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

Ilias Filippou
Florida State University
ifilippou@fsu.edu

Ioannis Psaradellis
University of Edinburgh
ipsarade@ed.ac.uk

David E. Rapach*
Federal Reserve Bank of Atlanta
dave.rapach@gmail.com

Lazaros Zografopoulos
University of St. Andrews
lazografopoulos@gmail.com

February 16, 2025

*Corresponding author. Send correspondence to David Rapach, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street NE, Atlanta, GA 30309; email: dave.rapach@gmail.com.

We are grateful to seminar participants at the University of Edinburgh for insightful comments. The views expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

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

Abstract

We study the critical issue of handling missing entries in hedge fund data. We introduce a deep learning approach, the BRITS, for recovering data for fund returns and 23 fund predictors. We compare its performance with popular imputation methods, such as the cross-sectional mean and singular value thresholding. BRITS' ability to capture information from past and future values in time series and the whole cross-section of observations yields the highest imputation fidelity in our simulations. The recovered information improves predictions of nonlinear and linear methods. At the same time, it helps to select top-performing funds that earn significant out-of-sample annual alphas of 13.4% net of all costs.

JEL classifications: G12, G14, G23.

Keywords: Hedge funds, machine learning, missing values, BRITS, imputation.

1. Introduction

A common problem in financial datasets, especially those for hedge funds, is the presence of missing values. Missing values present challenges for a number of topics in asset pricing, including model estimation and out-of-sample return forecasting. Thus, effective imputation of missing data has the potential to substantially improve empirical asset pricing, including out-of-sample asset return prediction and investment decisions. Despite its importance, the issue of missing values in financial datasets has only recently started to receive significant attention in the finance literature (e.g., Giglio, Liao, and Xiu 2021; Freyberger et al. 2025; Bryzgalova et al. 2025; Beckmeyer and Wiedemann 2023; Chen and McCoy 2024).

We examine the imputation of missing values in hedge fund datasets, a critical issue arising from the voluntary nature of hedge fund disclosures. Missing data can distort empirical analyses, affecting return prediction, fund selection, and broader asset pricing research. We assess imputation methods for hedge fund returns and characteristics, evaluating their accuracy and impact on out-of-sample forecasting. Hedge funds' flexibility and limited regulation often lead to incomplete reporting, complicating performance evaluation. Their diverse strategies and proprietary approaches further obscure true managerial skill, making fund selection challenging. Addressing missing data is essential for improving predictive accuracy and empirical asset pricing.

To deal with missing data in hedge fund datasets, we employ a deep learning model that considers both time-series and cross-sectional properties to impute missing values for hedge fund returns and a large set of fund characteristics. Specifically, we use the *bidirectional recurrent imputation network for time series* (BRITS, Cao et al. 2018) method that applies to general settings for missing data. A number of properties of BRITS make it attractive for filling in missing values for hedge fund datasets. First, when training the BRITS network, information from both the past and future values of a variable is used to impute missing values. This bidirectional flow provides a richer information set that is especially useful in sparse datasets with relatively large numbers of missing values and helps the model achieve the same quality of imputation at the beginning as well as the end of a variable's time series. Second, the BRITS network uses information from the entire cross-section of available observations for a variable, which aligns with the dependency of financial variables on common factors such as business-cycle fluctuations. Third, BRITS attaches less weight to observations in the distant past when imputing missing values by using a *temporal decay* factor. This seems reasonable, as observations from the recent past are likely to have more of an effect on a variable's current value than those from the more distant past. Fourth, BRITS is a type of recurrent neural network that can effectively model data characterized by nonlinear dynamics without making strict assumptions about the data. Such a feature is in advance of previous methods used for recovering missing hedge fund returns in the literature, which require strong assumptions and low-rankness of the data (e.g., singular value decomposition) (Giglio, Liao, and Xiu 2021). Overall, the properties of BRITS make it well suited for imputing missing values for financial data, including hedge fund data.

We consider data for 3,800 hedge funds from January 1994 to December 2021. Our dataset consists of a time series for hedge fund returns, along with 23 hedge 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 measures of fund manager skill. After making standard adjustments to the hedge fund return data and accounting for the availability of hedge fund characteristic data, the in-sample period is January 1998 to December 2012, while January 2013 to December 2021 serves as the out-of-sample period for evaluating hedge fund return forecasts and use them to select out-performing funds based on the fund characteristics and their interactions with a set of economic variables.

We first conduct an experiment over the in-sample period to investigate the accuracy of BRITS for imputing missing values for hedge fund returns and characteristics relative to a set of benchmark methods. For benchmarks, we consider the cross-sectional mean (e.g., Haugen and Baker 1996; Green, Hand, and Zhang 2017; Light, Maslov, and Rytchkov 2017; Chen and McCoy 2024), the time-series mean, and the singular value thresholding algorithm for matrix completion (e.g., Cai, Candès, and Shen 2010; Giglio, Liao, and Xiu 2021). In the experiment, we randomly drop 10% and 20% of hedge fund returns and their corresponding predictors and fill in the actual missing values and the artificial missing values using an imputation method. For the artificial missing values, we use the filled-in and actual values to compute the root mean squared error (RMSE), thereby providing a measure of the accuracy of the imputation method. The BRITS method is compared with benchmark methods, including the time-series mean, cross-sectional mean, and Singular Value Thresholding (SVT), a matrix completion technique. We find that the BRITS method produces a lower RMSE than the benchmarks for the artificial missing values, so BRITS imputes missing values for hedge fund variables with greater fidelity.

Next, we assess the performance of BRITS in an out-of-sample forecasting context, where we forecast 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). We generate a sequence of monthly hedge fund return forecasts for all available funds in a given month using a rolling window of 15 years (180 months). Specifically, to generate return forecasts for month $t+1$, we first consider available data for hedge fund returns and characteristics for month $t-179$ to month t. We apply the BRITS method to fill in missing values for the return and characteristic data. We then use the complete set of available and filled-in observations for the returns and characteristics (as well as the economic variables, which do not have missing values) for month $t-179$ to month t to train a prediction model that generates a set of return forecasts for month $t+1$. We do this sequentially to generate hedge fund return forecasts for January 2013 to December 2021.

We employ three neural network architectures for the prediction models, including deep neural networks.[1] Out-of-sample hedge fund return forecasts based on filling in missing

[1] We also use linear penalized regressions, such as the seminal *least absolute shrinkage and selection operator* Tibshirani (LASSO, 1996) as well as two of its variants: adaptive LASSO Zou (2006), and sparse

values (for both hedge fund returns and characteristics) with BRITS are substantially more accurate in terms of RMSE than return forecasts that rely on filling in missing values via the cross-sectional mean (i.e., XMean) and the forecasts generated without using an imputation method. In the context of forecasting hedge fund returns with a large set of predictors and machine learning methods, BRITS provides an effective strategy for dealing with the plethora of missing values in hedge fund datasets.

We first assess the prediction of hedge fund realized returns using a deep neural network, comparing BRITS-imputed data with other imputation methods, as well as a no-imputation benchmark. The results show that BRITS significantly improves forecasting accuracy, with the neural network ensemble model (NN ensemble) yielding an out-of-sample R^2 of 2.16%, significantly outperforming other imputation methods and the no-imputation approach, which has an out-of-sample R^2 of only 0.16%. In line with the no-imputation method, which performs poorly, the ensemble using XMean also yields negative or insignificant R^2 values.

Next, we evaluate the performance of long-only hedge fund portfolios formed based on these forecasts. The top-decile portfolios of funds, constructed using BRITS-imputed data, deliver the highest realized returns, with an annualized mean return of 23.20%, significantly outperforming other methods. BRITS-imputed portfolios also show the highest alphas, with the Fung-Hsieh alpha at 10.47% and the Chen et al. (2024) alpha at 13.41% both statistically significant. These portfolios also report the highest risk-adjusted performance, including a Sharpe ratio of 1.30 and an upside potential ratio of 0.82. In contrast, the no-imputation portfolios show a much lower annualized mean return of 11.58%, with a Fung-Hsieh alpha of 5.12% and a Sharpe ratio of 0.78. These findings highlight BRITS' strong informational advantage in both forecasting and portfolio management, particularly for long-only strategies where the top-decile portfolio shows superior performance.

group LASSO Simon et al. (2013). The relevant findings are at least qualitatively consistent with those of neural networks, and they are presented in the Internet Appendix.

We further analyze the performance of hedge fund portfolios across prediction-weighted deciles to assess the informational advantage of the imputation methods. Tables 5, and 6 present the annualized mean returns and alphas for each decile portfolio and the top-minus-bottom spread. BRITS demonstrates the highest performance, with a significant annualized return spread of 25.43%, followed by SVT at 21.99%. In contrast, XMean and non-imputed portfolios show lower and statistically insignificant spreads. The alpha results support these findings, with BRITS generating the highest statistically significant alphas—16.97% Fung and Hsieh (2004) and 18.75% Chen et al. (2024).

We also examine the persistence of hedge fund portfolio performance over three years and find that the mean returns do not follow a monotonic decline but exhibit cyclical behavior, especially for the BRITS and SVT imputation methods. Initially, returns drop for about eight months, but BRITS-imputed portfolios recover, reaching a significant 11.83% return at 36 months. This is particularly relevant as hedge funds typically have a three-month lock-up period for new investors. SVT-imputed portfolios also show a similar trend, with a 7.75% return at 36 months. Non-imputed and XMean portfolios display a counter-cyclical pattern, while BRITS portfolios generate the highest returns over the three-year period.

