference is x in equation (3) is replaced by x' in equation (5). The goal of the denoising autoencoder is then to minimize the distance between z and the original clean x. Therefore, denoising refers to transforming a noise-corrupted input back to a noise-free one, and the goal of an ordinary autoencoder and that of a denoising autoencoder are the same: reconstruct the clean input at the output. It was claimed in Vincent *et al.* (2010), with positive experimental results and supporting geometric interpretations, that this set of transformations from x' to y to z learns better latent representation of x than that of the ordinary autoencoders.

To build a DNN using stacked denoising autoencoders (SdAE), denoising autoencoders are cascaded as before and the DNN ends with the supervised regression layer at the highest layer as in DNN using SAE. Thus the DNN using SdAE that we used in our experiments is identical to the DNN shown in figure 2 with ordinary autoencoders replaced by denoising autoencoders.

There are a couple of choices by which x can be noise-corrupted to produce x'. Two popular choices are the *masking noise* and the *additive isotropic Gaussian noise* processes. In the former, a fraction of a randomly chosen elements of the input x is set to 0, while in the latter, a zero-mean Gaussian noise of some variance is added to x. The former can be interpreted as replacing missing or unknown elements by some default value normally set to 0, while the latter is a common choice for real-valued inputs. It was shown in Vincent *et al.* (2010) that for the image classification application considered in that paper, DNN using SdAE with masking noise resulted in significant performance improvement over DNN using SAE.

In our experiments, we used the most recent 10% of training set data as the validation set to pick the best noise hyperparameter. The accruracy measure used in the validation set was the mean square error (MSE) measure just as when choosing the hyperparameters for the SVR. Choices for the noise level of the masking noise were $\{0, 0.1, 0.25, 0.4\}$ and those of the Gaussian noise were $\{0.1, 0.15, 0.3, 0.5\}$.

4. Performance results

In this section, we demonstrate the empirical results of the strategies described in this paper using the data set from Kenneth French's website (French 2018). The website contains the data set of returns of every decile portfolio for month t in which each decile portfolio is created as follows. Cumulative returns from month t-12 to month t-2 of stocks in NYSE, AMEX and NASDAQ are sorted and the corresponding stocks are placed into one of the 10 equal-weighted decile portfolios. We used monthly rebalancing and the holding periods span from January 1953 to December 2017. We used the returns of top and bottom decile portfolios to implement our strategies, and in addition to WML and WO strategies, we have used the following notations for the five selective long/short momentum strategies:

(i) SWMLa—look-back period-based selective long/short strategy

(ii) SWMLb—SVR-based selective long/short strategy
(iii) SWMLc—DNN-using SAE-based selective long/short strategy
(iv) SWMLd—DNN-using SdAE with masking noise-based selective long/short strategy
(v) SWMLe—DNN-using SdAE with Gaussian noise-based selective long/short strategy

Description of the parameter selection for the DNN architecture is in order. First, we set the number of hidden layers to three, and we used 6-6-6 hidden layers as shown in figure 2. We also tried 6-5-5, 6-5-4 versions, however 6-6-6 version produced the best results. The learning rates for the unsupervised pretraining and that for the supervised fine-tuning were both set to 0.0005. The number of pretraining epochs was set to 300, and batch size was set to 100 with three batch runs forming one epoch as the number of input vectors used to train a DNN is equal to 300. Finally, we used minmax scaling for the input to the first layer of the DNN in which

$$x \rightarrow (x - x_{\min})/(x_{\max} - x_{\min})$$

transformation is made for every input value x. Here, $x_{\min}$ and $x_{\max}$ are equal to the minimum and the maximum values, respectively, of the input values used to predict the return of current month.

All returns of the seven momentum strategies shown in this paper are net of transaction fees calculated by adopting the turnover for both the plain and the risk-managed strategies to be about 75% from Barroso and Santa-Clara (2015), and the one-way transaction cost of 0.2% and 0.1% from January 1953 to December 1997 and from January 1998 to December 2017, respectively, from Hurst *et al.* (2017). To compare the performance of the described strategies that incur transaction fees with the buy-and-hold strategy that does not incur any transaction fee except possibly the first buy trade, we have used 'Mkt' to denote this strategy which can be realized by, say, buying the index ETF. So, Mkt here represents the market of all NYSE, AMEX and NASDAQ firms as explained in and whose returns can be downloaded from French's data library (French 2018), and the returns of Mkt shown in this paper were not subject to any transaction fees accordingly.

