Improving Hedge Fund Return Prediction: Dealing with Missing Data via Deep Learning

March 15, 2026

Abstract

We study the important issue of handling missing values in hedge fund data. We employ a deep learning approach, BRITS, for filling in missing values for both hedge fund returns and fund characteristics used to predict returns. BRITS harnesses information from past and future observations in the time-series dimension as well as contemporaneous observations in the cross-sectional dimension. Compared to alternative data imputation methods, BRITS provides high fidelity for data imputation in experiments. Moreover, BRITS substantially improves out-of-sample hedge fund return prediction with neural networks in terms of filling in missing values for the training samples. BRITS appears especially valuable for generating hedge fund return predictions to guide the dynamic selection of top-performing hedge funds.

JEL classifications: G12, G14, G17, G23

Keywords: Hedge funds, Deep Learning, Missing values, Data imputation, BRITS, Bias-variance trade-off

Improving Hedge Fund Return Prediction: Dealing with Missing Data via Deep Learning

Abstract

We study the important issue of handling missing values in hedge fund data. We employ a deep learning approach, BRITS, for filling in missing values for both hedge fund returns and fund characteristics used to predict returns. BRITS harnesses information from past and future observations in the time-series dimension as well as contemporaneous observations in the cross-sectional dimension. Compared to alternative data imputation methods, BRITS provides high fidelity for data imputation in experiments. Moreover, BRITS substantially improves out-of-sample hedge fund return prediction with neural networks in terms of filling in missing values for the training samples. BRITS appears especially valuable for generating hedge fund return predictions to guide the dynamic selection of top-performing hedge funds.

JEL classifications: G12, G14, G17, G23

Keywords: Hedge funds, Deep Learning, Missing values, Data imputation, BRITS, Bias-variance trade-off

1. Introduction

A common problem in financial datasets is the presence of missing values. Missing values present keen challenges for important topics in empirical asset pricing, including model estimation, out-of-sample return forecasting, and asset allocation. Thus, the effective imputation of missing data has the potential to substantially improve empirical asset pricing. Despite its importance, the issue of missing values and data imputation in finance has only recently begun to receive significant attention in the literature (e.g., Giglio, Liao, and Xiu 2021; Beckmeyer and Wiedemann 2025; Chen and McCoy 2024; Bryzgalova et al. 2025; Freyberger et al. 2025).

In this paper, we examine the imputation of missing values in hedge fund data, a critical issue arising from the voluntary nature of hedge fund disclosures. Missing values are a notorious issue for the empirical analysis of hedge funds, which have substantial flexibility and face limited regulations when it comes to data reporting, rendering hedge fund data disclosures effectively voluntary. The voluntary nature of hedge fund disclosures gives rise to pervasive missing values in hedge fund returns as well as the fund characteristics that are used to predict their returns. The surfeit of missing data can adversely affect the empirical analysis of the important topics mentioned above when it comes to hedge funds, such as return prediction and fund selection from an asset allocation perspective. Furthermore, the diverse strategies and proprietary approaches employed by hedge funds make it inherently difficult to identify true managerial skill, exacerbating the challenges posed by missing data.

We consider monthly data for 3,800 hedge funds from 1994:01 to 2021:12. Our dataset consists of hedge fund returns together with 23 fund characteristics that are potentially relevant for predicting future hedge fund returns. Among the characteristics that we consider are past returns, return autocorrelations, higher return moments, and various measures of fund managerial skill. After making standard adjustments to the hedge fund return data and accounting for the availability of fund characteristic data, the sample period spans 1998:01 to 2021:12.

We can view filling in missing values for hedge fund return and characteristic data as performing matrix completion. Consider an $N \times T$ matrix $\boldsymbol{X}$ of observations for hedge fund returns or a fund characteristic, where each column contains the N-dimensional vector of fund observations for period t. For hedge fund data, $\boldsymbol{X}$ often contains a (very) large number of missing observations, so the matrix can be quite sparse. The aim is to fill in the missing values in $\boldsymbol{X}$ using the information in the non-missing values. To accomplish this, we employ a deep learning model that utilizes both time-series and cross-sectional information for matrix completion. Specifically, we use the bidirectional recurrent imputation network for time series (BRITS, Cao et al. 2018). A number of properties make BRITS attractive for filling in missing values for hedge fund data. First, when training the BRITS network, information from both past and future values of a variable is used to impute missing values. This bidirectional flow provides a rich information set that is especially useful in sparse datasets with relatively large numbers of missing values; furthermore, it helps the model achieve the same quality of imputation at the beginning and the end of a variable's time series. Second, the BRITS network harnesses cross-sectional information from non-missing values when filling in missing values, which aligns with the substantive contemporaneous correlations across many financial variables. Third, BRITS does not make specific assumptions about the data-generating process, so we do not need to make strict assumptions about the data. Overall, BRITS has a number of features that make it a promising strategy for imputing missing values for hedge fund data.

We consider the cross-sectional mean (CS-Mean) and singular value thresholding (SVT, Cai, Candès, and Shen 2010) as alternative data imputation methods. We include the CS-Mean because it is a popular and simple method for matrix completion. It entails filling in the missing values for a column of $\boldsymbol{X}$ with the mean of the non-missing values in the column. SVT is a well-known and powerful matrix completion algorithm that is designed to handle very sparse matrices. SVT assumes that the data are characterized by low-rankness, so it makes stronger assumptions about the data-generating process than BRITS. The best

method for data imputation for out-of-sample hedge fund return prediction is ultimately an empirical issue, essentially determined by the bias-variance trade-off. Because the CS-Mean is the simplest method, it is likely to have a relatively large bias but small variance compared to the more complex SVT and BRITS methods. Because of the low-rankness assumption, we expect SVT to have a larger bias but a smaller variance than BRITS. Due to its complex network structure, BRITS should have the lowest bias but the largest variance, which increases the risk of overfitting. We undertake an extensive analysis of out-of-sample hedge fund return prediction to examine which matrix completion method provides the best performance in light of the bias-variance trade-off.

As a preliminary exercise, we conduct an experiment to investigate the accuracy of CS-Mean, SVT, and BRITS for directly imputing missing values for hedge fund returns and characteristics. In the exercise, we randomly drop 10% or 20% of non-missing values for a matrix of hedge fund return or characteristic observations and fill in the pseudo missing values using one of the imputation methods. For the pseudo missing values, we use the root mean squared error (RMSE) to measure the accuracy of the imputed values in tracking the actual values. For hedge fund returns and the clear majority of fund characteristics, BRITS provides the most accurate imputations in terms of RMSE. SVT provides the most accurate imputations for the other fund characteristics and performs comparably to BRITS for a number of characteristics. In general, the results suggest that BRITS provides the greatest fidelity for imputing missing values for hedge fund data, followed by SVT.

Next, we assess the performance of CS-Mean, SVT, and BRITS in the context of out-of-sample hedge fund return prediction. We predict hedge fund returns using a large set of predictors comprised of the 23 fund characteristics and their interactions with four economic variables (for a total of 115 predictors). The economic variables include equity market and economic policy uncertainty indices (Baker, Bloom, and Davis 2016), the TED spread, and the VIX. We generate a sequence of monthly hedge fund return forecasts for all funds with non-missing values for return observations in a given month using a rolling estimation window

of 180 months (15 years). Specifically, to generate return forecasts for month $t+1$, we use available data for hedge fund returns and the 23 fund characteristics as well as the economic variables for months $t-179$ to t. We apply an imputation method to fill in missing values for the hedge fund return and characteristic data.[1] We then use the complete set of non-missing and filled-in observations for the returns and characteristics (as well as the economic variables, which do not have missing values) for months $t-179$ to t to train a prediction model and generate a set of return forecasts for month $t+1$. Using an initial in-sample period of 1998:01 to 2012:12, we proceed sequentially in this manner to generate a series of monthly out-of-sample hedge fund return predictions for 2013:01 to 2021:12.

As prediction models for hedge fund returns, we employ three deep neural networks as well as an ensemble of the neural networks.[2] We assess the statistical accuracy of the deep neural network return predictions for the different imputation methods as well as a no-imputation approach that only includes data for hedge funds with no missing values for any variable when training the prediction model (i.e., "complete cases"). The results show that more complex data imputation methods improve the statistical accuracy of out-of-sample hedge fund return predictions. Forgoing data imputation or using the simplest CS-Mean approach yields insignificant improvements in forecasting accuracy, while the more complex SVT and BRITS imputation methods generate larger and significant improvements in forecasting accuracy. With respect to predicting hedge fund returns with neural networks, the increased variance associated with more complex data imputation methods does not adversely affect performance in light of the bias-variance trade-off.

We further evaluate the out-of-sample deep neural network return forecasts based on the different imputation methods in terms of economic value. First, we form a long-only top-decile hedge fund portfolio by sorting funds in terms of their return predictions. Specifically,

[1] We train the BRITS networks that we use to fill in the missing values using data for months $t-179$ to t, so there is no look-ahead bias when imputing the missing values.

[2] We also use linear penalized regressions, such as the seminal least absolute shrinkage and selection operator (LASSO, Tibshirani 1996) as well as three variants: elastic net (Zou and Hastie 2005), adaptive LASSO (Zou 2006), and sparse group LASSO (Simon et al. 2013). The results are consistent with those for the neural networks and are reported in the Internet Appendix.

each month we sort hedge funds with non-missing return observations in terms of their return forecasts for the next month and go long the 10% of funds with the highest predicted returns. We consider equal, rank, and prediction weighting of the funds in the top decile, where the latter two strategies place greater weight on funds with higher predicted returns within the top decile. Particularly with respect to rank and prediction weighting, we find that portfolio performance improves when we use the more complex data imputation methods. BRITS provides the best overall performance, delivering significant and large alphas for both the Fung and Hsieh (2004) seven-factor and Chen et al. (2025) nine-factor models of hedge fund risks as well as sizable annualized Sharpe ratios well above one. We also construct long-short portfolios that take long (short) positions in the top (bottom) decile of hedge funds sorted based on their return forecasts for the next month. The results are similar to those for the long-only top-decile portfolios, with even starker improvements in performance when we employ the more complex imputation methods as well as rank and prediction weighting. For the long-short portfolios, BRITS again delivers the best overall performance.

In sum, we make two key contributions to the hedge fund literature. First, when it comes to out-of-sample hedge fund return prediction, in terms of both statistical accuracy and economic value, we show that it pays to rely on more complex data imputation methods to deal with the many missing values that plague hedge fund data. Especially when it comes to economic value, BRITS is an effective deep learning tool for imputing missing values. Because missing values are a pervasive problem in hedge fund data, this makes BRITS a valuable resource for empirical research on hedge funds. The efficacy of BRITS in filling in missing values stems from its ability to harness information along both the time-series and cross-sectional dimensions; it also accommodates general data processes, so it does not require strong assumptions regarding the underlying data. BRITS is based on a complex recurrent neural network structure, which increases its variance and thus the risk of overfitting. However, our results indicate that BRITS does not lead to substantive overfitting

when it comes to data imputation for hedge fund return prediction, so its complexity is justified in light of the bias-variance trade-off.

Second, we provide the most comprehensive analysis to date on out-of-sample hedge fund return predictability. Our analysis incorporates a large number of hedge fund return predictors, including numerous fund characteristics and their interactions with a set of economic variables, as well as an array of neural networks.[3] When we impute missing values for hedge fund data via BRITS, we find that the combination of a large number of predictors and machine learning methods generates significant gains in out-of-sample hedge fund return predictability both statistically and economically. BRITS-based neural network hedge fund returns predictions appear especially valuable for guiding the dynamic selection of top-performing hedge funds.

The issue of missing values has been most prominent in regard to cross-sectional stock return prediction based on large sets of firm characteristics, with hundreds of firm characteristics having been studied in the literature (e.g., Hou, Xue, and Zhang 2015; Chen and Zimmermann 2022). The most popular approach for filling in missing values for firm characteristics is the CS-Mean (e.g., Haugen and Baker 1996; Green, Hand, and Zhang 2017; Light, Maslov, and Rytchkov 2017; Gu, Kelly, and Xiu 2020; Han et al. 2024). Another frequently used option is to omit firms with missing characteristic values (e.g., Freyberger, Neuhierl, and Weber 2020). Chen and McCoy (2024) find that CS-Mean imputation performs well for imputing missing firm characteristics relative to more sophisticated expectation-maximization methods, given the nature of missing data for firm characteristics and the risk of overfitting when relying on more sophisticated methods. With respect to filling in missing data for predicting cross-sectional hedge fund returns, Wu et al. (2021) rely on the CS-Mean approach. In a different context, Giglio, Liao, and Xiu (2021) employ the SVT algorithm to impute missing values for the residuals of hedge fund returns in factor models for identifying funds with significant alphas while controlling the false discovery rate. In the present paper, we

[3]Existing studies that investigate out-of-sample hedge fund return predictability include Avramov et al. (2011), Avramov, Barras, and Kosowski (2013), and Wu et al. (2021).

find that imputing missing data via BRITS to predict hedge fund returns delivers substantial out-of-sample performance gains.

The rest of the paper is organized as follows. Section 2 describes the data. Section 3 outlines the BRITS network for imputing missing values. Section 4 presents an experiment to analyze imputation accuracy. Section 5 reports the main empirical results on out-of-sample hedge fund return prediction using different imputation methods. Section 6 concludes.

2. Data

We use hedge fund data from the Lipper Trading Advisor Selection System (TASS) database for 1994:01 to 2021:12. Following the standard procedure in the hedge fund literature, we apply a set of filters to the data before using it for our analysis (e.g., Fung and Hsieh 2000; Aragon 2007; Bali, Brown, and Caglayan 2012; Bali et al. 2021; Chen, Han, and Pan 2021; Wu et al. 2021). We follow Chen, Han, and Pan (2021) and consider only US-oriented hedge funds (to avoid fund duplicates in different currencies) as well as funds that report net-of-fee returns. In terms of filters, we preclude survivorship bias by including both live and defunct funds. To account for backfill bias, we delete each fund's first twelve months of returns. In addition, to remove multi-period sampling bias, we require each fund to have at least 30 return observations.[4] Finally, we discard funds with assets under management below $5 million, so we focus on large funds that are more relevant to investors and less likely to manipulate reporting to TASS. After applying the filters, we have 3,800 hedge funds for our analysis.

For hedge fund return predictors, we consider 23 fund characteristics, many of which have been found to predict fund returns (e.g., Titman and Tiu 2011; Bali, Brown, and Caglayan 2012, 2019; Heuson, Hutchinson, and Kumar 2020; Wu et al. 2021). We group the characteristics into four categories in Table 1. The first category (Panel A) consists of lagged

[4]For robustness purposes, we follow Giglio, Liao, and Xiu (2021) and require each fund to have at least 60 return observations across the full sample period for the imputation experiment in Section 4.

cumulative returns ranging from one to 36 months, which are designed to capture short-, medium-, and long-term momentum effects. The second category (Panel B) is comprised of the first- through third-order autocorrelations for a fund's return over the last twelve months, which provide additional measures of persistence in hedge fund returns.