Finally, we examine the importance of fund predictors and the interactions of those predictors with macroeconomic variables to predict fund returns via Shapley Additive exPlanations (SHAP). We focus on the predictions generated using BRITS recovered data, as the superior imputation method. We find that the deep neural network predictions rely heavily on the interactions of fund-specific predictors with macroeconomic variables, specifically the interactions of second-order auto-correlations, total volatility, and cumulative fund returns with the Economic policy uncertainty (EPU) and the VIX indices.

In sum, we make two primary contributions to the literature. First, we show that BRITS is an efficacious deep-learning tool for dealing with missing values in hedge fund datasets. Because missing values is a pervasive problem in hedge fund datasets, this makes BRITS a valuable resource for empirical research on hedge funds. The effectiveness of BRITS in filling in missing values lies in its ability to glean information along both the time and cross-sectional dimensions and to accommodate nonlinear dynamics. BRITS also does not require strong data assumptions, so it provides a flexible approach for filling in missing values in financial datasets. 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 hedge fund characteristics and their interactions with a set of economic variables, as well as a broad array of machine learning models.[2] When we impute missing values for hedge fund datasets via BRITS, we find that the combination of a large number of predictors and machine learning methods generates significant improvements in out-of-sample hedge fund return predictability.

The rest of the paper is organized as follows. Section 2 describes the data. Section 2 describes the data. Section 3 outlines the BRITS network for imputing missing values. Section 4 offer a simulation study that discusses the imputation accuracy. Section 5 discusses the prediction models. Section 6 reports the empirical results. Section 7 concludes.

2. Data

We use hedge fund data from the Lipper Trading Advisor Selection System (TASS) database for January 1994 to December 2021. 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 the 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

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

each fund to have at least 30 return observations.[3] Finally, we discard funds with assets under management below $5 million, so we focus on large funds that are more relevant to investors and that are less likely to manipulate reporting to TASS. After applying the filters, we have 3,800 hedge funds for our analysis.

Table 1. Hedge fund return predictors

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

(1) Variable	(2) Abbreviation
Panel A: Previous 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: Previous 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`

With respect to hedge fund return predictors, we consider a collection of 23 characteristics that are plausible fund return predictors and/or 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 hedge fund characteristics into four categories, as

[3] 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.

indicated in Panels A through D of Table 1. The first category consists of lagged cumulative returns ranging from one to 36 months, which are designed to capture short-, medium-, and long-term momentum effects. The second category is comprised of the first- through third-order autocorrelations in hedge fund returns over the last twelve months. These provide additional measures for capturing persistence in hedge fund returns. The next category is comprised of various return moments, including measures of volatility, coskewness, skewness, and kurtosis. Characteristics that provide proxies for managerial skill make up the fourth category. This category includes alphas based on different multifactor models for hedge fund returns, the R^2 statistic in the context of the well-known Fung and Hsieh (2004) seven-factor model, assets under management, and the maximum return of the last twelve months.

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
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
R^2 over last 12 months for Fung and Hsieh (2004) 7-factor model	Rsq
Assets under management	AUM
Maximum return over the last 12 months	MaxRet
Panel E: Economic variables	
Equity market uncertainty index	EMU
Economic policy uncertainty index	EPU
TED spread	TED
VIX	VIX

In addition, we consider four economic variables (see Panel E of Table 1, which we interact with the fund characteristics. This allows predictive relations between the characteristics and future hedge fund returns to vary with economic conditions. The four economic variables are the equity market uncertainty index (Economic Policy Uncertainty website), the economic policy uncertainty index Baker, Bloom, and Davis (2016), the TED spread, and the CBOE volatility index (VIX). After allowing the hedge fund characteristics to interact with the economic variables, we have a total of $23 + 4 \times 23 = 115$ predictors.

3. BRITS Methodology

Our work expands on Cao et al. (2018) by utilizing the BRITS architecture, a specialized bidirectional LSTM for multivariate time series imputation, applied to hedge fund datasets. BRITS processes data in both forward (positive time) and backward (negative time) directions, allowing for more accurate imputation by jointly training these layers. This bidirectional processing helps capture patterns from past and future data, limiting the *bias exploding* issue while considering the delay of the error calculation of estimated missing entries based on observed values. At the same time, BRITS considers the correlation of the observed entries in the cross-section and also provides cross-sectional imputations, which is particularly valuable for hedge funds and financial datasets in general. The time-series and cross-sectional estimations are combined to generate the final imputation output. Unlike other methods limited to past data or linear assumptions, BRITS can fully recover missing values with high fidelity according to the recurrent dynamics. The following sections explain the data setup, forward and backward layers, hyperparameter optimization, and performance evaluation through simulations, forecasting, and a trading application.

3.1. Initial setup

We denote a dataset of multivariate time series of hedge fund returns and predictors as

$$\{x_{t_1}, x_{t_2}, \ldots, x_{t_N}\} \equiv \mathbf{X}.$$

These time series are observed at regularly-spaced intervals $\{t_1, t_2, \ldots, t_N\}$, and for each observation it holds that $x_t \in \mathbb{R}^D$ (i.e., the t-th observation x_t consists of D elements $\{x_t^1, x_t^2, \ldots, x_t^D\}$). In our hedge fund dataset, it is often the case that some x_t's have missing values for some or all of the D elements. Table $\mathbf{X}$ below provides a representation of our data setting, illustrating the data's irregular missing value patterns. We illustrate an arbitrary form of missingness that is present in both our hedge fund returns and predictors datasets:

$\mathbf{X} =$

Date	**Hedge Fund$_1$**	**Hedge Fund$_2$**	**Hedge Fund$_3$**	**Hedge Fund$_4$**	**Hedge Fund$_5$**
Month 1	observed value	observed value	observed value	observed value	observed value
Month 2	Missing value X?	Missing value X?	Missing value X?	observed value	observed value
Month 3	Missing value X?	Missing value X?	Missing value X?	Missing value X?	observed value
Month 4	Missing value X?	observed value	Missing value X?	Missing value X?	observed value
Month 5	Missing value X?	observed value	observed value	Missing value X?	observed value
Month 6	observed value	Missing value X?	observed value	Missing value X?	Missing value X?
Month 7	Missing value X?	observed value	Missing value X?	observed value	Missing value X?

For each month t, we denote the observed values and the missing values for the multivariate time series matrix $\mathbf{X}$. Missing entries can appear randomly.

The multivariate time series $\mathbf{X}$ is incomplete, and a masking matrix $m_t(M)$ is used to represent the missing components. In some cases, consecutive x_t's ($\in \mathbb{R}^D$) can be partially or fully missing for consecutive timesteps. Given that past time-series observations have a decaying influence on future observations, we aim to capture this feature by creating a specific matrix that tracks how long a value has been missing. Specifically, a tracking matrix $\delta_t(\Delta)$ is used to capture the gap from the last non-missing observation to the current timestep t.

Equations (1) and (2) below provide the calculation of the masking and tracking matrices for every timestep t:

$$m_t^d = \begin{cases} 0, & \text{if } x_t^d \text{ is missing,} \\ 1, & \text{if } x_t^d \text{ is not missing.} \end{cases} \tag{1}$$

$$\delta_t^d = \begin{cases} 1 + \delta_{t-1}^d, & \text{if } t > 1 \text{ and } m_{t-1}^d = 0, \\ 1, & \text{if } t > 1 \text{ and } m_{t-1}^d = 1, \\ 0, & \text{if } t = 1. \end{cases} \tag{2}$$

We show a practical application given the matrix $\mathbf{X}$. In the first column of the *masking* matrix ($\mathbf{M}$), we observe that only the first and the sixth months have a value of 1, while the rest of the months have a value of zero. The sequence of ones and zeros follows the presence of missing values in $\mathbf{X}$, and, evidently, in the first column, we only have observed values for Month 1 and Month 6.

$\mathbf{M} =$

Date	Hedge Fund$_1$	Hedge Fund$_2$	Hedge Fund$_3$	Hedge Fund$_4$	Hedge Fund$_5$
Month 1	1	1	1	1	1
Month 2	0	0	0	1	1
Month 3	0	0	0	0	1
Month 4	0	1	1	0	1
Month 5	0	1	1	0	1
Month 6	1	0	1	0	0
Month 7	0	1	0	1	0

The frequent presence of missing values for the first hedge fund (i.e., the first column) of matrix $\mathbf{X}$ is also depicted in the *tracking* matrix ($\boldsymbol{\Delta}$). For instance, by inspecting the first column and Month 6 of the tracking matrix, we see the value of five. Given the current timestep t (i.e., Month 6), the last observation took place in Month 1, and therefore δ_t is calculated as $\delta_6 = 6 - 1 = 5$.

$\mathbf{\Delta} =$

Date	Hedge Fund$_1$	Hedge Fund$_2$	Hedge Fund$_3$	Hedge Fund$_4$	Hedge Fund$_5$
Month 1	0	0	0	0	0
Month 2	1	1	1	1	1
Month 3	2	2	2	1	1
Month 4	3	3	3	2	1
Month 5	4	1	1	3	1
Month 6	5	1	1	4	1
Month 7	1	2	1	5	2

We impute the hedge fund returns and each predictor matrices separately by forming 24 corresponding cross-sectional matrices, upon which we apply the BRITS methodology. Also, the masking and tracking matrices are calculated separately for the returns and each predictor. The aforementioned setup is essential for training our deep learning architecture, which we describe in the following section.