Tables 1–4 show the performance summaries of the seven plain momentum strategies along with the buy-and-hold strategy from January 1953 to December 1977, January 1978 to December 1997, January 1998 to December 2017, and the entire period from January 1953 to December 2017, respectively. We have divided the entire period into three periods of roughly equal duration with each reflecting different economic condition. In particular, the periods get progressively more volatile with the first period from 1953 to 1977 being relatively benign without serious economic turmoils. The second period from 1978 to 1997 contained the 1987 market crash, while the last period contained the dot-com bubble burst and the financial crisis. The 10 performance measures that we have included are average monthly return, maximum monthly return, minimum monthly return, annualized return, annualized volatility, Sharpe ratio, information ratio, kurtosis, skewness, and maximum drawdown (MDD).

Table 1 indicates that in this period, SWMLe performs the best in most of the performance measures. In particular, it shows the best performance in both the plain and risk-adjusted returns, and it shows the second best kurtosis, skewness, and maximum drawdown characteristics. Tables 2–4 show that for the latter two periods and for the entire period, SWMLd produced the best performance across nearly all performance measures. Note that for these plain momentum

Table 1. Momentum strategies performance summary on S&P 500 data from 1953 to 1977.

	Mkt	WML	WO	SWMLa	SWMLb	SWMLc	SWMLd	SWMLe
Avg. Return	0.008	0.009	0.015	0.013	0.011	0.013	0.016	0.017
Return- Max	0.166	0.175	0.231	0.231	0.231	0.231	0.231	0.231
Return- Min	− 0.122	− 0.280	− 0.180	− 0.462	− 0.280	− 0.212	− 0.212	− 0.180
Ann. Return	0.092	0.102	0.174	0.136	0.116	0.142	0.183	0.206
Ann. Vol.	0.139	0.164	0.196	0.246	0.196	0.191	0.192	0.192
Sharpe ratio	0.663	0.621	0.889	0.552	0.592	0.743	0.953	1.073
Info. Ratio	6.690	0.094	1.101	0.180	0.240	0.472	0.818	1.011
Kurtosis	4.119	10.225	4.258	11.005	7.901	5.372	5.199	4.763
Skewness	− 0.014	− 1.559	− 0.086	− 1.585	− 1.077	− 0.568	− 0.408	− 0.131
MDD	0.464	0.399	0.480	0.467	0.594	0.528	0.366	0.445

Table 2. Momentum strategies performance summary on S&P 500 data from 1978 to 1997.

	Mkt	WML	WO	SWMLa	SWMLb	SWMLc	SWMLd	SWMLe
Avg. Return	0.014	0.011	0.019	0.014	0.015	0.017	0.022	0.020
Return- Max	0.129	0.126	0.170	0.167	0.170	0.170	0.170	0.170
Return- Min	− 0.226	− 0.198	− 0.326	− 0.326	− 0.261	− 0.326	− 0.261	− 0.261
Ann. Return	0.165	0.122	0.219	0.151	0.166	0.199	0.267	0.239
Ann. Vol.	0.150	0.146	0.220	0.237	0.199	0.213	0.195	0.199
Sharpe ratio	1.099	0.832	0.997	0.636	0.835	0.936	1.366	1.201
Info. Ratio	8.283	0.066	1.026	0.230	0.285	0.602	1.014	0.811
Kurtosis	6.656	6.394	7.749	7.443	6.674	8.719	6.496	6.090
Skewness	− 0.794	− 0.866	− 1.214	− 1.476	− 1.026	− 1.429	− 0.872	− 0.820
MDD	0.299	0.364	0.371	0.474	0.577	0.326	0.302	0.329

Table 3. Momentum strategies performance summary on S&P 500 data from 1998 to 2017.

	Mkt	WML	WO	SWMLa	SWMLb	SWMLc	SWMLd	SWMLe
Avg. Return	0.007	0.002	0.012	0.005	0.006	0.010	0.012	0.011
Return- Max	0.114	0.192	0.313	0.313	0.313	0.313	0.313	0.313
Return- Min	– 0.172	– 0.601	– 0.222	– 0.649	– 0.649	– 0.248	– 0.248	– 0.222
Ann. Return	0.076	– 0.019	0.119	– 0.001	0.021	0.105	0.121	0.109
Ann. Vol.	0.154	0.278	0.239	0.322	0.279	0.215	0.223	0.215
Sharpe ratio	0.491	– 0.067	0.500	– 0.004	0.074	0.490	0.544	0.507
Info. Ratio	2.820	– 0.305	0.445	– 0.232	– 0.184	0.181	0.246	0.199
Kurtosis	4.202	19.034	5.348	15.489	22.597	6.935	6.779	5.877
Skewness	– 0.712	– 2.767	– 0.129	– 1.942	– 2.373	– 0.037	0.032	0.308
MDD	0.504	0.849	0.604	0.754	0.699	0.514	0.386	0.290

Table 4. Momentum strategies performance summary on S&P 500 data from 1953 to 2017.