Table 1. Hedge fund return predictors

The table provides hedge fund characteristics and economic variables that are used to predict hedge fund returns.

(1) Variable	(2) Abbreviation
Panel A: Lagged cumulative returns	
1-month return	`Ret_1mo`
3-month cumulative return	`CRet_3mo`
6-month cumulative return	`CRet_6mo`
9-month cumulative return	`CRet_9mo`
12-month cumulative return	`CRet_12mo`
36-month cumulative return	`CRet_36mo`
Panel B: Lagged return autocorrelations	
Lag 1 autocorrelation for returns over last 12 months	`AC_Lag1`
Lag 2 autocorrelation for returns over last 12 months	`AC_Lag2`
Lag 3 autocorrelation for returns over last 12 months	`AC_Lag3`
Panel C: Return moments	
Volatility over last 36 months	`Vol`
Idiosyncratic volatility over last 36 months	`IdioVol`
Systematic volatility over last 36 months	`SysVol`
Coskewness over last 36 months	`CoSkew`
Idiosyncratic skewness over last 36 months	`IdioSkew`
Skewness over last 36 months	`Skew`
Kurtosis over last 36 months	`Kurt`

The next category (Panel C) contains various return moments, including volatility measures, coskewness, skewness, and kurtosis.[5] Characteristics that provide proxies for managerial skill make up the fourth category (Panel D). This category includes alphas for various

[5]Idiosyncratic volatility is the standard deviation of the residuals for the CAPM fitted over the last 36 months, while systematic volatility is volatility over the last 36 months minus idiosyncratic volatility. Coskewness is the beta for the squared market factor for the quadratic CAPM fitted over the last 36 months; idiosyncratic skewness is the skewness of the residuals for the quadratic CAPM fitted over the last 36 months.

multifactor models for hedge fund risks (Fung and Hsieh 2004; Bali et al. 2021; Chen, Han, and Pan 2021) computed over the last twelve, 24, and 36 months as well as the R^2 statistic over the last twelve months for the popular Fung and Hsieh (2004) seven-factor model. Other proxies of managerial skill in the fourth category are the maximum return over the last twelve months and assets under management.

Table 1 (continued)

(1) Variable	(2) Abbreviation
Panel D: Managerial skill	
Alpha over last 12 months for Fung and Hsieh (2004) 7-factor model	`AlphaFH7_12mo`
R^2 over last 12 months for Fung and Hsieh (2004) 7-factor model	`R2FH7_12mo`
Alpha over last 36 months for Fung and Hsieh (2004) 7-factor model	`AlphaFH7_36mo`
Alpha over last 24 months for Bali et al. (2021) 9-factor model	`AlphaBBCC9_24mo`
Alpha over last 36 months for Chen, Han, and Pan (2021) 11-factor model	`AlphaCHP11_36mo`
Maximum return over the last 12 months	`MaxRet_12mo`
Assets under management (log)	`AUM`
Panel E: Economic variables	
Economic policy uncertainty index	`EPU`
Equity market uncertainty index	`EMU`
TED spread	`TED`
VIX	`VIX`

We also consider four economic variables (Panel E), which we interact with the fund characteristics. This is in the spirit of Gu, Kelly, and Xiu (2020), who predict individual stock returns using firm characteristics and their interactions with a set of economic variables. The interactions allow the relations between fund characteristics and future hedge fund returns to vary with economic conditions, as captured by the economic variables. The first two economic variables are the economic policy uncertainty index (Baker, Bloom, and Davis 2016) and the equity market uncertainty index, both of which are from the Economic Policy Uncertainty website. We also include two leading measures of financial stress and/or risk: the difference between the three-month LIBOR and three-month Treasury bill yields (TED spread) and the Chicago Board Options Exchange volatility index (VIX). The TED spread and VIX are

from FRED. After allowing the fund characteristics to interact with the economic variables, we have a total of $23 + 4 \times 23 = 115$ predictors.

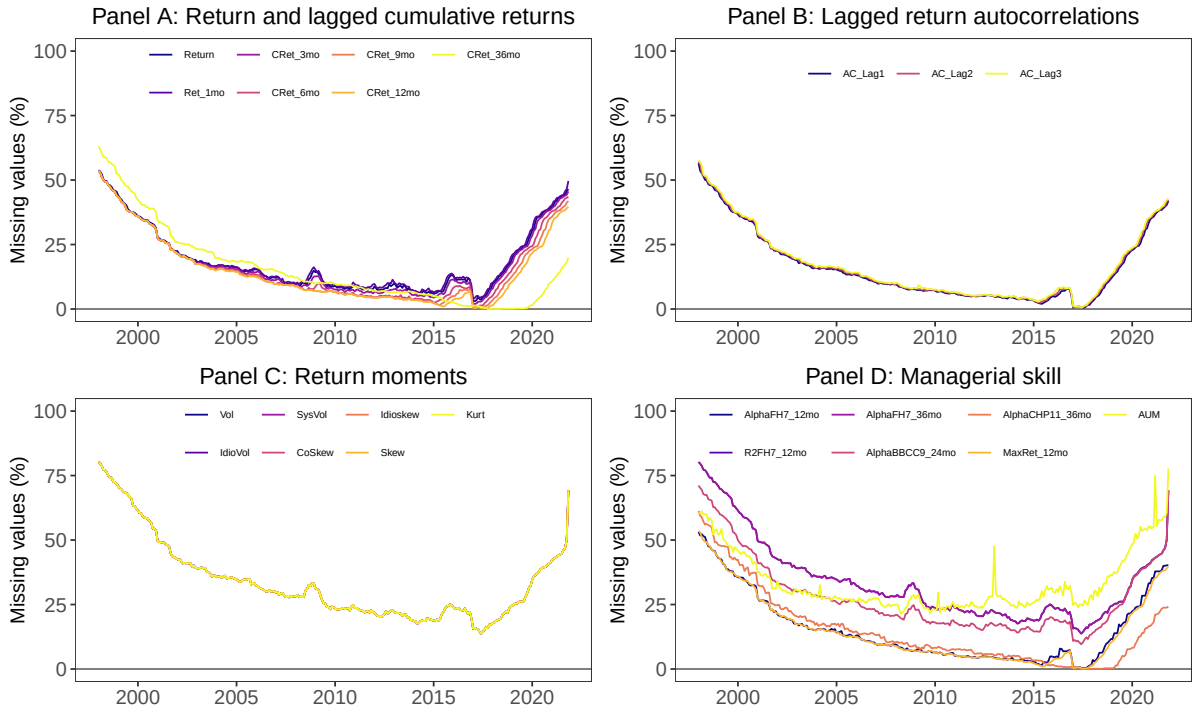


Figure 1. Missing value percentages. The figure depicts the percentage of missing values for hedge fund returns and each fund characteristic across all active funds for each month in the 1998:01 to 2021:12 sample period.

Figure 1 shows monthly time-series plots of the percent of missing values across active hedge funds for fund returns and characteristics. A few features of missing values for hedge fund data stand out in the figure. There are a number of months, especially early in the sample period, when there are large percentages of missing values for returns and many characteristics, reaching as high as approximately 60% to 80%. The percentages of missing values follow a U-shaped pattern. They generally decline from the start of the sample period until around 2017, when the percentages of missing values are substantially lower, often falling to less than 10%. Thereafter, the percentages of missing values typically increase, but they remain approximately 20 percentage points below the peaks at the start of the sample. Overall, Figure 1 confirms that missing values are a prevalent trait of hedge fund data.

3. BRITS

This section outlines BRITS (Cao et al. 2018), our main strategy of interest for filling in the (many) missing values in hedge fund data to improve hedge fund return prediction. BRITS is a specialized bidirectional recurrent neural network for multivariate time-series imputation. It analyzes data in both the forward and backward directions, allowing for more accurate imputation. BRITS further improves imputation by utilizing the cross-sectional correlations of the observed entries, which is valuable for imputing hedge fund data as well as financial data more generally, given the role of common influences in such data. The bidirectional time-series and cross-sectional components are combined to generate the final imputation output.

3.1. Preliminaries

Consider a generic variable $x_{i,t}$ (a fund characteristic or return) that is observed for $i = 1, \ldots, N$ funds over $t = 1, \ldots, T$ periods. The $N \times T$ data matrix is denoted by $\boldsymbol{X} = [\ \boldsymbol{x}_1 \ \cdots \ \boldsymbol{x}_T\ ]$, where $\boldsymbol{x}_t = [\ x_{1,t} \ \cdots \ x_{N,t}\ ]'$ is the N-dimensional vector of variable observations for period t. The variable is observed at evenly spaced intervals, corresponding to monthly observations in our application. However, many of the $x_{i,t}$ elements have missing values, where the missing values can occur irregularly. Below is a simple example of a data matrix $\boldsymbol{X}$ with $N = 4$ and $T = 8$ and missingness of arbitrary form; a "?" in cell indicates a missing value:

$i = 1$	$x_{1,1}$	$x_{1,2}$	?	$x_{1,4}$	?	$x_{1,6}$	$x_{1,7}$	$x_{1,8}$
$i = 2$	$x_{2,1}$	?	?	?	$x_{2,5}$	$x_{2,6}$	$x_{2,7}$	$x_{2,8}$
$i = 3$	?	$x_{3,2}$	$x_{3,3}$	$x_{3,4}$	$x_{3,5}$	$x_{3,6}$	?	$x_{3,8}$
$i = 4$	$x_{4,1}$	$x_{4,2}$	?	?	$x_{4,5}$	?	$x_{4,7}$	$x_{4,8}$
	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_3$	$\boldsymbol{x}_4$	$\boldsymbol{x}_5$	$\boldsymbol{x}_6$	$\boldsymbol{x}_7$	$\boldsymbol{x}_8$

Due to the missing values, the data matrix $\boldsymbol{X}$ is incomplete. The goal is to complete the matrix by filling in the missing values.

An $N \times T$ masking matrix $\boldsymbol{M} = [\ \boldsymbol{m}_1 \ \cdots \ \boldsymbol{m}_T\ ]$ tracks the missing components in $\boldsymbol{X}$, where $\boldsymbol{m}_t = [\ m_{1,t} \ \cdots \ m_{N,t}\ ]'$, and $m_{i,t} = 0$ if $x_{i,t}$ is missing and one otherwise. For the above $\boldsymbol{X}$ example, $\boldsymbol{M}$ is given by

$i = 1$	1	1	0	1	0	1	1	1
$i = 2$	1	0	0	0	1	1	1	1
$i = 3$	0	1	1	1	1	1	0	1
$i = 4$	1	1	0	0	1	0	1	1
	$\boldsymbol{m}_1$	$\boldsymbol{m}_2$	$\boldsymbol{m}_3$	$\boldsymbol{m}_4$	$\boldsymbol{m}_5$	$\boldsymbol{m}_6$	$\boldsymbol{m}_7$	$\boldsymbol{m}_8$

In addition, a time gap $\delta_{i,t}$ defines how many periods it has been since a non-missing value $x_{i,t}$ was observed (where $\delta_{i,1} = 0$). The $N \times T$ matrix of time gaps is denoted by $\boldsymbol{\Delta} = [\ \boldsymbol{\delta}_1 \ \cdots \ \boldsymbol{\delta}_T\ ]$, where $\boldsymbol{\delta}_t = [\ \delta_{1,t} \ \cdots \ \delta_{N,t}\ ]'$. For the $\boldsymbol{X}$ and $\boldsymbol{M}$ example matrices above, $\boldsymbol{\Delta}$ is given by

$i = 1$	0	1	1	2	1	2	1	1
$i = 2$	0	1	2	3	4	1	1	1
$i = 3$	0	1	1	1	1	1	1	2
$i = 4$	0	1	1	2	3	1	2	1
	$\boldsymbol{\delta}_1$	$\boldsymbol{\delta}_2$	$\boldsymbol{\delta}_3$	$\boldsymbol{\delta}_4$	$\boldsymbol{\delta}_5$	$\boldsymbol{\delta}_6$	$\boldsymbol{\delta}_7$	$\boldsymbol{\delta}_8$

We construct the matrix triplet $\{\boldsymbol{X}, \boldsymbol{M}, \boldsymbol{\Delta}\}$ for the hedge fund returns as well as each of the 23 fund characteristics in our panel of data. The matrix triplet is subsequently used in the BRITS algorithm to fill in the missing values and thereby complete the data matrix $\boldsymbol{X}$ for each of the 24 variables.

3.2. Algorithm

BRITS employs a recurrent neural network (RNN) architecture, and it utilizes both forward and backward imputation to fill in missing values along the time-series dimension. The bidirectional approach improves imputation by substantially increasing the speed of convergence when training the network and by reducing the effects of sequential prediction errors made early and transferred to later parts of the network. It also improves estimates of missing values near the start and end of the sample period. Instead of a constant (as in other approaches), BRITS treats each missing value as a variable of an RNN graph during backpropagation, which improves stability during training. In addition to the time-series dimension, BRITS harnesses information along the cross-sectional dimension to further improve imputation. This is relevant for hedge fund data, given the common influences often underlying financial data.

We outline basic aspects of the BRITS algorithm; see Cao et al. (2018, Section 4) for details. First, consider forward imputation with correlation across the elements of $\boldsymbol{x}_t$ in the context of an RNN architecture. Denote the K-dimensional hidden state vector by $\boldsymbol{h}_t$. The algorithm updates time-series and cross-sectional predictions of $\boldsymbol{x}_t$, denoted by $\hat{\boldsymbol{x}}_t$ and $\hat{\boldsymbol{z}}_t$, respectively, which are then blended to form a combination prediction of $\boldsymbol{x}_t$, denoted by $\hat{\boldsymbol{c}}_t$.

We begin with the regression component of the RNN that transforms the previous hidden state to the time-series prediction $\hat{\boldsymbol{x}}_t$:

$$\hat{\boldsymbol{x}}_t = \boldsymbol{b}_x + \boldsymbol{W}_x \boldsymbol{h}_{t-1}, \tag{1}$$

where $\boldsymbol{b}_x$ is an N-dimensional vector of bias weights, $\boldsymbol{W}_x$ is an $N \times K$ matrix of weights, $\boldsymbol{h}_0 = \boldsymbol{0}_N$, and $\boldsymbol{0}_N$ is an N-dimensional vector of zeros.[6] Using $\boldsymbol{m}_t$, the complement vector

[6] Bias is this context is different from its statistical meaning (e.g., the bias-variance trade-off).

for $\boldsymbol{x}_t$ is defined as follows:

$$\boldsymbol{x}_t^c = \boldsymbol{m}_t \odot \boldsymbol{x}_t + (\boldsymbol{\iota}_N - \boldsymbol{m}_t) \odot \hat{\boldsymbol{x}}_t, \tag{2}$$

where $\odot$ indicates element-wise multiplication, and $\boldsymbol{\iota}_N$ is an N-dimensional vector of ones. The complement vector replaces the missing values in $\boldsymbol{x}_t$ with the corresponding values from the time-series prediction $\hat{\boldsymbol{x}}_t$. The cross-sectional prediction is given by

$$\hat{\boldsymbol{z}}_t = \boldsymbol{b}_z + \boldsymbol{W}_z \boldsymbol{x}_t^c, \tag{3}$$

where $\boldsymbol{b}_z$ is an N-dimensional vector of bias weights, and $\boldsymbol{W}_z$ is an $N \times N$ matrix of weights with zeros along the main diagonal. The $\boldsymbol{W}_z$ matrix uses information in the elements $x_{j,t}^c$ for $i \neq j$ to estimate $x_{i,t}$.