3.2. BRITS Algorithm

This section provides an overview of the BRITS (Bidirectional Recurrent Imputation for Time Series) algorithm. This technique employs a recurrent neural network (RNN) architecture, which utilizes both forward and backward imputation and a cross-sectional imputation to accurately estimate the missing values. By utilizing historical and future data points, the dual-directional approach improves the imputation process in various ways. First, it improves the slow convergence of the training due to the error estimation delay until an observed entry appears. Second, it reduces the effect of sequential prediction mistakes made early and transferred to the model as inputs (i.e., *bias exploding* issue). Third, the recovered returns and fund predictor values are treated as a learnable parameter, and not a constant, at each timestep. Hence the recovered values are validated by future observations. The cross-sectional improves the imputation process by considering the correlation of a fund's returns with the rest of the funds in the cross-section. Such a feature captures the overall market effect on top of a fund's time-series dynamics.

Consider the dataset of hedge fund returns and predictors represented as $\mathbf{X} = \{x_{t_1}, x_{t_2}, \ldots, x_{t_N}\}$, where at each time step t, certain values may be missing. The mask matrix m_t is used to track missing values, where $m_t = 0$ indicates a missing value and $m_t = 1$ denotes an observed value.

The BRITS algorithm consists of three main components:

1. **Forward Imputation**: In this step, missing values are imputed using the information from the previous time step. At each timestep t, the hidden state h_{t-1} from the previous timestep, combined with the current observation x_t, is used to predict the missing values:

$$(\hat{x}_t)^{forw} = W_x h_{t-1} + b_x$$

   where W_x is the weight matrix and b_x is the bias term. The hidden state h_{t-1} (i.e., recurrent component) is updated using a gated mechanism similar to Long Short Term Memory (LSTM) cells, which allows the model to capture temporal dependencies and impute missing data effectively.[4]

2. **Backward Imputation**: This step involves processing the data in reverse order, starting from the last timestep and moving backward. The backward imputation uses information from future timesteps to improve the accuracy of the imputed values:

$$(\hat{x}_t)^{back} = W_x h_{t+1} + b_x$$

   where h_{t+1} is the hidden state from the subsequent timestep, and W_x and b_x are shared parameters with the forward imputation.

[4] A LSTM network is used as the applied recurrent network avoiding the gradient vanishing issue observed in a plain RNN.

3. **Cross-sectional Imputation**: The final step involves processing a fund's data based on the other funds' entries cross-sectionally:

$$(\hat{x}_t)^{cross} = W_c x_t^c + b_c$$

where x_t^c are either the observed cross-sectional return values at a timestep t or their time-series recovered values when those are missing.

The final imputed value $\hat{x}_t^{Combo}$ is obtained by combining the time-series (i.e., forward and backward looking) and the cross-sectional imputations results. This dual imputation strategy enhances the model's ability to accurately predict missing values by considering past and future was well as cross-sectional data:

$$\hat{x}_t^{Combo} = \beta_t \cdot (\hat{x}_t)^{cross} + (1 - \beta_t) \cdot \hat{x}_t$$

where $\hat{x}_t = ((\hat{x}_t)^{forw} + (\hat{x}_t)^{back})/2$, and β_t is the optimal weighting between the time-series and the cross-sectionally based estimation ($0 \leq \beta_t \leq 1$).

Including cross-sectional information requires the adjustment of the hidden layer h_t to the $\hat{x}_t^{Combo}$ since the hidden layer is involved in the time-series imputation. Additionally, to account for the duration of missing values, a temporal decay factor γ_t is introduced. This factor controls the influence of past observations on the imputation at timestep t, depending on how long the value has been missing, $\gamma_t = e^{-\max(0, W_\gamma \delta_t + b_\gamma)}$, where δ_t represents the gap since the last non-missing observation, and W_γ and b_γ are parameters learned during training. The temporal decay factor, γ_t is involved in both the backward and forward estimations and the estimation of β_t, and as a result, the optimal weighting can be adjusted when a value has been missing for several timesteps and, therefore, its influence diminishes. For instance, suppose a fund's return has been missing for several months, and we need to impute its value at time t in which cross-sectional values are observed. Then, at this point, the BRITS adjusts β_t by assigning a larger weight towards the cross-sectional estimation (i.e., $(\hat{x})_t^{cross}$)

and a smaller weight to the time-series estimation (i.e., $\hat{x}_t$) given that the fund's return has been missing for a long period.[5]

Finally, a reconstruction loss function is used, which focuses on minimizing the error for the imputed values while excluding the observed values:

$$l_t^{final} = \langle m_t, L_e(x_t, \hat{x}_t)\rangle + \langle m_t, L_e(x_t, (\hat{x}_t)^{cross})\rangle + \langle m_t, L_e(x_t, (\hat{x}_t)^{Combo})\rangle$$

where Ł$_e$ is the estimation error.

Essentially, the l_t^{final} accumulates the model's error from three different error sources (i.e., forward, backward and cross-sectional imputation) so that the weight matrices can be updated effectively and reach their optimal values.[6] The BRITS model is then updated by minimizing the accumulated error. Following Cao et al. (2018), we use the mean absolute error (MAE) as our estimation error. The BRITS tuning and hyperparameters are presented in the Internet Appendix A.

4. Imputation Accuracy

We examine the performance of the BRITS methodology by estimating the imputation error via a simulation study in which we randomly drop a percentage of observed fund values and impute them along with the original missing entries. We track the location of the artificially dropped values via indicative masking, which is essential for measuring imputation performance. We follow the recent literature on financial data imputation (Bryzgalova et al. 2025) and adapt the root-mean-square error (RMSE) formula to calculate the imputation error for the simulation study. A lower error reveals higher imputation performance and, therefore, higher imputation fidelity. We develop two simulation experiments for our in-sample dataset (i.e., January 1998 to December 2012). In the first one, we randomly drop

[5] For further details of the algorithm please see Cao et al. (2018).

[6] Separate estimations are performed for backward and forward components (i.e., $l_t^{forw} = \langle m_t, L_e(x_t, (\hat{x}_t)^{forw})\rangle$ and $l_t^{forw} = \langle m_t, L_e(x_t, (\hat{x}_t)^{forw})\rangle)$ before generating the time-series imputations, $\hat{x}_t$. To ensure that the forward and backward imputations are consistent, a consistency loss function is employed, penalizing discrepancies between the forward and backward imputations (i.e., $l_t^{cons} = \left\|(\hat{x}_t)^{forw} - (\hat{x}_t)^{back}\right\|$).

10% of hedge fund returns and the 23 fund-specific predictors' datasets, and we fill in their missing entries along with those that are initially missing. We repeat the same process in the second experiment, but we randomly remove 20% of the original entries this time.

We also compare the performance of BRITS in imputing hedge fund missing entries against famous counterparts used by the financial literature. For that purpose, we employ as benchmarks the time series mean and the cross-sectional mean (Haugen and Baker 1996; Green, Hand, and Zhang 2017; Light, Maslov, and Rytchkov 2017; Beckmeyer and Wiedemann 2023; Bryzgalova et al. 2025) and the singular value thresholding (SVT) for nuclear norm minimization (Giglio, Liao, and Xiu 2021), which belongs to a class of algorithms known as the matrix completion methods, which are particularly used to fill in missing values in the data (Cai, Candès, and Shen 2010).[7]

Table 2 presents the 10% and 20% simulation study results, respectively, for the hedge fund returns and the universe of our predictors. Panel A reports results for a 10% data reduction, while Panel B shows results for a 20% reduction. We report the mean RMSE criterion. For the predictor's case, we provide aggregate metrics, such as average RMSE and the standard deviation of the RMSE across all predictors. A lower standard deviation indicates consistent imputation performance across all predictors. We provide the exact imputation performance for each predictor separately in Appendix B.

Tables 2 shows that BRITS generates superior imputation performance than the benchmark methods for hedge fund returns and fund-specific predictors. Concerning the returns imputation, BRITS yields the smallest RMSE, 3.448 for the 10% simulation and 3.586 for the 20% simulation, respectively, than all benchmark imputation methods. BRITS also yields the lowest standard deviation of imputation error compared to the benchmark models, as shown in the same tables. The second performing model is the cross-sectional mean. Such a finding aligns with the recent evidence of Chen and McCoy (2024), who find that

[7] Giglio, Liao, and Xiu (2021) study differs from our study as they use a singular value decomposition method to impute missing values of the residuals of hedge fund returns obtained by factor models for identifying funds with significant alphas via false discovery rate control.

Table 2. Imputation Fidelity

The table presents the results of an imputation simulation in which we artificially drop a portion of the observed values for hedge fund returns and each of the 23 fund-specific predictors. Panel A reports results for a 10% data reduction, while Panel B shows results for a 20% reduction. We evaluate imputation accuracy using the root mean square error (RMSE), reporting the RMSE for hedge fund returns, as well as the average RMSE and standard deviation (Std) of the error across all predictors.