	Mkt	WML	WO	SWMLa	SWMLb	SWMLc	SWMLd	SWMLe
Avg. Return	0.010	0.008	0.015	0.011	0.010	0.013	0.016	0.016
Return- Max	0.166	0.192	0.313	0.313	0.313	0.313	0.313	0.313
Return- Min	– 0.226	– 0.601	– 0.326	– 0.649	– 0.649	– 0.326	– 0.261	– 0.261
Ann. Return	0.109	0.069	0.170	0.096	0.101	0.148	0.188	0.185
Ann. Vol.	0.147	0.202	0.217	0.270	0.226	0.206	0.203	0.202
Sharpe ratio	0.738	0.343	0.784	0.357	0.445	0.719	0.926	0.917
Info. Ratio	4.637	– 0.084	0.830	0.025	0.073	0.404	0.643	0.636
Kurtosis	4.969	23.891	5.905	14.253	19.480	7.059	6.233	5.513
Skewness	– 0.509	– 2.739	– 0.472	– 1.836	– 1.911	– 0.674	– 0.376	– 0.185
MDD	0.504	0.849	0.604	0.754	0.699	0.528	0.386	0.445

strategies, WO performs much better than Mkt which in turn outperforms WML. However, SWMLd outperforms WO in all of the performance measures but kurtosis while the difference was trivial. Also note that the SVR-based SWMLb does not perform clearly better than the look-back period-based SWMLa, however, the DNN-based strategies SWMLc, SWMLd, SWMLe show performance that dwarfs that of both SWMLa and SWMLb. Furthermore, as expected from the definition of SdAE using masking noise, SWMLd outperforms SWMLc across all performance measures as SWMLc is a special case of SWMLd with noise level set to 0.

In figures 3–6, we show the cumulative return curves for the three periods and the entire period, respectively, for the seven plain momentum strategies along with the buy-and-hold strategy. As can be expected from tables 1–4, SWMLe shows the best return performance in figure 3, while in all of the other figures, SWMLd shows the best performance. The figures show that SWMLd outperforms SWMLe which in turn outperforms WO by a larger gap, in general. Alternatively, tables 1–4 and figures 3–6 show that both versions of DNN using noise-added stacked autoencoders perform strictly better than the noise-‘unadded’ counterpart which in turn shows strictly better performance than the best shallow learning-based method, namely, SWMLb, and the look-back period-based method SWMLa. For these plain strategies, WML gave the worst performance among the eight strategies considered.

In table 5, we list the performance differential of the annualized return and the Sharpe ratio between the best performing selective WML strategy, namely, the SWMLd strategy, and the buy-and-hold strategy Mkt and the two baseline momentum strategies WML and WO. Across all periods of differing economic conditions, SWMLd showed consistent performance improvement over the baseline strategies.

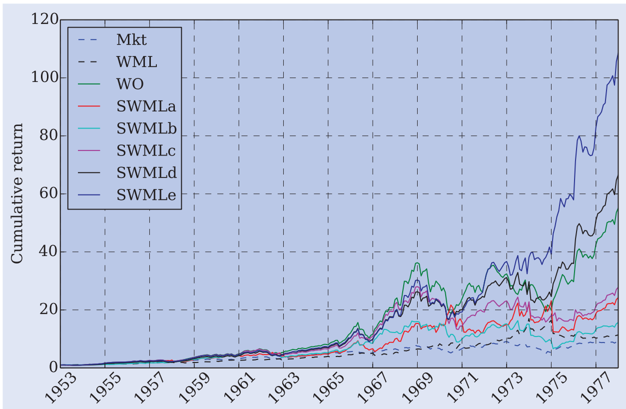

Figure 3. Cumulative return curves for plain strategies during 1953–1977 period.

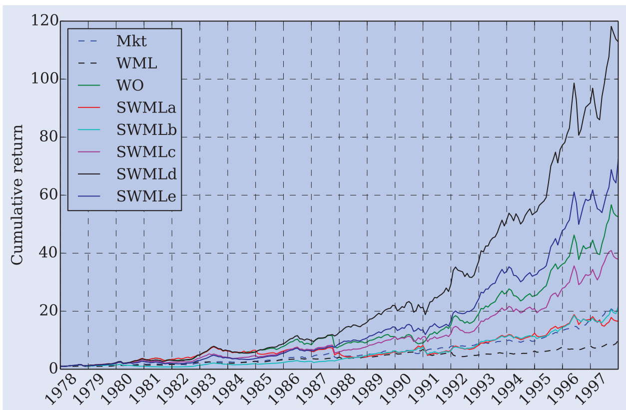

Figure 4. Cumulative return curves for plain strategies during 1978–1997 period.