To incorporate the time gaps, a K-dimensional vector of temporal decay factors is defined as

$$\boldsymbol{\gamma}_t = \exp[-\max(\boldsymbol{0}_K, \boldsymbol{b}_\gamma + \boldsymbol{W}_\gamma \boldsymbol{\delta}_t)], \tag{4}$$

where $\boldsymbol{b}_\gamma$ is a K-dimensional vector of bias weights, $\boldsymbol{W}_\gamma$ is a $K \times N$ matrix of weights, and the exponential function in Equation (4) operates element-wise on its vector argument. To combine the time-series and cross-sectional predictions in Equations (1) and (3), respectively, the following vector is first defined:

$$\boldsymbol{\beta}_t = \sigma(\boldsymbol{b}_\beta + \boldsymbol{W}_\beta[\boldsymbol{\gamma}_t \circ \boldsymbol{m}_t]), \tag{5}$$

where $\boldsymbol{b}_\beta$ is an N-dimensional vector of bias weights, $\boldsymbol{W}_\beta$ is an $N \times (K+N)$ matrix of weights, $\circ$ is the vertical concatenation operator, and $\sigma(\cdot)$ is the sigmoid function, which operates element-wise on its vector argument. Using Equation (5), the combination prediction is

given by

$$\hat{\boldsymbol{c}}_t = \boldsymbol{\beta}_t \odot \hat{\boldsymbol{z}}_t + (\boldsymbol{\iota}_N - \boldsymbol{\beta}_t) \odot \hat{\boldsymbol{x}}_t, \tag{6}$$

while the complement vector for $\boldsymbol{c}_t$ is defined as

$$\boldsymbol{c}_t^c = \boldsymbol{m}_t \odot \boldsymbol{x}_t + (1 - \boldsymbol{m}_t) \odot \hat{\boldsymbol{c}}_t. \tag{7}$$

Finally, the hidden state vector is updated via

$$\boldsymbol{h}_t = \sigma(\boldsymbol{b}_h + \boldsymbol{W}_h[\boldsymbol{h}_{t-1} \odot \boldsymbol{\gamma}_t] + \boldsymbol{U}_h[\boldsymbol{c}_t^c \circ \boldsymbol{m}_t]), \tag{8}$$

where $\boldsymbol{b}_h$ is a K-dimensional vector of bias terms, and $\boldsymbol{W}_h$ ($\boldsymbol{U}_h$) is a $K \times K$ ($K \times 2N$) matrix of weights.[7]

Running the RNN forward produces $\{\hat{\boldsymbol{x}}_t\}_{t=1}^T$, $\{\hat{\boldsymbol{z}}_t\}_{t=1}^T$, and $\{\hat{\boldsymbol{c}}_t\}_{t=1}^T$, the sequences of time-series, cross-sectional, and combination predictions of the columns of $\boldsymbol{X}$. To improve convergence speed when training the network, Cao et al. (2018) recommend minimizing the following objective for the forward direction:

$$\mathcal{L}^{\text{forward}} = \sum_{t=1}^{T} \boldsymbol{m}_t'[\phi(\boldsymbol{x}_t, \hat{\boldsymbol{x}}_t) + \phi(\boldsymbol{x}_t, \hat{\boldsymbol{z}}_t) + \phi(\boldsymbol{x}_t, \hat{\boldsymbol{c}}_t)], \tag{9}$$

where $\phi(\cdot, \cdot)$ is a loss function that operates element-wise on its vector argument. Using the same weights, the RNN is then run in the backward direction; denote the corresponding sequences of predictions by $\{\hat{\boldsymbol{x}}_t^{\downarrow}\}_{t=1}^T$, $\{\hat{\boldsymbol{z}}_t^{\downarrow}\}_{t=1}^T$, and $\{\hat{\boldsymbol{c}}_t^{\downarrow}\}_{t=1}^T$. When running the algorithm in the backward direction, an analogous objective to Equation (9) is used:

$$\mathcal{L}^{\text{backward}} = \sum_{t=1}^{T} \boldsymbol{m}_t' \left[\phi\left(\boldsymbol{x}_t, \hat{\boldsymbol{x}}_t^{\downarrow}\right) + \phi\left(\boldsymbol{x}_t, \hat{\boldsymbol{z}}_t^{\downarrow}\right) + \phi\left(\boldsymbol{x}_t, \hat{\boldsymbol{c}}_t^{\downarrow}\right)\right]. \tag{10}$$

[7] In practice, a long short-term memory (LSTM) is used as the recurrent component in Equation (8) to circumvent the gradient exploding or vanishing problem that can plague conventional RNNs.

An additional objective is included to promote consistency between the backward and forward combination predictions:

$$\mathcal{L}^{\text{consistency}} = \sum_{t=1}^{T} \phi\left(\hat{\boldsymbol{c}}_t, \hat{\boldsymbol{c}}_t^{\downarrow}\right). \tag{11}$$

Given $\boldsymbol{X}$, $\boldsymbol{M}$, and $\boldsymbol{\Delta}$, we train the RNN via the Adam optimizer (Kingma and Ba 2017) with the objective of minimizing the sum of Equations (9) to (11). Following Cao et al. (2018), we use the absolute error for the loss function $\phi(\cdot, \cdot)$ in Equations (9) to (11), and based on the fitted network, the final estimates of the missing elements in $\boldsymbol{x}_t$ is the average of the corresponding elements in $\hat{\boldsymbol{c}}_t$ and $\hat{\boldsymbol{c}}_t^{\downarrow}$.[8]

4. Imputation Accuracy

As a preliminary exercise, we examine the performance of CS-Mean, SVT, and BRITS in terms of imputation error. In the experiment, we randomly select a percentage of non-missing fund values and treat them as missing; then, we use one of the imputation methods to fill in values for the pseudo missing values (along with the original missing values). For the pseudo missing values and their corresponding filled-in values, we compute the RMSE to measure the accuracy of the imputations. We carry out the experiment for 1998:01 to 2012:12 (the initial in-sample period) for the hedge fund returns and each of the 23 fund characteristics. We conduct the experiment twice, randomly selecting 10% and then 20% of the non-missing values to treat as missing.

Table 2 reports the results for the data imputation experiment. When we randomly treat 10% of the non-missing values as missing in Panel A, BRITS stands out as the most accurate imputation method. Specifically, among the three imputation methods, BRITS

[8] We tune the following hyperparameters for the BRITS network: learning rate, number of epochs, number of hidden units, and early stopping patience levels. We tune the hyperparameters over the initial in-sample period, so there is no look-ahead bias in the tuning process when generating the out-of-sample return predictions.

Table 2. Imputation fidelity

The table reports results for an imputation experiment in which we randomly select a percentage of non-missing hedge fund return and characteristic values and treat them as missing; then, we use an imputation method to fill in values for the pseudo missing values (along with the original missing values). For the pseudo missing values and their corresponding filled-in values, each entry reports the RMSE to measure the accuracy of the imputations. The imputation methods are the cross-sectional mean (CS-Mean), singular value thresholding (SVT), and bidirectional recurrent imputation network for time series (BRITS). Bold indicates the imputation method with the lowest RMSE. Panel A (B) reports results for 10% (20%) pseudo missing values. The sample period is 1998:01 to 2012:12.

(1) Variable	(2) CS-Mean	(3) SVT	(4) BRITS	(5) Variable	(6) CS-Mean	(7) SVT	(8) BRITS
Panel A. 10% pseudo missing values							
Return	4.0381	4.0656	**3.4477**	SysVol	0.8665	**0.0645**	0.1515
Ret_1mo	7.9104	7.9584	**7.7001**	CoSkew	0.0415	0.0152	**0.0105**
Cret_3mo	10.1245	10.0896	**6.9363**	IdioSkew	0.8386	**0.2128**	0.2233
Cret_6mo	11.7554	11.9334	**5.8979**	Skew	0.9303	**0.2236**	0.2521
Cret_9mo	16.0935	16.2844	**8.8056**	Kurt	2.9423	**1.5121**	1.7106
Cret_12mo	20.7144	20.9878	**9.8611**	AlphaFH7_12mo	3.9689	3.7360	**3.3112**
Cret_36mo	28.3747	28.2551	**15.7967**	R2FH7_12mo	0.1962	**0.0318**	0.1099
AC_Lag1	0.3120	0.1747	**0.1738**	AlphaFH7_36mo	1.3524	1.0298	**0.2635**
AC_Lag2	0.3160	0.1746	**0.1729**	AlphaBBCC9_24mo	1.2570	0.6770	**0.2622**
AC_Lag3	0.3369	0.1853	**0.1829**	AlphaCHP11_36mo	29.9949	29.8859	**18.9780**
Vol	4.8429	4.0007	**1.5872**	MaxRet_12mo	8.9909	8.0718	**3.6069**
IdioVol	4.5944	3.9721	**1.5348**	AUM	1.6154	**0.3464**	7.1020

produces the lowest RMSE for the hedge fund returns and 17 of the 23 fund characteristics; in many cases, BRITS provides a sizable improvement in imputation accuracy. For the other six fund characteristics, SVT produces the lowest RMSE. For SysVol, R2FH7_12mo, and AUM, SVT performs well, generating the most accurate imputations by a substantive margin. The RMSEs for SVT and BRITS are very similar for a handful of characteristics, including AC_Lag1, AC_Lag2, AC_Lag3, CoSkew, and IdioSkew. CS-Mean fails to generate the most accurate imputation for any variable, and in a number of cases, its RMSE is substantially larger than those for SVT and BRITS.

The results are similar when we randomly treat 20% of the non-missing values as missing in Panel B of Table 2. BRITS again leads to the lowest RMSE for hedge fund returns,

Table 2 (continued)

(1) Variable	(2) CS-Mean	(3) SVT	(4) BRITS	(5) Variable	(6) CS-Mean	(7) SVT	(8) BRITS
Panel B. 20% pseudo missing values							
Return	4.1207	4.1901	**3.5863**	SysVol	0.8731	**0.0740**	0.1577
Ret_1mo	6.2703	6.3494	**6.0267**	CoSkew	0.0426	0.0161	**0.0109**
Cret_3mo	8.9340	9.1814	**5.9644**	IdioSkew	3.8454	3.1905	**1.4431**
Cret_6mo	13.0564	13.6099	**7.0421**	Skew	0.9285	**0.2530**	0.2755
Cret_9mo	20.1071	20.7028	**8.8869**	Kurt	2.9374	1.7881	**1.6865**
Cret_12mo	19.0195	19.8569	**9.1863**	AlphaFH7_12mo	3.8699	3.6632	**3.3433**
Cret_36mo	32.0088	32.4138	**15.4648**	R2FH7_12mo	0.1966	**0.0335**	0.1096
AC_Lag1	0.3131	0.1814	**0.1813**	AlphaFH7_36mo	1.1574	0.8043	**0.2793**
AC_Lag2	0.3162	0.1840	**0.1829**	AlphaBBCC9_24mo	1.3866	0.9237	**0.3255**
AC_Lag3	0.3370	0.1944	**0.1943**	AlphaCHP11_36mo	88.5379	88.4574	**85.5289**
Vol	4.1345	3.2885	**1.5818**	MaxRet_12mo	7.6046	6.8102	**3.1418**
IdioVol	0.8305	0.2371	**0.2356**	AUM	1.6190	**0.4480**	7.0319

while it generates the lowest RMSE for 19 of the 23 fund characteristics. SVT delivers the most accurate imputations for the remaining four characteristics. Again, BRITS produces substantive improvement in imputation accuracy by a considerable margin for a number of characters, while SVT performs well for SysVol, R2FH7_12mo, and AUM; the RMSEs of SVT and BRITS are again quite similar for some variables. Overall, the results in Table 2 suggest that BRITS provides the greatest fidelity for data imputation (as measured by RMSE), while SVT also performs relatively well. Thus, in terms of the bias-variance trade-off, we have preliminary evidence that the added complexity of BRITS vis-à-vis SVT and CS-Mean improves performance. However, we are not primarily interested in data imputation per se; rather, we are concerned with data imputation as a means to an end, namely, hedge fund return prediction. To this end, in the next section, we investigate the efficacy of the different data imputation methods when it comes to out-of-sample hedge fund return prediction.

5. Out-of-Sample Results

This section describes the construction of hedge fund excess return predictions based on fund characteristics and econommic variables. Then, it reports the main set of out-of-sample results regarding the statistical accuracy and economic value of hedge fund return predictions based on the different imputation methods.

5.1. Prediction Models

A general prediction model for the one-month-ahead hedge fund excess return can be expressed as follows:

$$r_{i,t+1} = f(\boldsymbol{x}_{i,t}; \boldsymbol{\theta}) + \varepsilon_{i,t+1}, \tag{12}$$

where $r_{i,t}$ is the month-t excess return for hedge fund i, $\boldsymbol{x}_{i,t} = [\ x_{i,1,t} \quad \cdots \quad x_{i,K,t}\ ]'$ is a K-dimensional vector of predictor variables, $f(\boldsymbol{x}_{i,t}; \boldsymbol{\theta})$ is the conditional expectation (i.e., prediction) function that depends on the parameter vector $\boldsymbol{\theta}$, and $\varepsilon_{i,t+1}$ is a mean-zero and serially uncorrelated additive disturbance term. Note that Equation (12) is a pooled model, since the prediction function and parameter vector is the same across funds. Pooling substantially reduces the number of parameters that need to be estimated—as it eliminates the need to estimate a separate prediction model for each fund—which is beneficial in terms of the bias-variance trade-off, especially given the inherently small predictable component in asset returns as well as the limited historical data for many individual hedge funds. Pooled prediction models are widely used in the literature (e.g., Freyberger, Neuhierl, and Weber 2020; Gu, Kelly, and Xiu 2020; Filippou et al. 2024). The out-of-sample excess return forecast corresponding to Equation (12) is given by

$$\hat{r}_{i,t+1} = \hat{f}\left(\boldsymbol{x}_{i,t}; \hat{\boldsymbol{\theta}}\right), \tag{13}$$

where $\hat{f}(\cdot;\hat{\boldsymbol{\theta}})$ is the fitted prediction function based on data through month t.