Panel A: 10% Simulation			
Imp. Method	Fund Returns	Fund Predictors	
		Average error	Std of error
TSMEAN	4.503	6.894	9.365
XMEAN	4.038	6.885	8.955
SVT	4.065	6.514	9.155
BRITS	3.448	4.114	5.308

Panel B: 20% Simulation			
Imp. Method	Fund Returns	Fund Predictors	
		Average error	Std of error
TSMEAN	4.560	9.422	19.181
XMEAN	4.121	9.492	19.031
SVT	4.190	9.246	19.221
BRITS	3.586	6.882	17.617

cross-sectional mean is still one of the most robust methods for recovering missing entries in financial datasets, even standard ML methods. Our findings show that an ML method that considers both time-series and cross-sectional characteristics for recovering missing entries, such as BRITS, has higher imputation fidelity than the cross-sectional mean and more powerful matrix completion methods, such as the SVT. Concerning the predictors' imputation, the results align with those of hedge fund returns, highlighting that BRITS yields the smallest average RMSE value at the 4.114 level for the 10% simulation case and the 6.882 level for the 20% simulation case. Once again, BRITS generates the lowest standard deviation of imputation error. This time, the SVT method reports the second-best imputation performance. The above results reveal that BRITS is a robust method for recovering missing data for returns and predictor variables.

5. Prediction Model

We consider a feedforward neural network as our prediction model for the hedge fund realized returns. Most recent studies highlight the superior performance of deep neural networks in predicting both equities' and fund' returns compared to other linear and nonlinear methods (Gu, Kelly, and Xiu 2020; Wu et al. 2021; Kaniel et al. 2023). Contrary to other machine learning approaches, neural networks have the advantage of providing valid confidence intervals, similar to regression analysis (Kaniel et al. 2023).

A generic model for predicting the one-month-ahead excess return for hedge funds can be described as follows:

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

where $r_{i,t}$ represents the excess return for hedge fund i during the month t, $\boldsymbol{x}_{i,t} = [x_{1,i,t}, \ldots, x_{k,i,t}]'$ is a vector of k predictor variables, and $f(\boldsymbol{x}_{i,t}; \boldsymbol{\eta})$ denotes the conditional expectation function (or prediction function), which depends on a parameter vector $\boldsymbol{\eta}$. The error term, $\varepsilon_{i,t+1}$, is assumed to be mean-zero and serially uncorrelated. This framework is based on a pooled model, where the prediction function and parameter values are identical across all hedge funds. Pooling reduces the number of parameters to estimate, as it eliminates the need for fund-specific parameter estimates. This is beneficial for improving out-of-sample performance due to the bias-variance trade-off, especially for hedge funds with limited historical data. Similar pooled approaches have been used in the context of asset return prediction, as seen in studies like Freyberger, Neuhierl, and Weber (2020), Gu, Kelly, and Xiu (2020), and Filippou et al. (2023).

The corresponding forecast for excess return is given by:

$$\hat{r}_{i,t+1} = \hat{f}(\boldsymbol{x}_{i,t}; \hat{\boldsymbol{\eta}}), \tag{2}$$

where $\hat{f}(\cdot;\hat{\boldsymbol{\eta}})$ is the fitted prediction function based on data up to month t, thus avoiding any look-ahead bias.

5.1. Deep Neural Network

A neural network consists of multiple layers of interconnected neurons. The first layer corresponds to the input data (predictors), followed by one or more hidden layers, and the final layer outputs the prediction.

Each hidden layer l consists of P_l neurons, where the mth neuron in layer l computes a nonlinear transformation of the inputs from the previous layer:

$$h_m^{(l)} = g\left(\omega_{m,0}^{(l)} + \sum_{j=1}^{P_{l-1}} \omega_{m,j}^{(l)} h_j^{(l-1)}\right) \quad \text{for } m = 1, \ldots, P_l,\ l = 1, \ldots, L, \tag{3}$$

where $h_j^{(0)} = x_{j,i,t}$ represents the input features, $\omega_{m,j}^{(l)}$ are the weights, and $g(\cdot)$ is a nonlinear activation function, typically the ReLU or leaky ReLU (Maas et al. 2017). The final output is a weighted sum of the last hidden layer:

$$\hat{r}_{i,t+1}^{\text{Net}} = \omega_0^{(L+1)} + \sum_{j=1}^{P_L} \omega_j^{(L+1)} h_j^{(L)}. \tag{4}$$

The network is trained by adjusting the weights using a stochastic gradient descent (SGD) algorithm, with the goal of minimizing the training loss. We use the Adam optimizer Kingma and Ba (2017) for training, and hyperparameters such as the number of neurons per layer, dropout rates, and learning rates are tuned using walk-forward cross-validation. As a robustness check, we implement a family of penalized regressions (i.e. LASSO, sparse group LASSO, and adaptive LASSO) as alternative linear approaches. The corresponding findings are similar to those of the neural network prediction, revealing the importance of BRITS for prediction and portfolio construction purposes. We report an in-depth comparison to the Internet Appendix E.

6. Empirical Analysis

After establishing BRITS' accuracy for recovering fund returns and characteristics, we examine whether the main futures of BRITS (i.e., time-series and cross-sectional components) in recovering missing data entries provide informational advantages in hedge fund return predictability and portfolio investment performance out-of-sample. First, we assess the prediction of hedge fund realized returns with the deep neural network, using information recovered from BRITS and the benchmark imputation methods, as well as the naive approach of not applying any imputation method (i.e., adopting the original dataset with missing values). Second, we assess the performance of long-only hedge fund realized portfolios formed based on the neural network forecasts and for each imputation method.

6.1. Forecasting Performance

As highlighted in Gu, Kelly, and Xiu (2020), using historical averages to predict future excess stock returns often performs substantially worse than a simple zero-return forecast. The high level of noise in historical mean returns creates an artificially low standard for what constitutes effective forecasting. To circumvent this issue, we use a zero forecast as the benchmark for our out-of-sample R^2 statistic.

We can compare the out-of-sample MSE for a forecast value of zero to that of a competing forecast via the out-of-sample R^2 statistic Fama and French (1989) and Campbell and Thompson (2008):

$$R^2_{\text{All,OS}} = 1 - \frac{\sum_{i=1}^{n} \sum_{s=1}^{T-t_{\text{in}}} \left(r_{i,t_{\text{in}}+s} - \hat{r}^{\text{Compete}}_{i,t_{\text{in}}+s} \right)^2}{\sum_{i=1}^{n} \sum_{s=1}^{T-t_{\text{in}}} r^2_{i,t_{\text{in}}+s}}, \tag{5}$$

where $\hat{r}^{\text{Compete}}_{i,t}$ generically denotes a competing forecast, and t_{in} (T) is the end of the initial in-sample period (total sample). Equation (5) is the proportional reduction in out-of-sample MSE for the competing forecast vis-à-vis the benchmark forecast across all of the hedge funds. We use the Diebold and Mariano (1995) and West (1996) (DMW) statistic to test whether the competing forecast delivers a statistically significant reduction in MSE relative to the benchmark. We compute the DMW statistic via the t-statistic corresponding to the intercept a in the following pooled regression:

$$\underbrace{r_{i,t}^2 - \left(r_{i,t} - \hat{r}_{i,t}^{\text{Compete}}\right)^2}_{d_{i,t}} = a + \varepsilon_{i,t} \quad \text{for } i = 1, \ldots, n;\ t = t_{\text{in}} + 1, \ldots, T. \tag{6}$$

where $d_{i,t}$ is the day-t loss differential (i.e., the difference between the squared errors for the benchmark and competing forecasts). We test H_0: $a \leq 0$ ($R^2_{\text{All,OS}} \leq 0$) against H_A: $a > 0$ ($R^2_{\text{All,OS}} > 0$). When computing the DMW statistic using Equation (6), we account for cross-sectional dependency by clustering the standard error by month and hedge funds.

6.1.1. Out-of-sample Results

Table 3 presents the out-of-sample $R^2_{i,\text{OS}}$ statistics for monthly return forecasts. Out-of-sample predictions begin on January 1, 2013, and continue until the end of the sample period. The results indicate that the no-imputation method (NoImp), which excludes observations with missing values, performs poorly. In contrast, as shown in the third column, using BRITS as an imputation method for both returns and predictors yields highly positive and always statistically significant out-of-sample R^2 statistics across different neural network specifications. Notably, the ensemble model, which integrates all specifications, achieves the second highest out-of-sample R^2 statistics of 2.16, which is statistically significant. The highest R^2 statistic (i.e., 2.28) is that of the ensemble using SVT as an imputation method. However, the statistic is not always statistically significant across different neural network models. Interestingly, the predictions with the cross-sectional mean (XMean) as an imputation generate the second lowest R^2 statistics, which are also not statistically significant. Such a finding contrasts with the recent findings Chen and McCoy (2024), who report the cross-sectional mean's comparable imputation performance with deep learning specifications on equity returns. The last column examines a variant of the third column (NoImpRetBRITS) where BRITS is applied only to impute missing predictor values, during the in-sample period. This approach results in equivalent out-of-sample neural network ensemble performance with BRITs and SVT. However, the R^2s are not always statistically significant across the neural network variations, suggesting that imputing both predictors and returns is crucial for maximizing predictive accuracy.

Overall, the results in Table 3 indicate that the unique properties of the BRITS imputation method facilitate the effective extraction of information from a broad range of predictors, significantly enhancing the accuracy of monthly hedge fund return predictions.