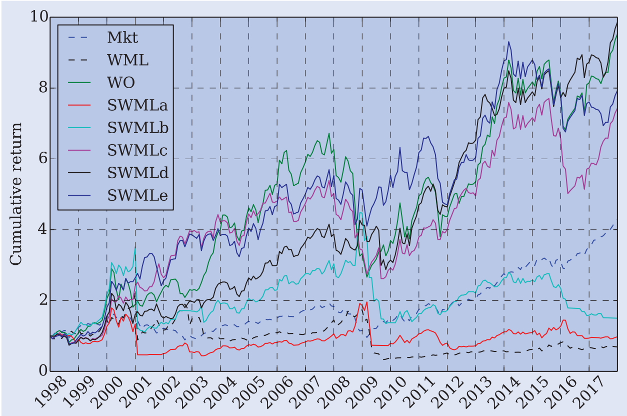

Figure 5. Cumulative return curves for plain strategies during 1998–2017 period.

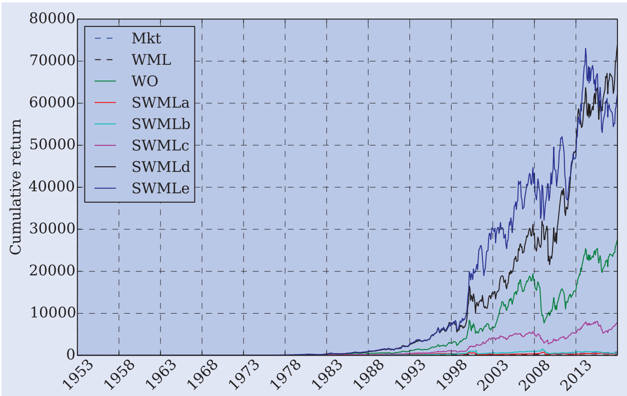

Figure 6. Cumulative return curves for plain strategies during 1953–2017 period.

In tables 6–9, we show the performance summaries of the seven volatility-scaled momentum strategies along with the buy-and-hold strategy from the same periods as in tables 1–4, respectively. The first finding to notice from the tables is that now WML, in general, outperforms WO which in turn outperforms Mkt. Not only have the returns improved for the WML strategy relative to the WO strategy, the maximum drawdown and the higher-order moments of kurtosis and skewness improved significantly for the WML strategy in these tables. Therefore, the fat left tail risk implied by high kurtosis and high left skewness is significantly reduced for the volatility-scaled WML strategy. On the other hand, consistent as in the case of plain momentum strategies, both types of DNN using noise-added stacked autoencoder-based strategies clearly outperform both WML and WO strategies. In particular, for the volatility-scaled versions, the outperformance of our proposed strategies is more pronounced in magnitude and more complete in the sense of being better in all of the 10 performance measures. So in summary, not only do our proposed selective WML strategies SWMLd and SWMLe outperform both of the baseline strategies WML and WO in terms of various forms of risk, but they also offer superior higher-order moments and maximum drawdown characteristics. Note especially that SWMLd showed maximum drawdown of at most approximately 0.3 in every single period we tested compared to the much improved WML whose corresponding measure was shown to be higher than 0.4 in some periods. Tables indicate that SWMLd is generally superior to SWMLe in all periods except the first one in most of the performance measures. In the first period, SWMLe slightly outperforms in the return measures but underperforms in higher-order moments and maximum drawdown measures compared to SWMLd. For all periods and performance measures, SWMLd and SWMLe

Table 5. Performance differential.

	SWMLd - Mkt	SWMLd - WML	SWMLd - WO
Annualized return			
1953–1977	0.091	0.081	0.009
1978–1997	0.102	0.145	0.048
1998–2017	0.045	0.140	0.002
1953–2017 (entire period)	0.079	0.119	0.018
Sharpe ratio			
1953–1977	0.290	0.332	0.064
1978–1997	0.267	0.534	0.369
1998–2017	0.053	0.611	0.044
1953–2017 (entire period)	0.188	0.583	0.142

Table 6. Volatility-scaled momentum strategies performance summary on S&P 500 data from 1953 to 1977.