Our core model for hedge fund return prediction is a feedforward neural network. Recent studies highlight the efficacy of deep neural networks for forecasting equity, mutual fund, and hedge fund returns (e.g., Gu, Kelly, and Xiu 2020; Kaniel et al. 2023; Wu et al. 2021, respectively). A neural network is comprised of multiple layers of interconnected neurons. The first layer consists of the predictors (which total 115 in our models, corresponding to the 23 fund characteristics and their interactions with the four economic variables). One or more hidden layers follow, while the final layer produces the prediction. Each hidden layer l consists of P_l neurons, where the mth neuron in the lth layer computes a nonlinear transformation of the inputs from the previous layer:

$$h_m^{(l)} = g\left(\omega_{m,0}^{(l)} + \sum_{j=1}^{P_{l-1}} \omega_{m,j}^{(l)} h_j^{(l-1)}\right) \quad \text{for } m = 1,\ldots,P_l;\ l = 1,\ldots,L, \tag{14}$$

where $h_j^{(0)} = x_{i,j,t}$ corresponds to the predictors, $\omega_{m,0}^{(l)}$ and $\omega_{m,j}^{(l)}$ are network weights, and $g(\cdot)$ is a nonlinear activation function. We use the Gaussian error linear unit (GELU) activation function (Hendrycks and Gimpel 2023): $g(x) = x\Phi(x)$, where $\Phi(x)$ is the standard Gaussian cumulative distribution function. GELU often offers performance gains relative to the popular rectified linear unit (ReLU) activation function. The network's final layer transforms the neurons in the last hidden layer into a prediction:

$$\hat{r}_{i,t+1}^{\text{NN}} = \omega_0^{(L+1)} + \sum_{j=1}^{P_L} \omega_j^{(L+1)} h_j^{(L)}, \tag{15}$$

where $\omega_0^{(L+1)}$ and $\omega_j^{(L+1)}$ are additional weights. We consider deep neural networks with $L = 3, 4, 5$ hidden layers. Each network is trained by estimating the weights using stochastic gradient descent (SGD) with the objective of minimizing the MSE over the training sample. We implement SGD via the Adam optimizer (Kingma and Ba 2017).[9] In addition to the

[9]We tune the hyperparameters for each network—including the number of neurons for the hidden layer, dropout rate, and learning rate—using a walk-forward cross-validation method that respects the time-series

three neural network predictions, we consider an ensemble prediction that is the average of the three predictions (NN-Ensemble).

We generate a sequence of out-of-sample neural network predictions using a rolling estimation window of 15 years (180 months) as follows. The initial in-sample period is 1998:01 to 2012:12. We use a data imputation method to fill in missing values for the hedge fund return and fund characteristic observations for 1998:01 to 2012:12. Using the filled-in fund return and characteristic data in conjunction with data for the economic variables, we train the neural network over the same period. Plugging the values for the fund characteristics and economic variables for 2012:12 into the fitted neural network, we compute hedge fund return forecasts for 2013:01. Next, we use a data imputation method to fill in missing values for the fund return and characteristic observations for 1998:02 to 2013:01, train the model over the same period using the filled-in return and characteristic data together with data for the economic variables, plug the values for the characteristics and economic variables for 2013:01 into the fitted neural network, and compute return forecasts for 2013:02. Continuing in this manner through the end of the sample period, we generate a sequence of out-of-sample hedge fund return forecasts for 2013:01 to 2021:12.

5.2. Statistical Accuracy

We use a pooled version of the popular Fama and French (1989) and Campbell and Thompson (2008) out-of-sample R^2 statistic to measure the out-of-sample accuracy of a hedge fund return forecast based on a neural network and data imputation method relative to a benchmark return forecast. The original out-of-sample R^2 statistic focuses on the aggregate equity market and uses the historical average as the benchmark return forecast. Gu, Kelly, and Xiu (2020) find that a zero benchmark return is more stringent for individual stocks, due in part to the difficulty in reliably estimating the mean return for relatively short samples for

dimension of the data. Carefully tuning the hyperparameters for neural networks is clearly important for forecasting asset returns to prevent overfitting, given the relatively low signal-to-noise ratio in asset return data.

numerous stocks. The same situation applies to a greater extent for hedge funds, given the even shorter samples available for many funds.

The pooled version of out-of-sample R^2 statistic for comparing a neural network excess return forecast to a zero benchmark forecast is given by

$$R^2_{\text{OS,Pool}} = 1 - \frac{\sum_{i=1}^{N}\sum_{s=1}^{T-t_{\text{in}}}\left(r_{i,t_{\text{in}}+s} - \hat{r}^{\text{NN}}_{i,t_{\text{in}}+s}\right)^2}{\sum_{i=1}^{N}\sum_{s=1}^{T-t_{\text{in}}} r^2_{i,t_{\text{in}}+s}}, \tag{16}$$

where t_{in} (T) is the end of the initial in-sample (complete sample) period. For notational simplicity, we assume a balanced panel in Equation (16). Of course, we only apply Equation (16) to the hedge fund return observations with non-missing values to measure forecast accuracy, which produces an unbalanced panel; it is straightforward to adjust Equation (16) for an unbalanced panel based on the non-missing return observations for each month. Equation (16) measures the proportional reduction in out-of-sample MSE for the neural network vis-à-vis the benchmark forecast across all of the hedge funds. We use a pooled version of the Diebold and Mariano (1995) and West (1996) (DMW) statistic to test whether the neural network forecast delivers a statistically significant reduction in MSE relative to the benchmark. The DMW statistic can be computed as the t-statistic corresponding to the intercept in the following pooled regression:

$$\underbrace{r^2_{i,t} - \left(r_{i,t} - \hat{r}^{\text{NN}}_{i,t}\right)^2}_{d_{i,t}} = a + \varepsilon_{i,t} \quad \text{for } i = 1,\ldots,N;\ t = t_{\text{in}}+1,\ldots,T, \tag{17}$$

where $d_{i,t}$ is the month-t loss differential for hedge fund i (i.e., the difference between the squared errors for the benchmark and NN forecasts). We test H_0: $a \leq 0$ ($R^2_{\text{OS,Pool}} \leq 0$) against H_A: $a > 0$ ($R^2_{\text{OS,Pool}} > 0$). When computing the pooled DMW statistic in Equation (17), we account for cross-sectional dependency by clustering the standard error by month and hedge funds.

Table 3. $R^2_{\text{OS,Pool}}$ statistics

The table reports $R^2_{\text{OS,Pool}}$ statistics in percent for hedge fund excess return forecasts based on different neural network specifications. NN3, NN4, and NN5 denote neural networks with three, four, and five hidden layers, respectively; NN-Ensemble is an average of the NN3, NN4, and NN5 forecasts. Results are reported for the different data imputation methods in the column headings when training the neural networks: complete cases (i.e., no data imputation), cross-sectional mean (CS-Mean), singular value thresholding (SVT), bidirectional recurrent imputation network for time series (BRITS), and BRITS excluding return imputation (BRITS-Partial). The $R^2_{\text{OS,Pool}}$ statistic measures the proportional reduction in out-of-sample MSE for the neural network forecast vis-à-vis the zero benchmark forecast; based on a pooled version of the Diebold and Mariano (1995) and West (1996) test, *, **, and *** indicate that the reduction in out-of-sample MSE is significant at the 10%, 5%, and 1% levels, respectively. The out-of-sample period is 2013:01 to 2021:12. The total number of hedge fund return observations with non-missing values that are available for evaluation is 137,505.

(1)	(2)	(3)	(4)	(5)	(6)
Model	Complete cases	CS-Mean	SVT	BRITS	BRITS-Partial
NN3	0.16	−0.09	1.68***	1.58***	1.26
NN4	0.01	0.05	1.43	1.59***	2.04***
NN5	0.19	0.58	2.76***	2.45***	2.68***
NN-Ensemble	0.16	0.29	2.28***	2.16***	2.25***

Table 3 reports $R^2_{\text{OS,Pool}}$ statistics for the various neural network forecasts as well as the different data imputation methods. In addition to the CS-Mean, SVT, and BRITS methods, we consider another approach—complete cases—that eschews data imputation and only uses data with non-missing observations for every variable when training the prediction model. Although popular, this approach substantially reduces the size of the sample available for training the prediction model, due to the large number of missing values for hedge fund data. We also consider a version of BRITS (BRITS-Partial) that does not fill in missing values for hedge fund returns when training the prediction model (but does fill in missing values for the corresponding fund characteristics).

The results in the second column of Table 3 indicate that the complete case approach provides only limited improvements in accuracy for the neural network forecasts relative to the zero benchmark forecast. Although all of the $R^2_{\text{OS,Pool}}$ statistics are positive, they are all less than or equal to 0.19% and none are significant. The results are similar for the

CS-Mean imputation method in the third column. The $R^2_{\text{OS,Pool}}$ statistic is negative for the NN3. While the $R^2_{\text{OS,Pool}}$ statistics are positive for the other neural networks, they are less than or equal to 0.58%, and again none are significant.

SVT performs substantially better in the fourth column of Table 3. The $R^2_{\text{OS,Pool}}$ statistics are all positive, ranging from 1.43% (NN4) to 2.76% (NN5), and they are significant at the 1% level for NN3, NN4, and NN-Ensemble. Given the inherently large unpredictable component in asset returns, an $R^2_{\text{OS,Pool}}$ statistic around 1.5% is quite sizable. BRITS also performs quite well across the neural network forecasts in the fifth column. All of the $R^2_{\text{OS,Pool}}$ statistics and positive and sizable, ranging from 1.58% (NN3) to 2.45% (NN5), and all are significant at the 1% level. BRITS-Partial provides large improvements in accuracy for NN4, NN5, and NN-Ensemble, with $R^2_{\text{OS,Pool}}$ statistics of 2.04%, 2.68%, and 2.25%, respectively, each of which is significant at the 1% level.

In terms of statistical accuracy, the results in Table 3 indicate that more complex data imputation methods improve out-of-sample hedge fund return prediction. The complete case approach, which does not impute any missing values, as well as CS-Mean, which is the simplest data imputation method, provide the smallest gains. SVT is considerably more complicated than CS-Mean and performs substantially better. Furthermore, the two BRITS methods are substantially more complicated than SVT, and these methods also perform impressively in Table 3. Although the bias-variance trade-off cautions that more complex methods do not necessarily lead to improved performance (and can substantively detract from performance due to overfitting), when it comes to hedge fund return prediction, more complex data imputation methods appear warranted.

5.3. Economic Value

Next, we analyze the neural network hedge fund return predictions based on the different data imputation methods in terms of economic value. We begin by forming long-only top-decile portfolios, which are relevant due to the difficulty of shorting hedge funds compared

to assets such as stocks. Specifically, for a given data imputation method, at the end of each month, we sort hedge funds with non-missing return observations in terms of their excess return forecasts for the next month and take long positions in the funds in the top decile. We consider three weighting strategies for the funds in the top decile. The first is equal weighting. The second is rank weighting, where we rank the funds in the top decile in ascending order and assign weights that are proportional to the funds' ranks.[10] Unlike equal weighting, rank weighting provides a strategy for increasing the allocations to hedge funds that appear more promising in terms of their predicted returns within the top decile. Finally, we also consider prediction weighting, which Kaniel et al. (2023) find performs well when allocating across mutual funds according to return predictions based on neural networks. For each hedge fund in the top decile, we compute

$$\hat{\delta}_{i,t+1} = \hat{r}_{i,t+1} - \min_{i \in \text{top}}(\hat{r}_{i,t+1}), \tag{18}$$

Scaling the weights to sum to one, the allocation to fund i for month $t+1$ based on information through month t for the top decile is given by

$$w^{\text{PW}}_{i,t+1} = \frac{\hat{\delta}_{i,t+1}}{\sum_{i \in \text{top}} \hat{\delta}_{i,t+1}}. \tag{19}$$

Equation (18) ensures that the top-decile portfolio is long only. Similarly to rank weighting, prediction weighting refines the allocation within the top decile by placing greater emphasis on funds with higher predicted returns.

Performance metrics for the top-decile portfolios are reported in Table 4 for the different imputation methods. To conserve space, the results in Table 4 are for the NN-Ensemble forecasts. The results are similar for the other neural network forecasts and are reported in the Internet Appendix. The table reports a variety of performance metrics, including the

[10] As a simple example, suppose that there are three funds in the top decile, so their ranks are 1, 2, 3, with the third fund having the highest return forecast and thus receiving the highest rank. Scaling the weights to sum to one, the rank weights are 1/6, 2/6, and 3/6 for the first, second, and third funds, respectively.

Table 4. Top-decile portfolio performance

The table reports out-of-sample performance metrics for long-only top-decile portfolios of hedge funds formed by sorting based on the NN-Ensemble hedge fund excess return forecasts for the next month. Results are reported for the different data imputation methods in the column headings when training the neural networks: complete cases (i.e., no data imputation), cross-sectional mean (CS-Mean), singular value thresholding (SVT), bidirectional recurrent imputation network for time series (BRITS), and BRITS excluding return imputation (BRITS-Partial). FH7 and CLTZ9 are the Fung and Hsieh (2004) seven-factor and Chen et al. (2025, CLTZ9) nine-factor models, respectively. Significance of the mean excess return and FH7 and CLTZ9 alphas are based on t-statistics computed using robust standard errors; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Panel A, B, and C report results for equal, rank, and prediction weighting, respectively.

(1)	(2)	(3)	(4)	(5)	(6)
Performance metric	Complete cases	CS-Mean	SVT	BRITS	BRITS-Partial
Panel A. Equal weighting					
Ann. mean excess return	10.36%***	7.66%***	13.84%***	14.66%***	14.01%***
Ann. FH7 α	4.89%***	3.55%**	5.33%**	6.70%***	7.11%***
Ann. CLTZ9 α	6.58%***	5.49%***	4.95%**	6.43%**	7.52%***
Maximum drawdown	14.83%	10.09%	18.00%	20.03%	18.39%
Ann. Sharpe ratio	1.16	1.15	1.14	1.20	1.27
Panel B. Rank weighting					
Ann. mean excess return	10.78%***	7.28%***	15.86%***	17.12%***	16.36%***
Ann. FH7 α	5.23%**	3.40%*	6.59%**	7.95%***	8.80%***
Ann. CLTZ9 α	8.09%***	5.79%***	6.41%**	8.34%***	9.71%***
Maximum drawdown	14.55%	10.01%	17.96%	18.38%	16.37%
Ann. Sharpe ratio	1.12	1.03	1.18	1.27	1.29
Panel C. Prediction weighting					
Ann. mean excess return	10.98%**	3.65%	20.20%***	22.59%***	23.62%***
Ann. FH7 α	6.20%	0.58%	9.69%**	12.50%**	15.15%***
Ann. CLTZ9 α	12.49%**	5.52%	11.98%**	15.66%***	17.73%***
Maximum drawdown	23.05%	42.69%	20.54%	20.36%	21.61%
Ann. Sharpe ratio	0.73	0.27	1.18	1.25	1.27

annualized mean excess return, annualized alphas for the Fung and Hsieh (2004) seven-factor and Chen et al. (2025) nine-factor models (FH7 and CLTZ9, respectively) for hedge fund risks, maximum drawdown, and annualized Sharpe ratio. Panels A, B, and C report results based on equal, rank, and prediction weighting, respectively.