Table 3. Out-of-sample R^2 statistics

The table reports out-of-sample R^2 statistics in percent for monthly excess return forecasts that are based on different NN specifications. Specifically, we show results for a NN with three layers (NN3), NN with four layers (NN4), NN with five layers (NN5) and the ensemble across all the specifications. The number of out-of-sample observations is given in the second column. The "NoImp" column reports out-of-sample R^2 statistics for an estimation that is based on no imputation. The "BRITS" column reports results when both returns and predictors are imputed using BRITS. The "SVT" column shows results for an estimation that is based on imputed series using SVT. The "XMean" column displays out-of-sample R^2 statistics for imputed variables based on the cross-sectional mean and the "NoImpRetBRITS" column reports results using imputed predictors based on BRITS and non-imputed returns. The out-of-sample R^2 statistic measures the proportional reduction in out-of-sample mean squared error (MSE) for the competing forecast in the column heading vis-à-vis the zero benchmark forecast; based on 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% level, respectively.

	N	NoImp	BRITS	SVT	XMean	NoImpRetBRITS
NN3	137505	0.16	1.58***	1.68***	-0.09	1.26
NN4	137505	0.01	1.59***	1.43	0.05	2.04***
NN5	137505	0.19	2.45***	2.76***	0.58	2.68***
NN ensemble	137505	0.16	2.16***	2.28***	0.29	2.25***

6.2. Out-of-sample Portfolio Performance

We assess the economic significance of the imputed hedge fund datasets in predicting "skilled" hedge funds. For each data imputation method, at the end of each month, we sort hedge funds in the cross-section into deciles based on the neural network forecasts generated previously. Then, we form prediction-weighted decile portfolios on funds' realized returns, following Kaniel et al. (2023). Prediction-weighted decile portfolios have the advantage of exploiting the heterogeneity in the prediction and assigning a higher relative weight to predictions that deviate more from the center of the decile.[8] The relevant results of the equally weighted decile portfolios, leading to similar conclusions, are presented in Appendix D.

6.2.1. Performance of Top-Decile Portfolio

Hedge fund investors are mainly interested only in the top-decile portfolio, which includes the top-performing hedge funds, due to investors' inability to short hedge funds in contrast with other assets (e.g., stocks). Hence, we evaluate the realized performance of investing all capital available by the end of each month in the top decile portfolio of funds, equivalent to holding long positions on the "skilled" funds.

Table 4 presents the out-of-sample realized performance of the top-decile portfolio of funds. We focus on presenting the performance of the portfolio constructed based on the forecasts of the averaging ensemble of neural networks for easy comparison. In particular, we present the annualized mean return, the annualized alphas of the Fung and Hsieh (2004) (FH) seven-factor and the Chen et al. (2024) (CLTZ) nine-factor models for hedge fund returns along with their corresponding Newey-West t-statistics with three lags. We also report a battery of risk and risk-adjusted performance measures, such as the maximum drawdown (MaxxDD), the annualized Sharpe ratio (SR), the upside potential ratio (UPR) and the corresponding information ratios of the FH and CLTZ alphas, respectively.

The top-decile portfolios using BRITS for recovering hedge fund returns and predictors or predictor data only yield the highest performance among the imputation and non-imputation methods. The corresponding top-decile portfolios generate the highest and statistically significant mean returns and alphas. The same portfolios also report the highest ratios. The

[8] The prediction-based weighting is determined by shifted and scaled weights. We subtract the smallest model fund prediction from each fund in the top-decile portfolio to ensure that the portfolio is long-only. We subtract the largest model fund prediction from each fund in the bottom-decile portfolio to ensure that the portfolio is short-only. We then standardize the normalized predictions to sum up to one. Please refer to the relevant footnote of Kaniel et al. (2023) for the detailed computation steps.

Table 4. Prediction-weighted Top-Decile Portfolio Performance

The table reports the out-of-sample performance of the prediction-weighted top-decile portfolio of hedge funds based on the NN predictions across all imputation approaches. Specifically, we show the performance of the top-decile portfolio of realized fund returns based on the predictions of Neural nets ensemble estimations. Each machine learning model is estimated on the imputed dataset of the "BRITS" and each benchmark imputation method along with the non-imputed dataset. We report the annualized mean return, the Fung and Hsieh (2004) (FH) and the Chen et al. (2024) (CLTZ) factor model annualized alphas, the maximum drawdown (MaxxDD) the Sharpe ratio, the upside-potential ratio (UPR) and the corresponding information ratios based on the Fung and Hsieh (2004) (FH IR) and the Chen et al. (2024) (CLTZ IR) factor model alphas. We use Newey and West (1987) standard errors with three lags to measure the statistical significance of mean returns and alphas. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	Mean (%)	FH alpha (%)	CLTZ alpha (%)	MaxxDD	SR	UPR	FH IR	CLTZ IR
NonImp	11.58**	5.12	10.07*	25.01	0.78	0.68	0.32	0.64
BRITS	23.20***	10.47*	13.41**	21.81	1.30	0.82	0.60	0.73
XMEAN	4.26	0.58	3.49	49.62	0.32	0.53	0.04	0.24
SVT	20.80***	8.32	10.73**	22.31	1.22	0.83	0.50	0.64
NonImpRetBRITS	24.22***	13.80**	14.59***	22.19	1.32	0.89	0.79	0.88

second-best imputation method for fund portfolio construction is the SVT. However, applying no imputation on fund datasets yields almost half lower portfolio performance, with a barely significant Chen et al. (2024) alpha, while the corresponding portfolios built based on the XMean imputation method generate the worst performance with insignificant mean returns and alphas.

6.2.2. Performance of Decile Portfolios

We further assess the informational advantage of the examined imputation methods on hedged fund sorting ability by reporting the portfolio realized performance across all prediction-weighted deciles. We anticipate a more sound monotonic increase and a statistically significant difference in the realized performance between the top (D10) and bottom (D1) decile portfolios for imputation methods with more significant informational advantage. Table 5 and Table 6 report the annualized mean returns and the annualized Fung and Hsieh (2004) and Chen et al. (2024) alphas for each decile portfolio and the top-bottom portfolio across imputation methods, respectively.

Our results again justify the informational superiority of BRITS in providing information for successfully differentiating high from low-performing hedge funds both by recovering

Table 5. Prediction-weighted Decile Portfolio Returns

The table provides the out-of-sample returns of the prediction-weighted decile portfolios, and top-minus-bottom spread portfolio of hedge funds based on NN predictions across all imputation approaches. Specifically, we show the performance of the portfolios of realized fund returns based on NN's predictions. NN is estimated on the imputed dataset of the "BRITS" and each benchmark imputation method along with the non-imputed dataset. We report the annualized mean returns of the decile portfolios. We use Newey and West (1987) standard errors with three lags to measure the statistical significance of mean returns and alphas. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D10–D1
NoImp	8.75**	4.30**	4.35**	4.87***	4.82**	5.42***	6.07***	6.91***	8.05***	11.58**	2.83
BRITS	-2.23	3.34*	4.01***	3.65**	4.66**	5.47***	6.48**	8.59***	9.62***	23.20***	25.43***
XMean	5.17	4.11*	5.60***	5.16**	5.75***	6.11***	6.29***	6.06**	9.27***	4.26	-0.91
SVT	-1.19	3.70**	4.36***	4.52***	5.00***	5.24***	5.57**	8.54***	9.45***	20.80***	21.99***
NoImpRetBRITS	1.24	2.91	4.38**	3.29*	5.23***	5.38***	6.65***	8.46***	9.40***	24.22***	22.98***

returns' and/or predictors' missing entries. The difference between the top and bottom decile portfolios is the largest among the other methods. For instance, this is 25.43% for the former and 22.98% for the latter, both statistically significant at 1%. Consistent with our previous findings, those differences are followed by those reported for SVT (i.e., 21.99%) which are also statistically significant at 1%. For the aforementioned imputation methods, there is a clear monotonic increase in the returns of the decile portfolios. On the contrary, the portfolios constructed with XMean imputation or no imputation the spread portfolio returns are the lowest and not statistically significant. Regarding the FH and CLTZ alphas, the overall picture is similar. We find in Table 6 that only the long-short decile portfolio of BRITS, NoImpRetBRITS, and the SVT imputation methods report statistically significant alphas, with those of BRITS being the highest. The spread portfolios of the neural network predictions based on the BRITS imputation method generated a Fung and Hsieh (2004) and a Chen et al. (2024) alpha of 19.97% and 18.75% respectively.