	Mkt	WML	WO	SWMLa	SWMLb	SWMLc	SWMLd	SWMLe
Avg. Return	0.008	0.016	0.014	0.016	0.013	0.016	0.017	0.018
Return- Max	0.166	0.189	0.165	0.186	0.165	0.312	0.312	0.312
Return- Min	− 0.122	− 0.266	− 0.168	− 0.238	− 0.266	− 0.224	− 0.224	− 0.224
Ann. Return	0.092	0.181	0.163	0.181	0.148	0.180	0.200	0.221
Ann. Vol.	0.139	0.194	0.177	0.191	0.193	0.198	0.196	0.200
Sharpe ratio	0.663	0.933	0.920	0.950	0.765	0.907	1.018	1.105
Info. Ratio	6.690	0.369	1.023	0.445	0.378	0.675	0.897	1.028
Kurtosis	4.119	6.640	3.854	5.060	5.795	6.780	6.751	6.969
Skewness	− 0.014	− 0.976	− 0.312	− 0.605	− 0.780	0.051	0.040	0.216
MDD	0.464	0.369	0.480	0.278	0.440	0.477	0.344	0.422

Table 7. Volatility-scaled momentum strategies performance summary on S&P 500 data from 1978 to 1997.

	Mkt	WML	WO	SWMLa	SWMLb	SWMLc	SWMLd	SWMLe
Avg. Return	0.014	0.021	0.019	0.018	0.022	0.022	0.025	0.024
Return- Max	0.129	0.183	0.217	0.183	0.217	0.217	0.217	0.217
Return- Min	− 0.226	− 0.230	− 0.388	− 0.388	− 0.285	− 0.388	− 0.285	− 0.285
Ann. Return	0.165	0.261	0.218	0.195	0.270	0.261	0.309	0.298
Ann. Vol.	0.150	0.208	0.232	0.242	0.225	0.242	0.214	0.224
Sharpe ratio	1.099	1.256	0.939	0.803	1.198	1.079	1.441	1.334
Info. Ratio	8.283	0.617	0.942	0.433	0.663	0.837	1.144	1.041
Kurtosis	6.656	4.936	9.456	8.350	5.390	8.793	5.842	5.258
Skewness	− 0.794	− 0.465	− 1.289	− 1.355	− 0.634	− 1.208	− 0.522	− 0.426
MDD	0.299	0.319	0.407	0.487	0.500	0.388	0.317	0.377

Table 8. Volatility-scaled momentum strategies performance summary on S&P 500 data from 1998 to 2017.

	Mkt	WML	WO	SWMLa	SWMLb	SWMLc	SWMLd	SWMLe
Avg. Return	0.007	0.007	0.008	0.007	0.007	0.007	0.008	0.006
Return- Max	0.114	0.137	0.186	0.186	0.186	0.212	0.212	0.212
Return- Min	− 0.172	− 0.355	− 0.223	− 0.259	− 0.259	− 0.223	− 0.223	− 0.223
Ann. Return	0.076	0.064	0.084	0.068	0.076	0.074	0.089	0.062
Ann. Vol.	0.154	0.182	0.148	0.166	0.151	0.148	0.147	0.152
Sharpe ratio	0.491	0.349	0.572	0.411	0.501	0.500	0.605	0.409
Info. Ratio	2.820	− 0.086	0.180	− 0.055	− 0.012	0.017	0.103	− 0.050
Kurtosis	4.202	12.125	7.513	8.014	9.996	9.189	9.517	8.693
Skewness	− 0.712	− 1.609	− 0.388	− 0.568	− 0.528	0.076	0.313	0.186
MDD	0.504	0.448	0.319	0.311	0.310	0.291	0.297	0.259

Table 9. Volatility-scaled momentum strategies performance summary on S&P 500 data from 1953 to 2017.

	Mkt	WML	WO	SWMLa	SWMLb	SWMLc	SWMLd	SWMLe
Avg. Return	0.010	0.015	0.014	0.013	0.014	0.015	0.017	0.016
Return- Max	0.166	0.189	0.217	0.186	0.217	0.312	0.312	0.312
Return- Min	− 0.226	− 0.355	− 0.388	− 0.388	− 0.285	− 0.388	− 0.285	− 0.285
Ann. Return	0.109	0.167	0.155	0.149	0.161	0.170	0.196	0.192
Ann. Vol.	0.147	0.196	0.189	0.202	0.194	0.201	0.190	0.196
Sharpe ratio	0.738	0.852	0.819	0.739	0.830	0.846	1.032	0.979
Info. Ratio	4.637	0.284	0.730	0.281	0.349	0.526	0.713	0.675
Kurtosis	4.969	7.350	8.456	7.838	6.359	8.748	6.826	6.549
Skewness	− 0.509	− 0.903	− 0.811	− 0.954	− 0.591	− 0.512	− 0.074	0.025
MDD	0.504	0.448	0.480	0.487	0.500	0.477	0.344	0.422

were the best and superior to SWMLc which was in turn superior to the shallow learning-based SWMLb. SWMLb was in more cases superior to the look-back period-based SWMLa. Finally, note from tables 1–4 and 6–9 that SWMLa performs inferior to one or both of the two baseline strategies in nearly all performance measures for every period considered. Therefore, it was crucial that we use deep learning-based prediction of returns in order to provide improvement, with respect to the various performance measures, to the baseline strategies.