A similar pattern to that in Table 3 prevails in Table 4, in that performance tends to improve as we move from the simpler data imputation methods to the more complex methods. The pattern is not as strong for equal weighting in Panel A, while it becomes quite pronounced for rank and prediction weighting in Panels B and C, respectively. For rank weighting, the mean returns are higher for the more complex SVT and BRITS methods relative to the complete case and CS-Mean approaches. The same holds for the FH7 alphas and Sharpe ratios. The CLTZ9 alpha is lower for the CS-Mean than that for SVT and the two BRITS methods, although the CLTZ9 alpha for the complete case approach is higher than that for SVT (but not the two BRITS methods). The BRITS methods deliver the best overall performance in Panel B. For example, BRITS-Partial produces FH7 and CLTZ9 annualized alphas of 8.80% and 9.71%, respectively, both of which are significant at the 1% level, as well as an annualized Sharpe ratio of 1.29. All of these are the highest values for these metrics in Panel B. For prediction weighting in Panel C, SVT and the BRITS methods perform impressively, with BRITS-Partial again evincing the best overall gains. For example, it generates very sizable FH7 and CLTZ9 annualized alphas of 15.15% and 17.73%, respectively (both of which are significant at the 1% level) and an annualized Sharpe ratio of 1.27. In sum, the results in Table 4 indicate that using BRITS to impute missing data for training neural networks generates hedge fund return predictions that have substantial economic value for guiding dynamic hedge fund selection, especially for rank and prediction weighting.[11]

To complement the analysis of long-only top-decile portfolios, we also consider long-short portfolios that take long (short) positions in the top (bottom) decile of hedge funds sorted according to their return predictions for the next month. We again consider the different data imputation methods as well as equal, rank, and prediction weighting. Again, to conserve space, we focus on the NN-Ensemble forecasts; results for the other neural network forecasts

[11]By way of comparison, an equal-weighted portfolio that invests in all hedge funds with non-missing return observations in a given month produces an annualized mean excess return of 5.47% (significant at the 5% level), FH7 and CLTZ9 annualized alphas of 1.31% and 1.12%, respectively (both of which are insignificant), a maximum drawdown of 14.00%, and an annualized Sharpe ratio of 0.89.

Table 5. Long-short portfolio performance

The table reports out-of-sample performance metrics for long-short portfolios of hedge funds formed by sorting based on the NN-Ensemble hedge fund excess return forecasts for the next month. The long-short portfolio goes long (short) funds in the top (bottom) decile. Results are reported for the different data imputation methods in the column headings when training the neural networks: complete cases (i.e., no data imputation), cross-sectional mean (CS-Mean), singular value thresholding (SVT), bidirectional recurrent imputation network for time series (BRITS), and BRITS excluding return imputation (BRITS-Partial). FH7 and CLTZ9 are the Fung and Hsieh (2004) seven-factor and Chen et al. (2025, CLTZ9) nine-factor models, respectively. Significance of the mean excess return and FH7 and CLTZ9 alphas are based on t-statistics computed using robust standard errors; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Panel A, B, and C report results for equal-, rank-, and prediction-weighting, respectively.

(1)	(2)	(3)	(4)	(5)	(6)
Performance metric	Complete cases	CS-Mean	SVT	BRITS	BRITS-Partial
Panel A. Equal weighting					
Ann. mean excess return	6.32%***	4.20%*	12.95%***	14.00%***	12.53%***
Ann. FH7 α	5.82%*	4.57%*	8.59%***	9.71%***	10.21%***
Ann. CLTZ9 α	9.06%***	8.29%***	6.75%**	7.90%**	9.36%***
Maximum drawdown	12.05%	17.48%	11.13%	10.46%	9.86%
Ann. Sharpe ratio	0.82	0.64	1.28	1.35	1.40
Panel B. Rank weighting					
Ann. mean return	5.74%**	3.56%	15.98%***	17.46%***	15.38%***
Ann. FH7 α	5.65%	4.48%	11.09%***	12.18%***	13.01%***
Ann. CLTZ9 α	9.37%***	8.63%***	9.62%**	10.85%***	12.37%***
Maximum drawdown	15.85%	25.59%	12.21%	9.31%	7.24%
Ann. Sharpe ratio	0.67	0.44	1.34	1.45	1.41
Panel C. Prediction weighting					
Ann. mean return	2.82%	−0.92%	21.99%***	25.44%***	22.98%***
Ann. FH7 α	3.77%	1.11%	16.18%***	18.92%***	20.70%***
Ann. CLTZ9 α	10.29%	7.33%	19.55%***	22.33%***	21.89%***
Maximum drawdown	47.07%	61.54%	18.06%	25.69%	16.10%
Ann. Sharpe ratio	0.18	−0.06	1.28	1.35	1.31

are similar and provided in the Internet Appendix. For rank weighting, relative to the top decile, we need to reverse the ranks for the bottom decile.[12] For prediction weighting,

[12]Again, as a simple example, suppose that there are three funds in the bottom decile. Relative to the situation in Footnote 10 for the top decile, we reverse their ranks to 3, 2, 1, so the first fund, which has the lowest predicted return, receives the highest rank. Scaling the weights to sum to −1 for the short leg, the rank weights are −3/6, −2/6, and −1/6 for the first, second, and third funds, respectively.

Equation (18) becomes

$$\hat{\delta}_{i,t+1} = \hat{r}_{i,t+1} - \max_{i \in \text{bottom}} (\hat{r}_{i,t+1}) \tag{20}$$

for the bottom decile. Equation (20) ensures that bottom-decile portfolio is short only. Scaling the weights to sum to -1, we adjust Equation (19) as follows to get the allocation to fund i for the bottom decile:

$$w_{i,t+1}^{\text{PW}} = \frac{\hat{\delta}_{i,t+1}}{\sum_{i \in \text{bottom}} |\hat{\delta}_{i,t+1}|}. \tag{21}$$

Given the difficulty in shorting hedge funds, we view the analysis of the long-short portfolios as a largely theoretical exercise that provides additional insight into the efficacy of the different data imputation methods.[13]

Table 5 reports performance metrics for the long-short portfolios. The superior performance of the more complex imputation methods is even more evident than in Table 4, especially for rank and prediction weighting in Panels B and C, respectively. For SVT and the two BRITS methods, the long-short portfolios in Table 5 typically increase the annualized alphas and Sharpe ratios and reduce the maximum drawdowns compared to the corresponding values for the long-only top-decile portfolios in Table 4. The long-short portfolios that employ SVT and the BRITS methods for data imputation, along with prediction weighting, deliver especially large alphas in Panel C of Table 5. For SVT, the FH7 and CLTZ9 annualized alphas are 16.18% and 19.55%, respectively (both of which are significant at the 1% level). The alphas are even larger for BRITS and BRITS-Partial. The former generates FH7 and CLTZ9 annualized alphas of 18.92% and 22.33%, respectively, while the latter produces corresponding alphas of 20.70% and 21.89%, respectively (all of which are significant at the 1% level). Echoing the results in Table 4, those in Table 5 show that using BRITS to impute

[13] In practice, with the advent of hedge fund ETFs, it is becoming more feasible to form the type of long-short portfolios for hedge funds in the exercise. Companies currently offering hedge fund ETFs include New York Life, ProShares, WisdomTree, AlphaClone, First Trust, Unlimited ETFs, DBi, and BlackRock.

missing data for training neural networks generates hedge fund return predictions that have considerable economic value for guiding dynamic hedge fund selection, particularly for rank and prediction weighting.

Finally, Figure 2 presents the out-of-sample cumulative returns of the prediction-weighted decile portfolios constructed from neural network forecasts using information from BRITS and benchmarks imputed datasets. Again, the two BRITS imputation approaches and the SVT provide more information on hedge fund sorting and predictability, resulting in a larger spread in the prediction-weighted portfolios across the whole out-of-sample period. An investor who used BRITS-imputed information to invest in the top 10% of funds, with a neural network as the forecaster, would have earned a cumulative abnormal return of 207% and 216% with BRITS and BRITS-partial, respectively. The same investor would have avoided the worst 10% of funds, yielding a corresponding cumulative return of -19% and 11%, respectively. SVT also performs well in identifying mainly top-performing funds, leading to a cumulative return of 185% for the top-decile portfolio. In contrast, the cumulative return of the bottom-decile portfolio is -11%. The complete cases and CS-Mean approaches report the smallest spreads of the best and worst funds. The cumulative returns of their corresponding top-decile portfolios are 103% and 38%, respectively.

5.3.1 Performance Persistence

We now evaluate the performance persistence of the hedge fund portfolios constructed under the BRITS and benchmark data imputation schemes over three years. We follow Kaniel et al. (2023) and hold the hedge fund top-decile portfolios formed with the neural network predictions for up to three years in an overlapping structure. For example, we hold the top-decile fund portfolios constructed monthly based on the one-month ahead prediction for months ranging from 1 to 36 months with overlapping returns. We calculate the annualized mean returns and the Chen et al. (2025) (CLTZ) factor model annualized alphas of the portfolios' overlapping returns along with their corresponding t-statistics with Newey and

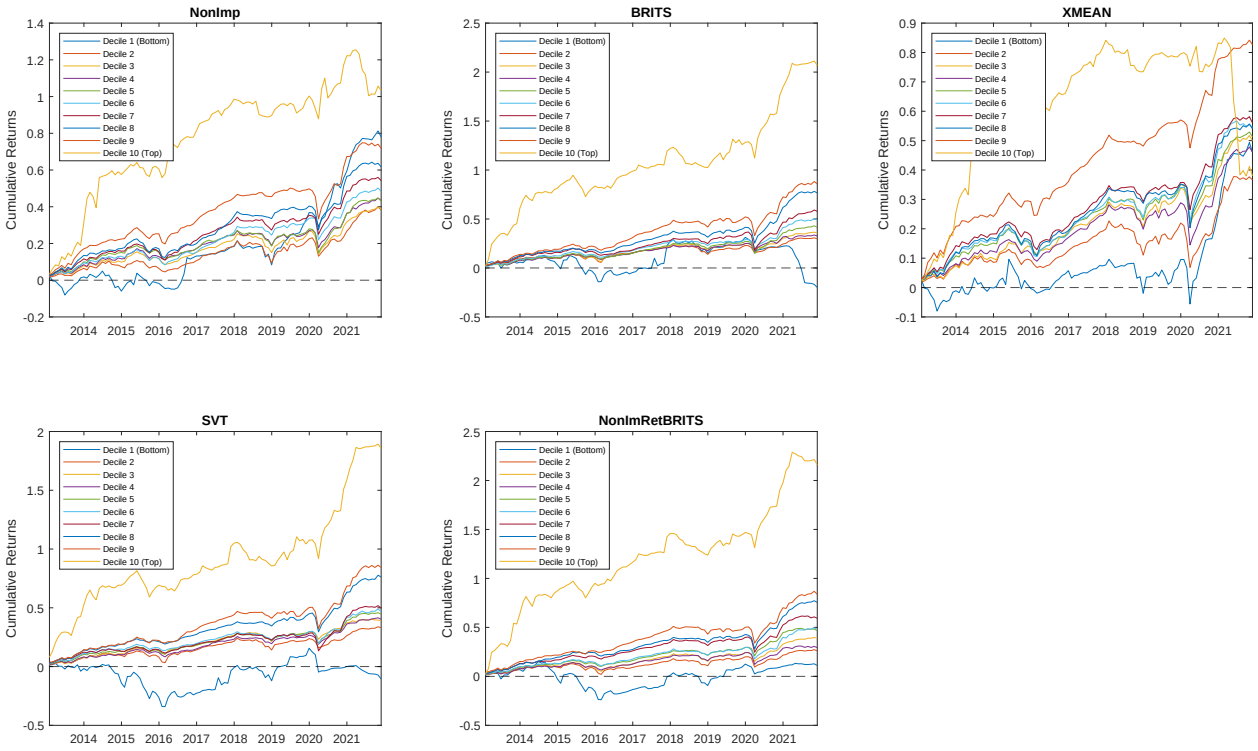


Figure 2. Prediction-weighting Decile Portfolio Cumulative Returns

The figure presents the out-of-sample realised cumulative returns of the neural network prediction-weighting decile portfolios for each imputation approach.

West (1987) standard errors.[14] Figure 3 presents the relevant performance persistence for each fund portfolio formed under the different imputation methods.