Figure 1 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 in-

Table 6. Prediction-weighted Decile Portfolio Alphas

The table presents out-of-sample annualized alphas from the Fung and Hsieh (2004) (FH) factor model (Panel A) for prediction-weighted decile portfolios and the top-minus-bottom spread portfolio of hedge funds based on NN predictions across all imputation approaches. Panel B reports alphas from the Chen et al. (2024) (CLTZ) factor model. Specifically, we evaluate the performance of realized fund return portfolios derived from NN's predictions. NN is estimated on the imputed datasets from "BRITS" and each benchmark imputation method, as well as the non-imputed dataset. We report the FH (CLTZ) factor model annualized alphas for each decile and the spread portfolio. Statistical significance is assessed using Newey and West (1987) standard errors with three lags. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Panel A: Fung and Hsieh (2004) Alphas											
	D1	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**	**D10–D1**
NoImp	2.42	-0.69	-0.55	0.09	-0.02	0.87	0.99	1.29	2.45	5.12	2.70
BRITS	-6.50	-0.81	0.47	0.19	0.31	0.61	0.52	1.92	1.78	10.47*	16.97***
XMean	-1.70	-1.74	0.45	-0.08	0.38	0.39	1.10	0.73	4.64***	0.58	2.28
SVT	-7.38**	-0.65	0.90	0.91	0.51	0.67	0.45	2.31*	2.16	8.32	15.70***
NoImpRetBRITS	-5.36**	-1.33	0.10	-0.36	0.99	0.35	1.18	2.38*	2.71**	13.80**	19.16***
Panel B: Chen et al. (2024) Alphas											
	D1	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**	**D10–D1**
NoImp	4.24	-0.61	0.14	0.64	0.64	1.67*	2.06**	3.20**	4.07**	10.08*	5.84
BRITS	-5.14	-0.35	0.97	0.27	1.13	1.98**	1.46	3.32**	3.62**	13.61**	18.75***
XMean	-0.69	-0.88	1.49	0.88	1.53	1.91*	2.30***	2.32**	5.35***	3.49	4.18
SVT	-5.17	-0.23	1.63**	0.97	1.72**	1.72*	1.36	3.70**	4.00**	10.73**	15.90***
NoImpRetBRITS	-2.85	-0.53	0.71	0.01	1.93***	1.62	1.61	3.48***	3.94***	14.59***	17.44***

vestor who has used BRITS imputed information to invest in the best 10% of funds using a neural network as a forecaster would have earned a cumulative abnormal return of 207% and 216% with BRITS and NoImpRetBRITS, respectively. The same investor would have avoided the worst 10% of funds yielding a corresponding -19% and 11% cumulative return, 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 NonImp and XMean 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.

6.2.3. 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

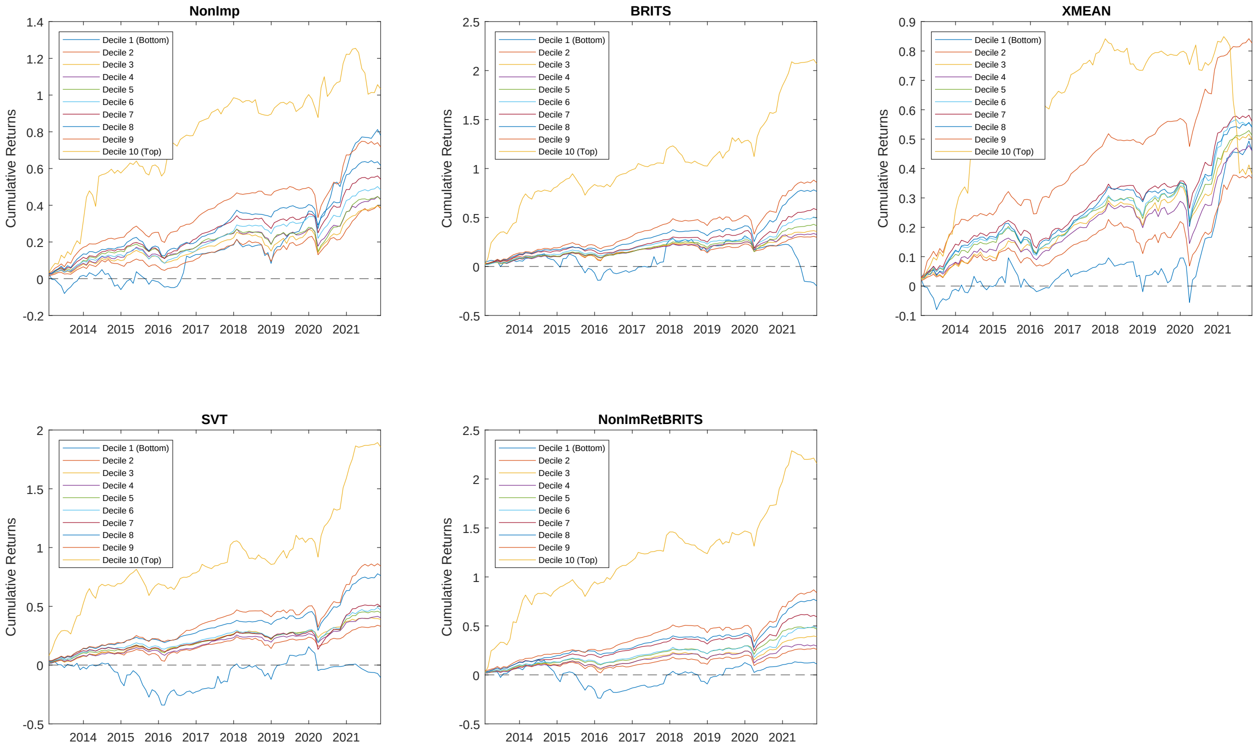

Figure 1. Prediction-weighted Decile Portfolio Cumulative Returns

The figure presents the out-of-sample cumulative realized returns of the NN prediction-weighted decile portfolios for each imputation approach.

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. (2024) (CLTZ) factor model annualized alphas of the portfolios' overlapping returns along with their corresponding t-statistics with Newey and West (1987) standard errors.[9] Figure 2 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

[9] We focus only the Chen et al. (2024) alphas due to lack of space. The relevant figures for the Fung and Hsieh (2004) alphas are similar and available upon request

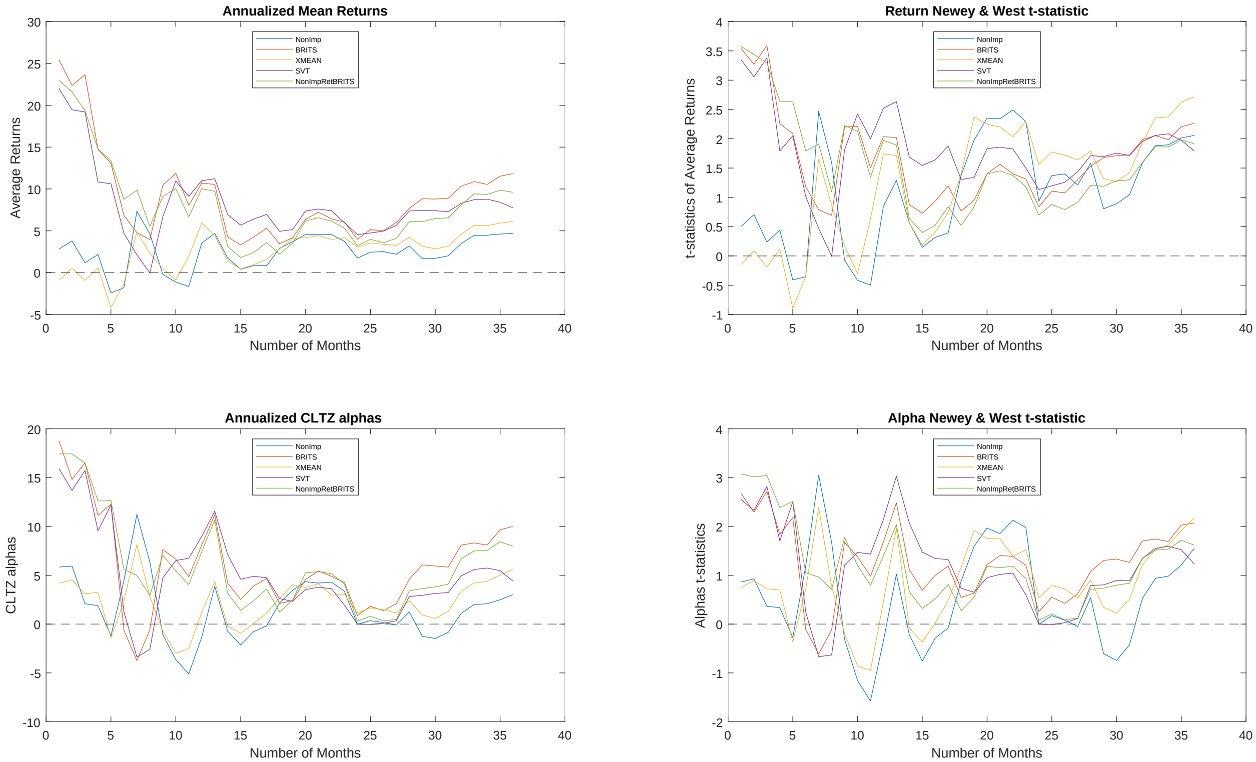

Figure 2. Prediction-weighted Top Decile Portfolio Performance Persistence

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.

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 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. (2024) alpha and its corresponding t-statistic are 10% and 2.07, respectively, after three years. Except for XMean, no other imputation method provides information that yields a statistically significant alpha (i.e., 5.61% and 2.17, respectively).

6.3. 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.[10]. 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 3 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

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

network specification separately and then compute the average SHAP values for each predictor across the three corresponding values of each neural network model.[11] 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 3 reports the relevant importance ranking.