In figures 7–10, we show the cumulative return curves for the three periods and the entire period, respectively, for the seven volatility-scaled momentum strategies along with the buy-and-hold strategy. In agreement with the results indicated in the tables, in all figures except the first one, SWMLd shows the clear best performance and the second best performance in the first figure. Thus our strategy SWMLd outperforms other selective WML strategies, namely, SWMLa, SWMLb and SWMLc, and two baseline strategies WML and WO in every single period tested. In these figures, the buy-and-hold strategy performed the worst in every single period followed by WO and SWMLa.

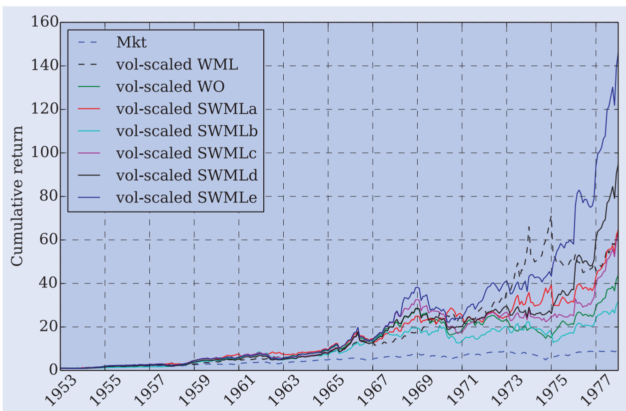

Figure 7. Cumulative return curves for volatility-scaled strategies during 1953–1977 period.

In table 10, we list the performance differential of the annualized return and the Sharpe ratio between the volatility-scaled versions of SWMLd, and Mkt, WML and WO, analogous to table 5. The table clearly shows that SWMLd

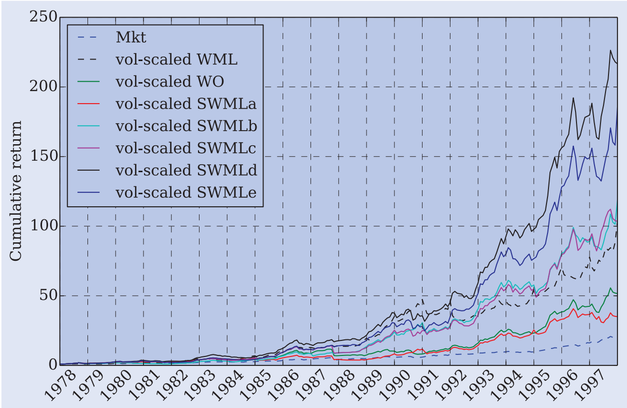

Figure 8. Cumulative return curves for volatility-scaled strategies during 1978–1997 period.

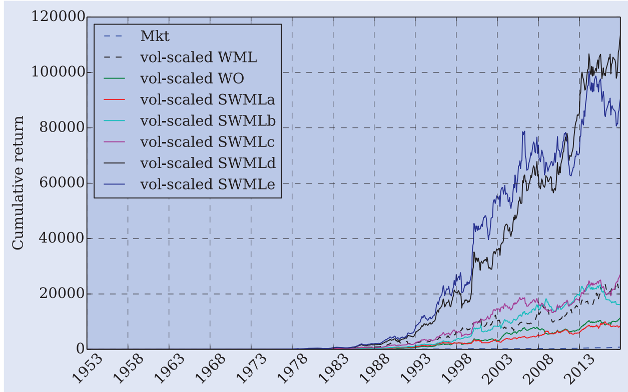

Figure 10. cumulative return curves for volatility-scaled strategies during 1953–2017 period.