Interestingly, mean returns do not monotonically decrease over time but reveal cyclical phases. This pattern is more apparent for portfolios constructed on predictions generated by the BRITS and SVT imputed datasets. There is a monotonic decrease for up to around eight months, but the corresponding t-statistics report significant returns for up to five months for the BRITS imputed dataset. Interestingly, the mean returns recover and become again significant at around nine months, while they again at around 14 months and increase after 30 months, leading to a remarkable 11.83% at 36 months. Such a finding is important for hedge fund prediction as most hedge funds impose, on average, a three-month lock-up period on first-time investors Liang (2001). The only imputation method closely following BRITS

[14]We focus only the Chen et al. (2025) alphas due to lack of space. The relevant figures for the Fung and Hsieh (2004) alphas are similar and available upon request

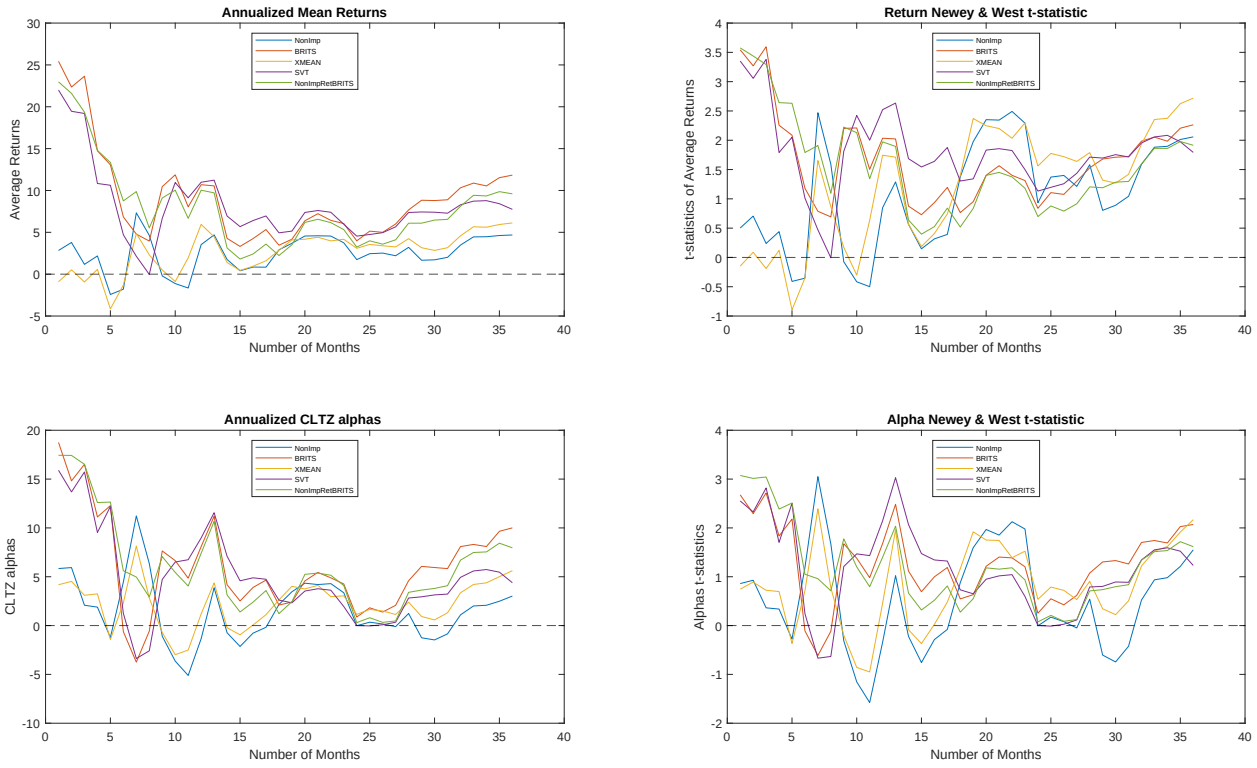


Figure 3. Prediction-weighted Top Decile Portfolio Performance Persistence

The figure presents the out-of-sample performance persistence of the NN prediction-weighted decile portfolios for each imputation approach. Every month, we sort funds into deciles based on the one-month NN predictions, and we hold the top-decile portfolio for up to 36 months with overlapping returns. We then calculate the annualized mean return, the CLTZ alpha and their corresponding Newey and West (1987) t-statistics.

is the SVT, but again its top-decile portfolio generates a mean return of 7.75% per annum at 36 months. The performance persistence of portfolios constructed on predictions of non-imputed and XMean imputed data is counter-cyclical of BRITS. Yet, BRITS imputed fund portfolios generate the highest mean return after three years compared to all benchmarks. The relevant alpha persistence results are similar to the mean returns, with BRITS yielding the highest alpha, but with statistical significance decreasing more quickly. Again, the BRITS fund prediction portfolio's annualized Chen et al. (2025) alpha and its corresponding t-statistic are 10% and 2.07, respectively, after three years. Except for CS-Mean, no other imputation method provides information that yields a statistically significant alpha (i.e., 5.61% and 2.17, respectively).

5.4. Style-Level Portfolio Performance

To examine the robustness and cross-sectional heterogeneity of our findings, we extend the analysis to individual hedge fund investment styles. This approach tests whether the predictive power of BRITS-imputed returns varies systematically across distinct investment mandates and provides granular insights into the sources of return predictability.

Within each Lipper TASS investment style, we impute missing returns using BRITS and apply our neural network forecasting framework to construct long-only portfolios. Portfolios are rank-weighted based on predicted next-month returns, ensuring that funds with higher expected performance receive proportionally larger allocations within their respective styles.

Table 6 presents the out-of-sample performance metrics for portfolios constructed using the BRITS-partial specification. The results reveal substantial economic and statistical significance across most investment styles. Multi-Strategy funds deliver the most compelling performance, generating an annualized excess return of 16.21% with alphas exceeding 10% relative to both the Fung and Hsieh (2004) seven-factor model and the CLTZ9 specification of Chen et al. (2025, CLTZ9). This exceptional performance likely reflects the flexibility of multi-strategy mandates to capitalize on predictive signals across diverse market segments. Emerging Markets and Long/Short Equity Hedge strategies similarly generate substantial alphas above 5%, indicating that imputation-enhanced forecasting is particularly effective for strategies with significant cross-sectional dispersion in returns.

Beyond mean returns, the portfolios exhibit markedly improved risk-adjusted performance. Several styles achieve Sharpe ratios exceeding unity, with Fund of Funds (1.44) and Multi-Strategy funds (1.31) demonstrating superior risk-return profiles. While certain strategies experience substantial drawdowns—most notably Convertible Arbitrage at −42%—the majority display maximum drawdowns comparable to or better than their respective unconditional benchmarks, suggesting that the enhanced predictability does not come at the expense of elevated tail risk.

Importantly, these performance patterns remain robust across alternative portfolio construction methodologies. Results using equal-weighted and prediction-weighted schemes yield qualitatively similar conclusions (untabulated results available upon request), confirming that the observed predictive content is not an artifact of the rank-weighting approach but rather reflects genuine information embedded in the imputed return series.

The style-level evidence demonstrates that sophisticated missing data imputation coupled with machine learning forecasting generates economically meaningful return predictability across a heterogeneous set of hedge fund strategies. These findings suggest that BRITS not only recovers individual fund-level dynamics but also uncovers systematic return patterns that persist across distinct investment mandates, highlighting the broader applicability of advanced imputation techniques for alternative asset management.

Table 6. Long-Only Portfolio Performance by Lipper TASS Investment Style

The table reports out-of-sample performance metrics for long-only portfolios of hedge funds formed by sorting based on the BRITS-partial predicted returns for the next month. Portfolios are constructed using rank weighting within each Lipper TASS investment style. FH7 and CLTZ9 are the Fung and Hsieh (2004) seven-factor and Chen et al. (2025, CLTZ9) nine-factor models, respectively. Significance of the mean excess return and FH7 and CLTZ9 alphas are based on t-statistics computed using robust standard errors; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

(1)	(2) Ann. mean excess return	(3) Ann. H7 α	(4) Ann. CLTZ9 α	(5) Maximum drawdown	(6) Ann. Sharpe ratio
Rank weighting					
Convertible Arbitrage	12.18%	1.95%	−1.38%	42.06%	0.42
Dedicated Short Bias	8.13%***	3.04%	2.91%	9.11%	0.95
Event Driven	9.42%***	3.99%**	5.07%**	11.83%	1.01
Emerging Markets	10.35%***	4.98%***	7.07%***	9.77%	1.31
Equity Market Neutral	12.67%***	3.83%	4.82%	18.37%	1.05
Fixed Income Arbitrage	15.76%***	10.58%*	13.90%**	19.10%	0.95
Fund of Funds	13.93%***	7.26%***	6.52%***	18.24%	1.44
Global Macro	7.67%***	1.88%	3.08%	15.87%	0.83
Long/Short Equity Hedge	11.79%***	5.37%***	4.82%***	17.54%	1.24
Multi-Strategy	16.21%***	11.81%***	10.88%***	10.08%	1.31

5.5. Predictor Importance

This section investigates the importance of fund-specific predictors and the interactions of those predictors with the macroeconomic variables in the prediction of hedge fund returns. We focus on the predictions generated from the neural network using information from the fund data recovered with BRITS because of its superiority in both return prediction and selection of top-performing funds against the benchmarks, as presented in the previous sections.[15]. We use the Shapley Additive exPlanations (SHAP) methodology of Lundberg and Lee (2017) to calculate the SHAP values of hedge fund return predictions. The SHAP method is based on a cooperative game theory framework, which aggregates the SHAP values across all predictors to measure the contribution of each predictor to the forecasting exercise. The method considers the fluctuations in the forecasting model output (i.e., the prediction) with and without including a specific predictor while retaining the rest of the predictors. The predictions from the two models are then compared, and the prediction difference is calculated. The SHAP values are estimated as the weighted average of all possible differences for each predictor Lundberg and Lee (2017).

Figure 4 provides the SHAP values for the ten most important fund predictors and the important interactions of predictors with macroeconomic variables (i.e., the ten highest SHAP values) for predicting hedge fund returns. Our findings focus on the prediction results of the neural network ensemble. In particular, we calculate the SHAP values for each neural network specification separately and then compute the average SHAP values for each predictor across the three corresponding values of each neural network model.[16] We estimate each predictor's and each interaction's SHAP values as the average of the absolute SHAP value across all out-of-sample observations. We evaluate the predictors' importance on the last out-of-sample window covering January 2021 to November 2021. Figure 4 reports the relevant importance ranking.

[15]The relevant predictor importance findings for the remaining imputation methods are available upon request

[16]The exact model structure affecting the output is essential in calculating the SHAP values.

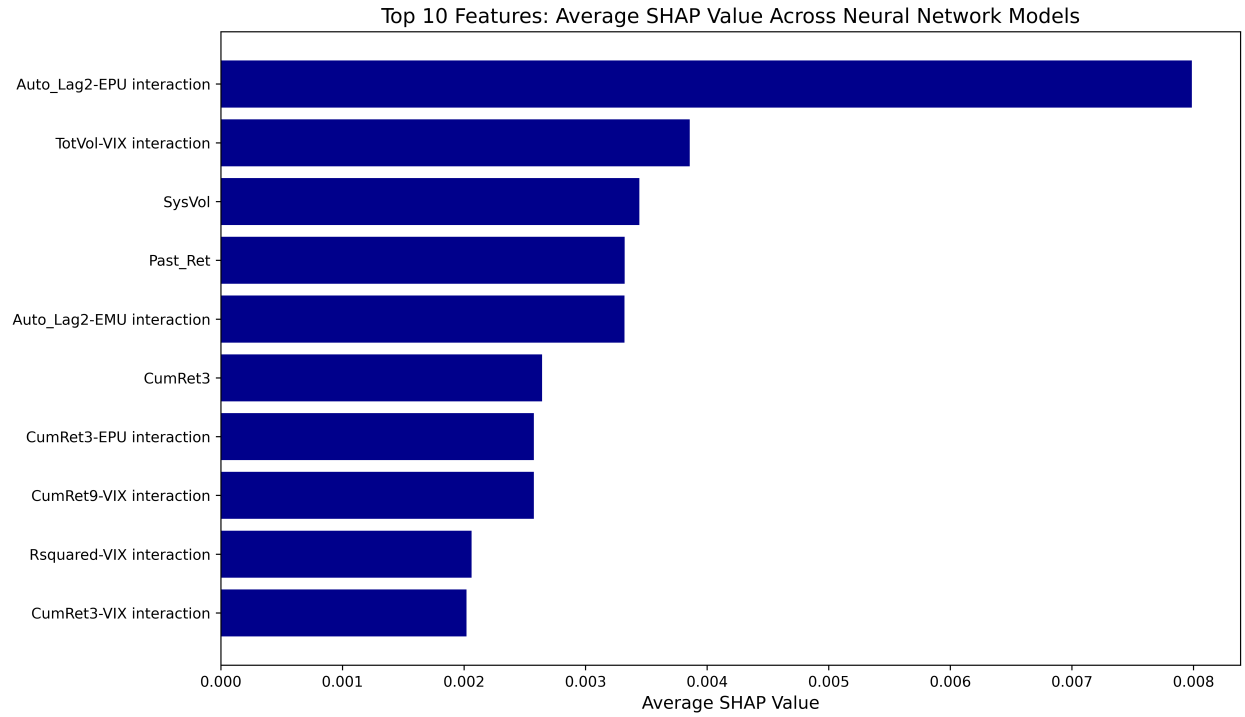


Figure 4. Predictor Importance - Top 10 Variables

The figure displays the SHAP values of the fund-specific predictors and their interactions with the macroeconomic variables out-of-sample under the neural network ensemble forecasting and the BRITS imputation method. The SHAP values are generated separately for each predictor in each neural network model, and the average SHAP values across the three neural network models are then computed for each predictor. The presented SHAP values correspond to the forecasting exercise using BRITS-imputed values.

We observe from Figure 4 that the interactions of fund-specific predictors with macroeconomic variables dominate the standalone fund-specific predictors regarding their importance in predicting hedge fund returns. There are only a few cases in which fund-specific predictors are among the most important (e.g., `SysVol`, `CRet_3mo`, `Past_Ret_1mo`). For example, the most important predictive variable is the interaction of the autocorrelation of the second lag with the Economic policy uncertainty index (`EPU`), followed by the interaction of total volatility with the VIX index. The fund's systematic volatility `SysVol` is the third most important predictor, justifying its significance in predicting hedge fund returns as found by Bali, Brown, and Caglayan (2012). The remaining of importance ranking is filled mainly with predictor interactions of cumulative returns (i.e., `CRet_3mo`, `CRet_9mo`) and the fund's

R-squared (`Rsq`) with the Economic policy uncertainty and the VIX indices. The above findings reveal the importance of predictor interactions in the prediction task, especially when using robust machine learning methodologies, such as a deep neural network, to capture nonlinearities in the dataset. We can also conclude that the most important fund-specific predictors (no interactions) of hedge fund returns (i.e., `SysVol`, `Ret_1mo`, `CRet_3mo`) belong in the family of return-based characteristics such as, previous returns, previous return autocorrelations and return moments, but no one from the family of predictors measuring managerial skill.

6. Conclusion

Missing values are pervasive in hedge fund data, which makes predicting hedge fund performance even more exacting than it inherently is. To address the challenge of missing values in hedge fund data, we employ BRITS, a complex recurrent neural network for data imputation. BRITS utilizes information from both past and future observations in the time-series dimension as well as contemporaneous observations in the cross-sectional dimension. We use BRITS to fill in missing values when training neural networks for predicting hedge fund returns based on a large number of fund characteristics. We find that BRITS is valuable for improving out-of-sample hedge fund prediction compared to alternative methods for dealing with missing data when training prediction models, including ignoring missing values by only using observations for funds with no missing values for their returns and characteristics, and imputing missing data via the CS-Mean. In light of the bias-variance trade-off, it is not obvious that a complex method like BRITS will outperform simpler methods, as more complex methods increase the risk of overfitting when imputing missing data. Nevertheless, when it comes to hedge fund data, we find that BRITS harnesses information from a variety of angles to impute missing data in a manner that avoids overfitting and substantially improves out-of-sample return prediction.

It would be interesting in future research to explore the use of BRITS to deal with missing data in other finance applications. This would help to establish if BRITS is of widespread value in finance when encountering missing data, or whether there are unique features of missing hedge fund data that make BRITS particularly useful for data imputation for this particular asset class.

References

Aragon, G. O. (2007). "Share Restrictions and Asset Pricing: Evidence from the Hedge Fund Industry." *Journal of Financial Economics* 83:1, 33–58.