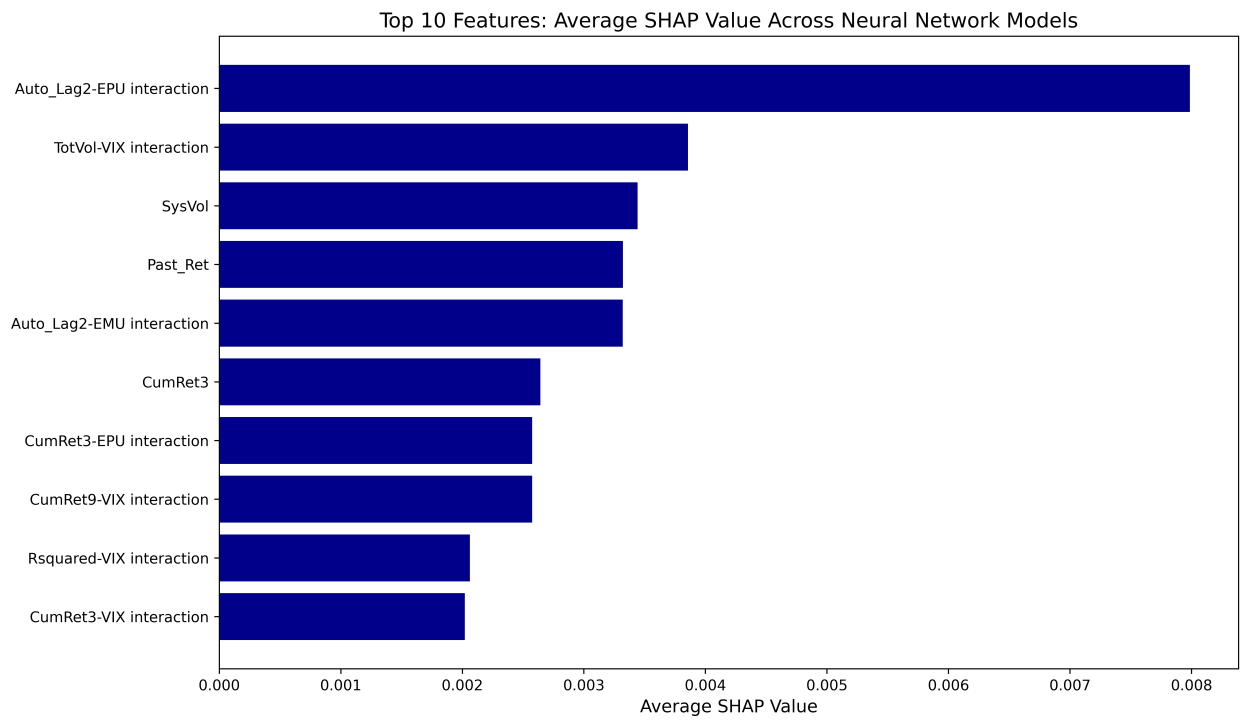

Figure 3. Predictor Importance - Top 10 Variables

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. under the BRITS imputation method. The SHAP values are separately generated for the predictors of each neural network, and then the average SHAP values across each neural network model are calculated for each predictor. The presented SHAP values correspond to the forecasting exercise with BRITS imputed values.
the forecasting models per neu for the neural network ensemble.
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

We observe from Figure 3 that the interactions of fund-specific predictors with macroeconomic variables dominate the standalone fund-specific predictors regarding their importance

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

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, 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 families of previous returns, previous return autocorrelations and return moments, but no one from the family of predictors measuring managerial skill.

7. Conclusion

We address the common issue of missing values in hedge fund datasets, a problem that poses challenges for empirical asset pricing and out-of-sample return forecasting. We show that the Bidirectional Recurrent Imputation Network for Time Series (BRITS) is an effective tool for imputing missing hedge fund returns and characteristics. BRITS outperforms traditional imputation methods, such as cross-sectional mean imputation and Singular Value Thresholding (SVT), by using both time-series and cross-sectional information to provide more accurate imputations, especially in the presence of missing data across different time periods. This dual approach enhances the imputation quality and is particularly useful in hedge fund datasets that typically exhibit sparse data.

Our results demonstrate that BRITS leads to significant improvements in the accuracy of hedge fund return forecasts. Specifically, we find that out-of-sample forecasts based on BRITS-imputed data exhibit substantially lower RMSE and higher out-of-sample R^2 values compared to other methods. Furthermore, portfolios constructed using BRITS-imputed data generate superior returns and risk-adjusted performance, including higher alphas and Sharpe ratios, compared to those based on other imputation techniques. These findings underline BRITS' ability to improve both the predictive accuracy of asset returns and the performance of hedge fund portfolios, particularly when using a large set of predictors and machine learning models.

Overall, this paper makes two key contributions to the literature. First, it establishes BRITS as a powerful tool for imputing missing values in hedge fund datasets, improving both model estimation and out-of-sample forecasting. Second, it provides a comprehensive analysis of hedge fund return predictability, incorporating a wide range of hedge fund characteristics and economic variables, as well as advanced machine learning methods. The findings underscore the importance of addressing missing data in financial research and offer a robust framework for improving financial forecasting and portfolio management. By leveraging BRITS, we contribute to a more accurate and reliable understanding of hedge fund returns and their drivers.

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, 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.

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.

Beckmeyer, H. and T. Wiedemann (2023). Empirical Asset Pricing with Missing Data. SSRN Working Paper No. 4003455.

Bryzgalova, S., S. Lerner, M. Lettau, and M. Pelger (2025). Missing Financial Data. *Review of Financial Studies*, forthcoming.

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.*

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

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 (2024). Anomalies as New Hedge Fund Factors. Working paper.

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 (2023). Economic Fundamentals and Short-Run Exchange Rate Prediction: A Machine Learning Perspective. Working Paper (available at https://ssrn.com/abstract=3455713).

Freyberger, J., B. Höppner, A. Neuhierl, and M. Weber (2025). Missing Data in Asset Pricing Panels. *Review of Financial Studies*, forthcoming.

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.

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

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.

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. ISSN: 0304-405X.

Kingma, D. P. and J. Ba (2017). Adam: A Method for Stochastic Optimization. *Working Paper*. arXiv: 1412.6980 [cs.LG].

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: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds. *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc.

Maas, A. L., P. Qi, Z. Xie, A. Y. Hannun, C. T. Lengerich, D. Jurafsky, and A. Y. Ng (2017). Building DNN acoustic models for large vocabulary speech recognition. *Computer Speech Language* 41: 195–213. ISSN: 0885-2308.

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.

Internet Appendix to

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

by

Ilias Filippou Ioannis Psaradellis David E. Rapach Lazaros Zografopoulos

(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 (Gu, Kelly, and Xiu 2020; Beckmeyer and Wiedemann 2023)). 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 variance of the estimated model. We use the Gaussian Error Linear Unit (GELU) as the

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

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]	3* 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. Out-of-sample Equally-Weighted Portfolio Performance

We present the empirical findings of portfolio performance and analysis for equally-weighted portfolios formed with neural network predictions as those generated using different imputation method's recovered datasets and the non-imputed dataset. Table A5 reports the performance of the corresponding top-decile equally-weighted portfolios. Similar to the relevant findings of the prediction-weighted portfolios the portfolios constructed with the BRITS imputation methods yield the highest performance compared to the benchmark imputation methods. Both mean returns and alphas are all statistically significant at 1% for most of the cases. The BRITS portfolio generates a mean return of 15.26% and 4.48% and 7.48% Fung and Hsieh (2004) and Chen et al. (2024), respectively. The BRITS portfolio in which only predictors are imputed reports a similar performance. The SVT method follows in terms of mean returns (i.e., 14.24% and statistically significant at 1%), while, interestingly, the portfolio using non-imputed data follows with respect to Fung and Hsieh (2004) and Chen et al. (2024) significant alphas (i.e., 3.74% and 6.16%, respectively).

Table A5. Equally-Weighted Top-Decile Portfolio Performance

The table reports the out-of-sample performance of the equally-weighted top-decile portfolio of hedge funds based on the NN predictions across all imputation approaches. Specifically, we show the performance of the top-decile portfolio of realised fund returns based on the predictions of neural network ensemble estimations. Each machine learning model is estimated on the imputed dataset of the "BRITS" and each benchmark imputation method along with the non-imputed dataset. We report the annualized mean return, the Fung and Hsieh (2004) (FH) and the Chen et al. (2024) (CLTZ) factor model annualized alphas, the maximum drawdown (MaxxDD) the Sharpe ratio, the upside-potential ratio (UPR) and the corresponding information ratios based on the Fung and Hsieh (2004) (FH IR) and the Chen et al. (2024) (CLTZ IR) factor model alphas. We use Newey and West (1987) standard errors with three lags to measure the statistical significance of mean returns and alphas. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively

	Mean (%)	FH alpha	CLTZ alpha	MaxxDD	SR	UPR	FH IR	CLTZ IR
NonImp	10.96***	3.74*	6.16***	15.08	1.24	0.72	0.66	1.29
BRITS	15.26***	4.48**	7.48***	20.42	1.26	0.76	0.65	1.19
XMEAN	8.27***	3.09*	5.07***	7.27	1.25	0.69	0.71	1.29
SVT	14.42***	3.53	6.16***	18.26	1.21	0.76	0.52	1.00
NonImpRetBRITS	14.59***	5.28**	7.76***	17.79	1.34	0.76	0.80	1.39

Regarding the portfolio analysis, Tables A6, A7 and A8 present the corresponding results for the decile portfolio mean returns, Fung and Hsieh (2004) and Chen et al. (2024) alphas, respectively. Again, BRITS imputed data provides information for successfully differentiating high from low-performing hedge funds against the benchmark imputation methods and applying no imputation method. The BRITS long-short predictive portfolio generates 14.00% annualized mean return statistically significant at 1% followed by the corresponding predictive portfolio of the SVT method. The Fung and Hsieh (2004) and Chen et al. (2024) alphas of the long-short predictive portfolios reveal a similar picture. The Fung and Hsieh (2004) and Chen et al. (2024) annualized alphas of the BRITS spread portfolios are 7.46 % and 9.45% significant at 1%, respectively. The BRITS applied only to the predictors, and the SVT imputation method spread portfolios yield the second and third-best alphas again significant at 1%. On the other hand, the XMean and the non-imputation methods always report the lowest long-short portfolio performance.