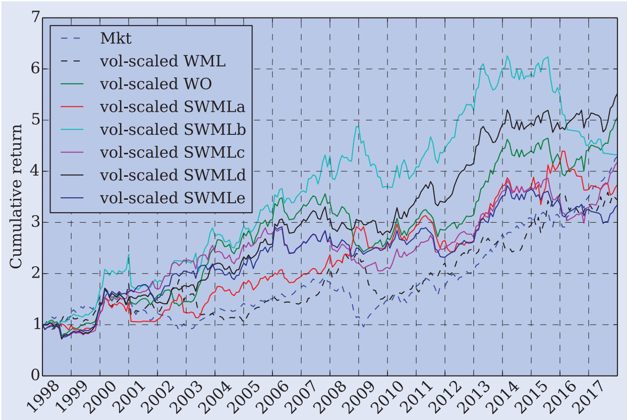

Figure 9. cumulative return curves for volatility-scaled strategies during 1998–2017 period.

is superior to the buy-and-hold strategy and both baseline momentum strategies in every single period we tested. Furthermore, tables 5 and 10 show that the performance differential between SWMLd and the better of the two baseline strategies, i.e. WO for the plain strategy and WML for the volatility-scaled strategy, is greater for the volatility-scaled version. In other words, while all strategies showed performance improvement through volatility-scaling, performance improvement of SWMLd was larger than that of WO or WML.

Finally, in tables 11–14, we exhibit the statistics of the long and short trades executed in the selective WML strategies. The tables show the fraction of times long and short positions are traded under the five selective WML strategies across the three periods and the entire period as before. SWMLd and SWMLe tend to pick long only trade more often than SWMLc which in turn tend to pick long only trade more often than SWMLa and SWMLb. The order of the frequencies of the trades taken was identical across all strategies and periods, specifically, long only first, long and short second, short only third, and no trade last. We note that when no trade is picked, one can allocate funds to Treasury bonds which can increase the return rates observed for the selective WML strategies in tables 1–4 and tables 6–9. These tables in current form were produced by equating no trade over a given holding period to zero return minus the transaction fee for that period. We also note that the fraction of times a particular trade is selected does not depend on whether the strategy is plain or volatility-scaled type, and hence the statistics shown reflect that of both types of strategies.

5. Conclusion

In this paper, we considered the portfolio strategy that is based on the premier anomaly in financial economics. We have proposed a variation of this strategy, namely, the selective momentum strategy and have investigated how different

Table 10. Performance differential.

	SWMLd - Mkt	SWMLd - WML	SWMLd - WO
Annualized return			
1953–1977	0.108	0.019	0.037
1978–1997	0.144	0.048	0.091
1998–2017	0.013	0.025	0.005
1953–2017 (entire period)	0.087	0.029	0.041
Sharpe ratio			
1953–1977	0.355	0.085	0.098
1978–1997	0.342	0.185	0.502
1998–2017	0.114	0.256	0.033
1953–2017 (entire period)	0.294	0.180	0.213

Table 11. Statistics of long/short in selective momentum strategies from 1953 to 1977.

	SWMLa	SWMLb	SWMLc	SWMLd	SWMLe
Long and short	0.183	0.253	0.177	0.127	0.137
Long only	0.627	0.610	0.780	0.853	0.833
Short only	0.177	0.073	0.020	0.013	0.023
No trade	0.013	0.063	0.023	0.007	0.007

Table 12. Statistics of long/short in selective momentum strategies from 1978 to 1997.

	SWMLa	SWMLb	SWMLc	SWMLd	SWMLe
Long and short	0.125	0.296	0.283	0.237	0.221
Long only	0.704	0.537	0.662	0.738	0.746
Short only	0.163	0.108	0.029	0.021	0.021
No trade	0.008	0.058	0.025	0.004	0.013

Table 13. Statistics of long/short in selective momentum strategies from 1998 to 2017.

	SWMLa	SWMLb	SWMLc	SWMLd	SWMLe
Long and short	0.125	0.237	0.267	0.237	0.217
Long only	0.613	0.537	0.688	0.733	0.746
Short only	0.179	0.096	0.013	0.017	0.017
No trade	0.083	0.129	0.033	0.013	0.021

Table 14. Statistics of long/short in selective momentum strategies from 1953 to 2017.

	SWMLa	SWMLb	SWMLc	SWMLd	SWMLe
Long and short	0.147	0.262	0.237	0.195	0.187
Long only	0.646	0.565	0.715	0.781	0.779
Short only	0.173	0.091	0.021	0.017	0.021
No trade	0.033	0.082	0.027	0.008	0.013

versions of this variation compare with the well-known existing momentum strategies. The integral part of our selective momentum strategy is how the return prediction method is defined. For this matter, we have investigated the efficacy of a number of competitive techniques from machine learning in our problem setting. The techniques included support vector regression, and deep neural networks using stacked autoencoders and denoising autoencoders.