Avramov, D., L. Barras, and R. Kosowski (2013). "Hedge Fund Return Predictability under the Magnifying Glass." *Journal of Financial and Quantitative Analysis* 48:4, 1057–1083.

Avramov, D., R. Kosowski, N. Y. Naik, and M. Teo (2011). "Hedge Funds, Managerial Skill, and Macroeconomic Variables." *Journal of Financial Economics* 99:3, 672–692.

Baker, S. R., N. Bloom, and S. J. Davis (2016). "Measuring Economic Policy Uncertainty." *Quarterly Journal of Economics* 131:4, 1593–1636.

Bali, T. G., S. J. Brown, M. O. Caglayan, and U. Celiker (2021). "Does Industry Timing Ability of Hedge Funds Predict Their Future Performance, Survival, and Fund Flows?" *Journal of Financial and Quantitative Analysis* 56:6, 2136–2169.

Bali, T. G., S. J. Brown, and M. O. Caglayan (2012). "Systematic Risk and the Cross Section of Hedge Fund Returns." *Journal of Financial Economics* 106:1, 114–131.

Bali, T. G., S. J. Brown, and M. O. Caglayan (2019). "Upside Potential of Hedge Funds as a Predictor of Future Performance." *Journal of Banking and Finance* 98:1, 212–229.

Beckmeyer, H. and T. Wiedemann (2025). "Empirical Asset Pricing with Missing Data." SSRN Working Paper 4003455.

Bryzgalova, S., S. Lerner, M. Lettau, and M. Pelger (2025). "Missing Financial Data." *Review of Financial Studies* 38:3, 803–882.

Cai, J.-F., E. J. Candès, and Z. Shen (2010). "A Singular Value Thresholding Algorithm for Matrix Completion." *SIAM Journal on Optimization* 20:4, 1956–1982.

Campbell, J. Y. and S. B. Thompson (2008). "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" *Review of Financial Studies* 21:4, 1509–1531.

Cao, W., D. Wang, J. Li, H. Zhou, Y. Li, and L. Li (2018). "BRITS: Bidirectional Recurrent Imputation for Time Series." In: *Proceedings of the 31st Conference on Advances in Neural Information Processing Systems*. Montreal, Canada, pp. 6776–6786.

Chen, A. Y. and J. McCoy (2024). "Missing Values Handling for Machine Learning Portfolios." *Journal of Financial Economics* 155:1, 103815.

Chen, A. Y. and T. Zimmermann (2022). "Open Source Cross-Sectional Asset Pricing." *Critical Finance Review* 27:2, 207–264.

Chen, L., M. Pelger, and J. Zhu (2024). "Deep Learning in Asset Pricing." *Management Science* 70:2, 714–750. eprint: https://doi.org/10.1287/mnsc.2023.4695.

Chen, Y., B. Han, and J. Pan (2021). "Sentiment Trading and Hedge Fund Returns." *Journal of Finance* 76:4, 2001–2033.

Chen, Y., Z. Li, Y. Tang, and G. Zhou (2025). "Anomalies as New Hedge Fund Factors." *Journal of Financial and Quantitative Analysis* 60:8, 3660–3693.

Diebold, F. X. and R. S. Mariano (1995). "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics* 13:3, 253–263.

Fama, E. F. and K. R. French (1989). "Business Conditions and the Expected Returns on Stocks and Bonds." *Journal of Financial Economics* 25:1, 23–49.

Filippou, I., D. E. Rapach, M. P. Taylor, and G. Zhou (2024). "Economic Fundamentals and Short-Run Exchange Rate Prediction: A Machine Learning Perspective." SSRN Working Paper 3455713).

Freyberger, J., B. Höppner, A. Neuhierl, and M. Weber (2025). "Missing Data in Asset Pricing Panels." *Review of Financial Studies* 38:3, 760–802.

Freyberger, J., A. Neuhierl, and M. Weber (2020). "Dissecting Characteristics Nonparametrically." *Review of Financial Studies* 33:5, 2326–2377.

Fung, W. and D. A. Hsieh (2000). "Performance Characteristics of Hedge Funds and Commodity Funds: Natural vs. Spurious Biases." *Journal of Financial and Quantitative Analysis* 35:3, 291–307.

Fung, W. and D. A. Hsieh (2004). "Hedge Fund Benchmarks: A Risk-Based Approach." *Financial Analysts Journal* 60:5, 65–80.

Giglio, S., Y. Liao, and D. Xiu (2021). "Thousands of Alpha Tests." *Review of Financial Studies* 34:7, 3456–3496.

Green, J., J. R. M. Hand, and X. F. Zhang (2017). "The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns." *Review of Financial Studies* 30:12, 4389–4436.

Gu, S., B. Kelly, and D. Xiu (2020). "Empirical Asset Pricing via Machine Learning." *Review of Financial Studies* 33:5, 2223–2273.

Han, Y., A. He, D. E. Rapach, and G. Zhou (2024). "Cross-Sectional Expected Returns: New Fama–MacBeth Regressions in the Era of Machine Learning." *Review of Finance* 28:6, 1807–1831.

Haugen, R. A. and N. L. Baker (1996). "Commonality in the Determinants of Expected Stock Returns." *Journal of Financial Economics* 41:3, 401–439.

Hendrycks, D. and K. Gimpel (2023). "Gaussian Error Linear Units (GELUs)." Working Paper arXiv:1606.08415.

Heuson, A. J., M. C. Hutchinson, and A. Kumar (2020). "Predicting Hedge Fund Performance When Fund Returns Are Skewed." *Financial Management* 49:4, 877–896.

Hou, K., C. Xue, and L. Zhang (2015). "Digesting Anomalies: An Investment Approach." *Review of Financial Studies* 28:3, 650–703.

Kaniel, R., Z. Lin, M. Pelger, and S. Van Nieuwerburgh (2023). "Machine-Learning the Skill of Mutual Fund Managers." *Journal of Financial Economics* 150:1, 94–138.

Kingma, D. P. and J. Ba (2017). "Adam: A Method for Stochastic Optimization." Working Paper arXiv:1412.6980v9.

Liang, B. (2001). "Hedge Fund Performance: 1990-1999." *Financial Analysts Journal* 57:1, 11–18. ISSN: 0015198X.

Light, N., D. Maslov, and O. Rytchkov (2017). "Aggregation of Information About the Cross Section of Stock Returns: A Latent Variable Approach." *Review of Financial Studies* 30:4, 1339–1381.

Lundberg, S. M. and S.-I. Lee (2017). "A Unified Approach to Interpreting Model Predictions." In: *Proceedings of the 31st Conference on Neural Information Processing Systems*. Long Beach, CA, pp. 4768–4777.

Newey, W. K. and K. D. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55:3, 703–708.

Simon, N., J. Friedman, T. Hastie, and R. Tibshirani (2013). "A Sparse-Group Lasso." *Journal of Computational and Graphical Statistics* 22:2, 231–245.

Tibshirani, R. (1996). "Regression Shrinkage and Selection via the LASSO." *Journal of the Royal Statistical Society. Series B (Methodological)* 58:1, 267–288.

Titman, S. and C. Tiu (2011). "Do the Best Hedge Funds Hedge?" *Review of Financial Studies* 24:1, 123–168.

West, K. D. (1996). "Asymptotic Inference About Predictive Ability." *Econometrica* 64:5, 1067–1084.

Wu, W., J. Chen, Z. Yang, and M. L. Tindall (2021). "A Cross-Sectional Machine Learning Approach for Hedge Fund Return Prediction and Selection." *Management Science* 67:7, 3985–4642.

Zou, H. (2006). "The Adaptive Lasso and Its Oracle Properties." *Journal of the American Statistical Association* 101:476, 1418–1429.

Zou, H. and T. Hastie (2005). "Regularization and Variable Selection via the Elastic Net." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67:2, 301–320.

Internet Appendix to

"Improving Hedge Fund Return Prediction: Dealing with Missing Data via Deep Learning"

(Not for publication)

A. BRITS Tuning Parameters

Similar to most neural network architectures, the weights of BRITS are optimized by the gradient descent iterative algorithm. In the BRITS methodology, the loss function, which is based on the MAE criterion, is minimized by the optimization process. At each step of the optimization, we evaluate the gradients and then update the BRITS weights in the opposite direction of the gradient (**BeckmeyerWiedemann2024**; Gu, Kelly, and Xiu 2020)). The practical implementation of BRITS requires hyperparameter optimization. The process involves deciding the highest-performing neural network specification for all candidate architectures. In our study, we examine 16 candidate architectures exploring different values for the number of hidden neurons, epochs, early stopping patience, and learning rates. We display the hyperparameter search space in Table A1.

Table A1. BRITS Hyperparameters. This table presents the different hyperparameter values for the candidate BRITS architectures. The optimized BRITS is the combination of parameters that achieves the lowest imputation error for the hedge fund returns and predictors datasets.

Tuning Parameters	Candidate Architectures
Number of hidden neurons	64, 128
Epochs	50, 200
Early stopping	5, 25
Learning rate	10^{-6}, 10^{-3}
Batch size	12
Epochs	Adam

For the number of hidden neurons, we explore two values in the power of two, as is common in the relevant literature (see, (Cao et al. 2018; Kaniel et al. 2023). Hence we use 64 and 128 hidden neurons. For the number of epochs, we experimented with a smaller number (i.e., 50 epochs) and a larger number (i.e., 200 epochs). In terms of early stopping, we follow Gu, Kelly, and Xiu (2020), and we apply five epochs of patience, as well as an even larger value of 25. As for the learning rate and optimizer, again, we follow Cao et al. (2018) and Gu, Kelly, and Xiu (2020), and we apply learning rate values of 10^{-6} and 10^{-3} and the Adam optimizer respectively. The batch size equals one year (i.e., 12 months). The

optimal architecture achieves the lowest imputation error, and it is the one we use for the remainder of our study.

B. Additional Simulation Results

In this section, we present the imputation simulation results of hedge fund predictors datasets. Tables A2 and A3 provide the BRITS and benchmark imputation methods' RMSE for each predictor separately and for the 10% and 20% simulations, respectively.

The above results reveal that our proposed model achieves the smallest imputation error for most predictors, specifically for 17 out of the 23 predictors for the 10% simulation study and 19 out of 23 for the 20% simulation study. SVT is the most accurate imputation method for the remaining predictors, while BRITS always holds the second-best position. Additional significant insights are derived from the above tables by comparing the imputation error's minimum and maximum values across all predictors. For example, the corresponding RMSE for BRITS ranges from 0.010 to 18.978 for the 10% simulation study. Its range is much higher than that of the second-best SVT (i.e., 0.015 – 29.889) for SVT. Those findings suggest that BRITS has the most consistent imputing ability with smaller performance deviations than all benchmarks.

C. Neural Network Tuning Parameters

Table A4 reports the tuning parameters for the neural network specifications. We follow the relevant literature to set the tuning parameters (see, (Gu, Kelly, and Xiu 2020; Chen, Pelger, and Zhu 2024; Kaniel et al. 2023)). We consider specifications with three variations concerning the layers (i.e., three, four and five). The number of neurons in each layer is half of those in each layer (e.g., 32, 16, 8 for a network with three layers). We also perform an estimation averaging over five model re-estimations to obtain stable and robust estimations. In such a way, we decrease the effect of a local suboptimal fit and it reduces the estimation

Table A2. Predictors' Imputation Fidelity - 10% Simulation

The table reports the imputation simulation results by artificially dropping 10% of observed values for each of the 23 predictors examined in our research. For all predictors, we present RMSE criterion. We also report the minimum (Min) and maximum (Max) RMSE values for each imputation method across all predictors.

Predictor	TSMEAN	XMEAN	SVT	BRITS
Panel A: Previous returns				
Ret_1mo	8.145	7.910	7.958	7.700
CRet_3mo	10.764	10.125	10.090	6.936
CRet_6mo	13.391	11.755	11.933	5.898
CRet_9mo	17.386	16.094	16.284	8.806
CRet_12mo	21.729	20.714	20.988	9.861
CRet_36mo	27.074	28.375	28.255	15.797
Panel B: Previous return autocorrelations				
AC_Lag1	0.311	0.312	0.175	0.174
AC_Lag2	0.322	0.316	0.175	0.173
AC_Lag3	0.349	0.337	0.185	0.183
Panel C: Return moments				
Vol	3.596	4.843	4.001	1.587
IdioVol	3.508	4.549	3.972	1.535
SysVol	0.496	0.866	0.065	0.152
CoSkew	0.037	0.041	0.015	0.011
IdioSkew	0.666	0.839	0.213	0.223
Skew	0.771	0.930	0.224	0.252
Kurt	2.405	2.942	1.512	1.711
Panel D: Managerial skill				
AlphaFH7_12mo	3.970	3.969	3.736	3.311
AlphaFH7_36mo	1.106	1.352	1.030	0.263
AlphaBBCC9_24mo	1.181	1.257	0.677	0.262
AlphaCHP11_36mo	32.200	29.995	29.886	18.978
Rsq	0.127	0.196	0.032	0.110
AUM	0.976	1.615	0.346	7.102
MaxRet	8.063	8.991	8.072	3.607
Panel E: Min-Max Error of All Predictors				
Min	0.037	0.042	0.015	0.010
Max	32.200	29.995	29.886	18.978

variance of the estimated model. We use the Gaussian Error Linear Unit (GELU) as the activation function, and we apply L1 (LASSO) and L2 (Ridge) regularisations to penalise large weights in our neural networks and prevent overfitting. We set a weight of 10^{-3} for

Table A3. Predictors' Imputation Fidelity - 20% Simulation

The table reports the imputation simulation results by artificially dropping 20% of observed values for each of the 23 predictors examined in our research. For all predictors, we present RMSE criterion. We also report the minimum (Min) and maximum (Max) RMSE values for each imputation method across all predictors.

Predictor	TSMEAN	XMEAN	SVT	BRITS
Panel A: Previous returns				
Ret_1mo	6.567	6.270	6.349	6.027
CRet_3mo	9.760	8.934	9.181	5.964
CRet_6mo	14.315	13.056	13.610	7.042
CRet_9mo	20.835	20.107	20.703	8.887
CRet_12mo	20.404	19.020	19.857	9.186
CRet_36mo	30.008	32.009	32.414	15.465
Panel B: Previous return autocorrelations				
AC_Lag1	0.312	0.313	0.181	0.181
AC_Lag2	0.323	0.316	0.184	0.183
AC_Lag3	0.349	0.337	0.194	0.194
Panel C: Return moments				
Vol	3.072	4.134	3.289	1.582
IdioVol	2.974	3.845	3.91	1.443
SysVol	0.500	0.873	0.074	0.158
CoSkew	0.038	0.043	0.016	0.011
IdioSkew	0.0664	0.830	0.237	0.236
Skew	0.771	0.929	0.253	0.276
Kurt	2.372	2.937	1.788	1.687
Panel D: Managerial skill				
AlphaFH7_12mo	3.880	3.870	3.663	3.433
AlphaFH7_36mo	0.972	1.157	0.804	0.279
AlphaBBCC9_24mo	1.319	1.387	0.924	0.325
AlphaCHP11_36mo	89.153	88.538	88.457	85.529
Rsq	0.127	0.197	0.033	0.110
AUM	0.986	1.619	0.448	7.032
MaxRet	6.966	7.605	6.810	3.142
Panel E: Min-Max Error of All Predictors				
Min	0.038	0.043	0.016	0.011
Max	89.153	85.538	88.457	85.529

both regularisations. Using the Adam optimiser, we train the network for 100 epochs for each combination of tuning parameters.

Table A4. Model Specifications for Neural Networks

Model Type	Model Specification
Neural Network Models	
Deep Neural Network (3 Hidden Layers) Neurons: [32, 16, 8]	Activation function: GELU Batch size: 10^4 Epochs: 100 Optimizer: Adam Early stopping: 5 epochs Learning rate: 10^{-3} L1, L2 regularization weight: 10^{-3}
Deep Neural Network (4 Hidden Layers) Neurons: [32, 16, 8, 4]	
Deep Neural Network (5 Hidden Layers) Neurons: [32, 16, 8, 4, 2]	

D. Additonal Empirical Results: LASSO-based Models

D.1. Linear-LASSO Models

D.1.1 LASSO model

We first consider a linear specification, where the model parameters are estimated using LASSO (Least Absolute Shrinkage and Selection Operator) Tibshirani (1996). LASSO applies ℓ_1 regularization, which both shrinks coefficient estimates and performs variable selection by forcing some coefficients to be exactly zero. The forecast for hedge fund excess returns using the LASSO model, omitting the intercept for simplicity, is expressed as:

$$\hat{r}_{i,t+1}^{\text{Lasso}} = \boldsymbol{x}_{i,t}' \hat{\boldsymbol{\eta}}, \tag{22}$$

where $\hat{\boldsymbol{\eta}}$ is the coefficient vector estimated via LASSO based on the data up to month t. Ordinary least squares (OLS) focuses on minimizing residuals but can suffer from overfitting, especially when predictors are highly correlated or when there is a low signal-to-noise ratio. LASSO regularization addresses this by shrinking the coefficients and performing variable selection.

The LASSO objective function is:

$$\arg\min_{\boldsymbol{\eta}} \frac{1}{2(t-1)n} \sum_{i=1}^{n} \sum_{s=1}^{t-1} \left(r_{i,s+1} - \boldsymbol{x}'_{i,s}\boldsymbol{\eta}\right)^2 + \lambda \|\boldsymbol{\eta}\|_1, \tag{23}$$

where λ is the regularization parameter that controls the degree of shrinkage applied to the coefficients, and $\|\boldsymbol{\eta}\|_1 = \sum_{j=1}^{k} |\eta_j|$ is the ℓ_1 norm of the coefficient vector $\boldsymbol{\eta}$. When $\lambda = 0$, LASSO reduces to the standard OLS estimation. The optimal value of λ is typically selected using cross-validation. We use a validation dataset to decide on the optimal value of λ and following Gu, Kelly, and Xiu (2020) our search space is $\{10^{-4}, 10^{-1}\}$.

D.1.2 Group LASSO Model

Yuan and Lin (2006) suggest the following Group Lasso estimator:

$$\hat{\theta}^* = \underset{\theta_0,\theta}{\operatorname{argmin}} \left(\|r - X^T\theta\|_2^2 + \lambda \sum_{\xi=1}^{K} \sqrt{d_\xi} \|\theta^\xi\|_2 \right)$$

where, K is the number of categories the predictors are divided into, the term $\sqrt{d_\xi}$ weights each category according to its size and d_ξ is the size of the ξ category, θ^ξ is a sub-vector of coefficients from θ with components that correspond to the covariates in ξ category. We use a validation dataset to decide on the optimal value of λ and following Gu, Kelly, and Xiu (2020) our search space is $\{10^{-4}, 10^{-1}\}$. In Section 3 of our main paper, we mention that our predictors can be divided into 3 major categories (i.e., past returns and autocorrelation, second and higher moments, and skill of hedge fund managers). We also introduce a fourth category by including macroeconomic factors and their interactions with the predictors (i.e., $K = 4$). We use a validation dataset to decide on the optimal value of λ and following Gu, Kelly, and Xiu (2020) our search space is $\{10^{-4}, 10^{-1}\}$.

D.1.3 Sparce Group LASSO Model

The Sparse Group Lasso model was introduced by Friedman et al. (2010) and combines the Lasso and Group Lasso penalization under the following mathematical formulation:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \left(\|r - X^T\theta\|_2^2 + \alpha\lambda \sum_{d=1}^{D} |\theta|_1 + (1-\alpha)\lambda \sum_{\xi=1}^{K} \sqrt{d_\xi} \|\theta^\xi\|_2 \right)$$

where, α is bounded in $[0, 1]$ and controls the penalization between Lasso and Group Lasso. We use a validation dataset to decide on the optimal value of λ, and following (Gu et al., 2020) our search space is $\{10^{-4}, 10^{-1}\}$. Regarding the tunable parameter α, we explore the following values: $\alpha \in \{0, 0.25, 0.5, 0.75, 1\}$.

D.1.4 Adaptive LASSO Model

The Adaptive Lasso was introduced in Zou (2006). The authors suggest that the model satisfies the so-called *oracle* properties and performs as well as if the true model was provided in advance. The Adaptive Lasso estimator is the following:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \left(\|r - X^T\theta\|_2^2 + \lambda \sum_{d=1}^{D} \hat{w}|\theta|_1 \right)$$

where $\hat{w} = 1/|\theta|^\gamma$ is a weight corresponding to θ coefficient. We use partial square regression (PLS) to determine $\hat{w}$. This is because the large number of variables in our pool of predictors can be correlated. To specify the desired amount of variability in X (i.e., the matrix of predictors) explained to determine the number of used PLS components, we select a number of PLS components that can explain 90% of the variability in our predictors' matrix. For the λ and γ tunable parameters, we again use a validation dataset to decide on the optimal combination given the defined search spaces. For λ this is $\{10^{-4}, 10^{-1}\}$ and for γ is $\{0, 8, 1\}$. All the Lasso-based model specifications and tuning parameters are presented in Table A5 below.

Table A5. Model Specifications for LASSO Methods

Model Type	Model Specification
Penalized Regressions	
Lasso	$\lambda \in \{10^{-4}, 10^{-1}\}$
Sparse Group Lasso	$\lambda \in \{10^{-4}, 10^{-1}\}$ $\alpha \in [0, 0.25, 0.5, 1]$
Adaptive Lasso	$\lambda \in \{10^{-4}, 10^{-1}\}$ $\gamma \in \{0, 8, 1\}$

D.2. Out-of-sample Portfolio Performance

We present a selection of key empirical findings comparing the informational value of the BRITS method against benchmarks, using linear LASSO as our prediction method. We begin with the performance report for the top-decile prediction-weighted portfolios constructed from LASSO ensemble predictions. These were generated by training the LASSO models on information from each imputation specification and from no imputation. We continue by presenting the cumulative returns of their corresponding decile portfolios.

Table A6 reports the relevant results for the equal-, rank- and prediction-weighted portfolios of top-performing funds as given by the LASSO predictions. Similar to the relevant results for deep neural network predictions, BRITS *informed* portfolios achieve the highest performance among all benchmark imputation methods and the no-imputation case, especially for the case of prediction-weighting top-decile portfolios. The BRITS *informed* prediction-weighting top-decile portfolio, based on both fund returns' and predictors' recovered data, yields an annualised mean return of 22.12% and a Fung and Hsieh (2004) and a Chen et al. (2025) alpha of 13.41% and 17.72%, all statistically significant at least at 5%. The same portfolio reports an outstanding Sharpe ratio of 1.23. For the remaining portfolio weighting schemes, the difference in economic value between BRITS and the other imputation methods is smaller, likely because the LASSO model's constraint on accounting for nonlinearities in fund returns and characteristics yields less accurate predictions. Interestingly, the second-best imputation methods are CS-Mean and SVT, which achieve similar

performance, and sometimes close to that of BRITS approaches, especially for equal- and rank-weighting portfolios, again due to the linear nature of the LASSO specifications.

Table A6. Top-decile portfolio performance: LASSO Ensemble

The table reports out-of-sample performance metrics for long-short portfolios of hedge funds formed by sorting based on the LASSO-Ensemble hedge fund excess return forecasts for the next month. The long-short portfolio goes long (short) funds in the top (bottom) decile. Results are reported for the different data imputation methods in the column headings when training the neural networks: complete cases (i.e., no data imputation), cross-sectional mean (CS-Mean), singular value thresholding (SVT), bidirectional recurrent imputation network for time series (BRITS), and BRITS excluding return imputation (BRITS-Partial). FH7 and CLTZ9 are the Fung and Hsieh (2004) seven-factor and Chen et al. (2025, CLTZ9) nine-factor models, respectively. Significance of the mean excess return and FH7 and CLTZ9 alphas are based on t-statistics computed using robust standard errors; *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Panel A, B, and C report results for equal-, rank-, and prediction-weighting, respectively.

(1)	(2)	(3)	(4)	(5)	(6)
Performance metric	Complete cases	CS-Mean	SVT	BRITS	BRITS-Partial
Panel A. Equal weighting					
Ann. mean excess return	11.46%***	11.96%***	12.28%***	13.22%***	12.53%***
Ann. FH7 α	3.29%	5.24%***	5.64%***	5.79%***	5.61%**
Ann. CLTZ9 α	4.54%*	7.13%***	6.99%***	7.18%***	4.52%***
Maximum drawdown	13.94%	11.13%	11.13%	13.28%	16.48%
Ann. Sharpe ratio	1.13	1.26	1.28	1.27	1.11
Panel B. Rank weighting					
Ann. mean return	13.79%***	15.39%***	15.76%***	16.48%***	15.88%***
Ann. FH7 α	4.54%*	7.91%***	7.82%***	8.69%***	7.15%***
Ann. CLTZ9 α	6.07%*	10.56%***	9.58%***	10.29%***	9.18%***
Maximum drawdown	16.82%	14.23%	12.83%	14.06%	16.80%
Ann. Sharpe ratio	1.17	1.46	1.42	1.47	1.31
Panel C. Prediction weighting					
Ann. mean return	19.28%***	20.73%***	19.69%***	22.12%***	23.96%***
Ann. FH7 α	9.76%*	12.55%**	11.24%**	13.41%**	14.43%**
Ann. CLTZ9 α	13.26%**	19.25%***	15.41%***	17.72%***	19.06%***
Maximum drawdown	24.85%	22.78%	19.57%	22.29%	25.66%
Ann. Sharpe ratio	1.22	1.24	1.17	1.23	1.19

Figure A1 presents the corresponding cumulative returns of the decile portfolios under each imputation method and no imputation. BRITS and SVT can provide information that enables the construction of long-short portfolios with the highest dispersion relative to the

other imputation methods. Such a finding justifies BRITS' ability to provide a *complete* dataset for predicting and accurately distinguishing between the best- and worst-performing funds. The BRITS method invests in the best-performing funds, yielding a cumulative return of 224%, while avoiding the worst-performing funds, yielding a return of 27%. The corresponding returns for the SVT and CS-Mean methods are 199% and 45% and 214% and 41%, respectively. Finally, the top decile portfolio's cumulative return is 203%, while the bottom decile's is 84% for LASSO predictions without any imputation methods.

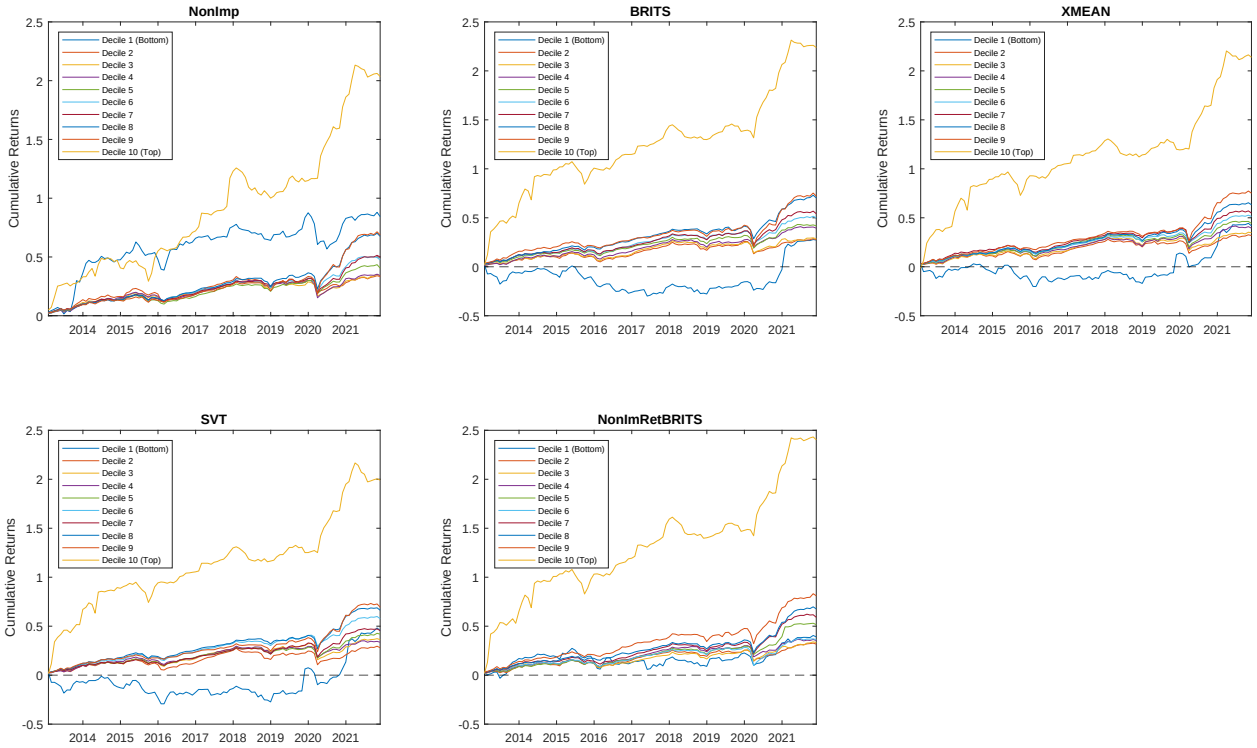


Figure A1. Prediction-weighted Decile Portfolio Cumulative Returns

The figure presents the out-of-sample cumulative realized returns of the equally-weighted decile portfolios or each imputation approach based on LASSO predictions.