Our empirical finding has shown that versions of our selective momentum strategy can outperform the benchmark existing momentum strategies persistently across different time periods of varying economic conditions. Our selective momentum strategies were also able to produce superior performance in both plain and risk-managed environments. Specifically, our proposed strategies exhibited performance that was superior to that of the benchmark strategies with respect to various forms of returns, higher-order moments, and maximum drawdown measures. Perhaps more importantly, the versions of our selective mommentum strategy with the described performance advantages across diverse set of experiment settings were ones defined by the latest deep learning technologies, namely, deep neural network using stacked denoising autoencoders. In contrast, the look-back period-based selective momentum strategy did not perform competitively to the benchmark strategies. In the future, we plan to incorporate generative models into our regression framework for improved return prediction.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1D1A1B03032722).

References

Asness, C.S., Moskowitz, T.J. and Pederson, L.H., Value and momentum everywhere. *J. Finance*, 2013, **68**(4), 929–985.

Bao, W., Yue, J. and Rao, Y., A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLoS ONE*, 2017, **12**(7), e0180944.

Barroso, P. and Santa-Clara, P., Momentum has its moments. *J. Financ. Econ.*, 2015, **116**, 111–120.

Bengio, Y., Lamblin, P., Popovici, D. and Larochelle, H., Greedy layer-wise training of deep networks. In *Neural Information Processing Systems*, Vancouver, Canada, 2006, 153–160.

Berner, J., Grohs, P. and Jentzen, A., Analysis of generalization error: Empirical risk minimization over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of Black–Scholes partial differential equations, 2018. Available online at: arXiv:1809.03062 (accessed 15 October 2018).

Chen, Y., Lin, Z., Zhao, X., Wang, G. and Gu, Y., Deep learning-baesd classification of hyperspectral data. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, 2014, **7**(6), 2094–2107.

Daniel, K. and Moskowitz, T.J., Momentum Crashes. Working Paper, Columbia University, New York, 2013.

Dixon, M., Polson, N.G. and Sokolov, V.O., Deep learning for spatio-temporal modeling: Dynamic traffic flows and high frequency trading, 2018. Available online at: arXiv:1705.09851 (accessed 15 October 2018).

Fama, E.F. and French, K.R., Multifactor explanations of asset pricing anomalies. *J. Finance*, 1996, **51**, 55–84.

Fama, E.F. and French, K.R., Size, value, and momentum in international stock returns. *J. Financ. Econ.*, 2012, **105**, 457–472.

French, K.R., 2018. Available online at: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html (accessed 28 March 2018).

Grundy, B. and Martin, J.S., Understanding the nature of the risks and the source of the rewards to momentum investing. *Rev. Financ. Stud.*, 2001, **14**, 29–78.

Heaton, J.B., Polson, N.G. and Witte, J.H., Deep learning in finance, 2016. Available online at: arXiv:1602.06561v2 (accessed 28 March 2018).

Hurst, B., Ooi, Y.H. and Pedersen, L.H., A century of evidence on trend-following investing. *J. Portfolio Manage.*, 2017, **44**(1), Fall.

Jegadeesh, N. and Titman, S., Returns to buying winners and selling losers: Implications for stock market efficiency. *J. Finance*, 1993, **48**, 65–91.

Jegadeesh, N. and Titman, S., Profitability of momentum strategies: An evaluation of alternative explanations. *J. Finance*, 2001, **56**, 699–720.

Kim, S., Volatility forecasting for low-volatility portfolio selection in the US and the Korean equity markets. *J. Exp. Theor. Artif. Int.*, 2018, **30**(1), 71–88.

Kim, S. and Heo, J., Time series regression-based pairs trading in the Korean equities market. *J. Exp. Theor. Artif. Int.*, 2017, **29**(4), 755–768.

Law, T. and Shawe-Taylor, J., Practical Bayesian support vector regression for financial time series prediction and market condition change detection. *Quant. Finance*, 2017, **17**(9), 1403–1416.

Moskowitz, T.J., Ooi, Y.H. and Pedersen, L.H., Time series momentum. *J. Financ. Econ.*, 2012, **104**, 228–250.

Sapankevych, N.I. and Sankar, R., Time series prediction using support vector machines: A survey. *IEEE Comput. Intell. Mag.*, 2009, **4**(2), 24–38.

Sirignano, J. and Cont, R., Universal features of price formation in financial markets: Perspectives from deep learning. *SSRN*, 2018, 3141294.

Sirignano, J., Sadhwani, A. and Giesecke, K., Deep learning for mortgage risk, 2018. Available online at: arXiv:1607.02470 (accessed 15 October 2018).

Vincent, P., Larochelle, H., Lajole, I., Bengio, Y. and Manzagol, P.-A., Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.*, 2010, **11**, 3371–3408.