The out-of-sample cumulative returns of the equally weighted predictive portfolios constructed with the examined imputation methods and no imputation are presented in Figure A1 below. Our findings are similar to those reported in Figure 1 for prediction-weighted

Table A6. Equally-weighted Decile Portfolio Returns

The table provides the out-of-sample returns of the equally-weighted decile portfolios, and top-minus-bottom spread portfolio of hedge funds based on NN predictions across all imputation approaches. Specifically, we show the performance of the portfolios of realised fund returns based on NN's predictions. NN is estimated on the imputed dataset of the "BRITS" and each benchmark imputation method along with the non-imputed dataset. We report the annualized mean returns of the decile portfolios. We use Newey and West (1987) standard errors with three lags to measure the statistical significance of mean returns and alphas. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D10 – D1
NoImp	4.63*	4.12*	4.86**	5.07***	4.94**	5.30***	6.42***	6.57***	7.84***	10.96***	6.34***
BRITS	1.26	3.29**	4.01***	4.31***	4.66**	4.88***	6.24***	8.26***	8.52***	15.26***	14.00***
SVT	1.50	3.49**	4.14***	4.62***	4.82***	5.05***	5.79**	7.77***	9.10***	14.43***	12.93***
XMean	4.07	4.29*	5.58***	5.36**	5.85***	5.90***	6.48***	6.15***	8.77***	8.27***	4.20*
NoImpRetBRITS	2.08	3.04*	4.34**	3.51**	5.29***	5.25***	5.98***	7.88***	8.76***	14.59***	12.51***

Table A7. Equally-weighted Decile Portfolio Fung and Hsieh (2004) alphas

The table provides the out-of-sample Fung and Hsieh (2004) (FH) factor model annualized alphas of the equally-weighted decile portfolios, and top-minus-bottom spread portfolio of hedge funds based on NN predictions across all imputation approaches. Specifically, we show the performance of the portfolios of realised fund returns based on NN's predictions. NN is estimated on the imputed dataset of the "BRITS" and each benchmark imputation method along with the non-imputed dataset. We report the Fung and Hsieh (2004) (FH) factor model annualized alphas for each decile and the spread portfolio. We use Newey and West (1987) standard errors with three lags to measure the statistical significance of mean returns and alphas. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D10 – D1
NoImp	-1.33	-0.83	0.04	0.19	0.02	0.59	1.43	0.97	1.88	3.72*	5.04*
BRITS	-3.01*	-0.58	0.72	0.55	0.33	0.13	0.66	2.09*	1.34	4.45*	7.46***
SVT	-3.65**	-0.62	0.89	0.77	0.46	0.37	0.70	1.93*	2.30*	3.53	7.19***
XMean	-2.10	-1.09	0.26	0.01	0.46	0.25	1.21	0.71	3.89***	3.10*	5.20**
NoImpRetBRITS	-3.10*	-1.07	0.36	-0.29	0.90	0.33	0.57	1.69	2.00	5.28**	8.39***

portfolios. In both BRITS applications, the SVT imputation methods provide more information on hedge fund sorting and the best and worst hedge fund predictability. An investor who has employed BRITS for imputing hedge fund data to invest in the best 10% of funds using a neural network as a forecaster would have earned a cumulative abnormal return of 136% and 130% with BRITS and NoImpRetBRITS, respectively. The same investor would have avoided the worst 10% of funds, generating a corresponding 11% and 18% cumulative return, respectively. The same investor would have realized 128% and 13% by investing in the best and worst 10% of funds, respectively. On the other hand, they would have earned

Table A8. Equally-weighted Decile Portfolio Chen et al. (2024) alphas

The table provides the out-of-sample Chen et al. (2024) (CLTZ) factor model annualized alphas of the equally-weighted decile portfolios, and top-minus-bottom spread portfolio of hedge funds based on NN predictions across all imputation approaches. Specifically, we show the performance of the portfolios of realised fund returns based on NN's predictions. NN is estimated on the imputed dataset of the "BRITS" and each benchmark imputation method along with the non-imputed dataset. We report the Fung and Hsieh (2004) (FH) factor model annualized alphas for each decile and the spread portfolio. We use Newey and West (1987) standard errors with three lags to measure the statistical significance of mean returns and alphas. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	D1	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**	**D10 – D1**
NoImp	-1.48	-0.84	0.78	0.85	0.99	1.65*	2.68***	2.80**	3.47**	6.61***	8.10***
BRITS	-1.97	-0.11	0.98	1.09*	1.04	1.30	1.58	3.23**	2.90**	7.48***	9.45***
SVT	-2.04	-0.17	1.12	1.02	1.52**	1.25	1.66	3.18**	3.81**	6.16***	8.21***
XMean	-2.15	-0.52	1.57	0.75	1.60	1.79*	2.25***	2.28**	4.90***	5.07***	7.22***
NoImpRetBRITS	-1.41	-0.40	0.66	0.25	1.79**	1.26	1.30	3.01**	3.30**	7.76***	9.17***

only 73% and 97% by investing in the best funds with the XMean imputation method and no imputation at all.

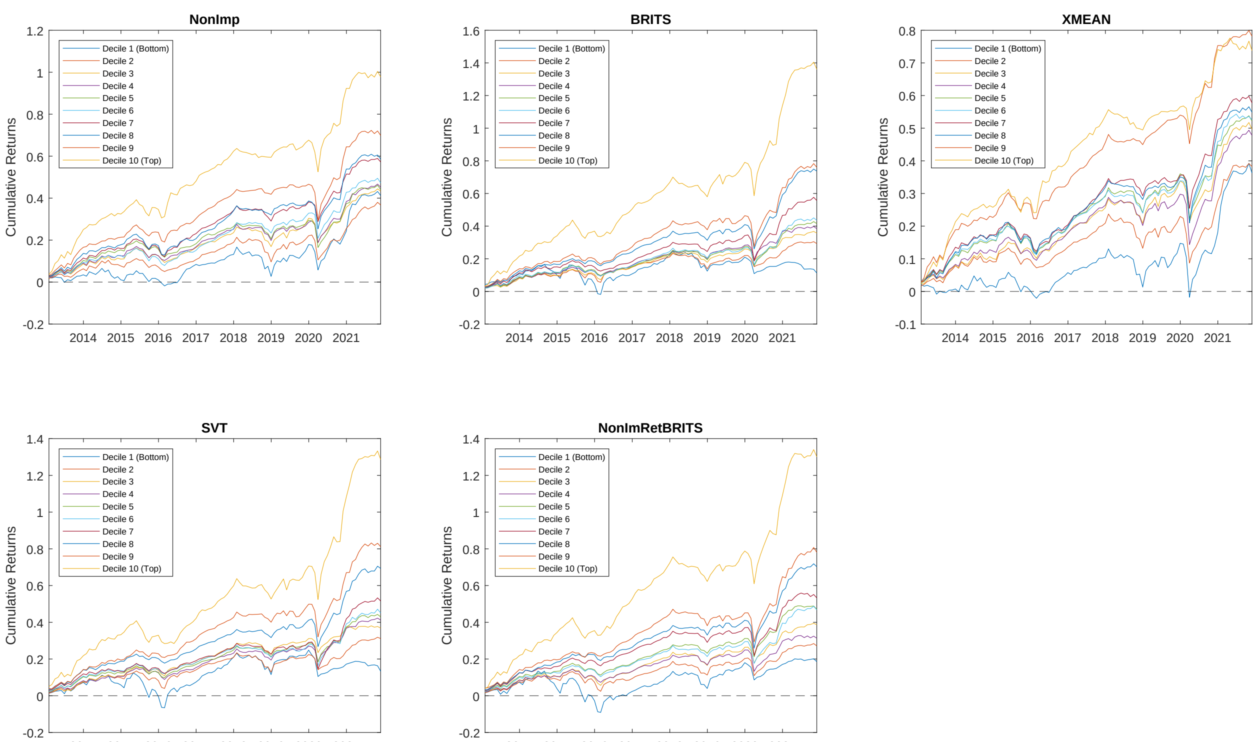

Figure A1. Equally-weighted Decile Portfolio Cumulative Returns

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

Figure A2 reports the performance persistence of the top-decile equally-weighted fund portfolio constructed on the deep neural network predictions using information from each imputation method. Similar to the findings of Figure 2 mean returns and alphas do not monotonically decrease over time but reveal cyclical phases. BRITS provides information that the top-decile portfolio generates the highest mean return and alphas, both statistically significant for at least the first three months, which is the average lock-up period on first-time investors. At the same time the portfolio constructed using no imputation performs poorly. Outstandinly, BRITS portfolios of funds generate the highest alphas after 36 months. Those alphas are statistically significant at least at 5% level.

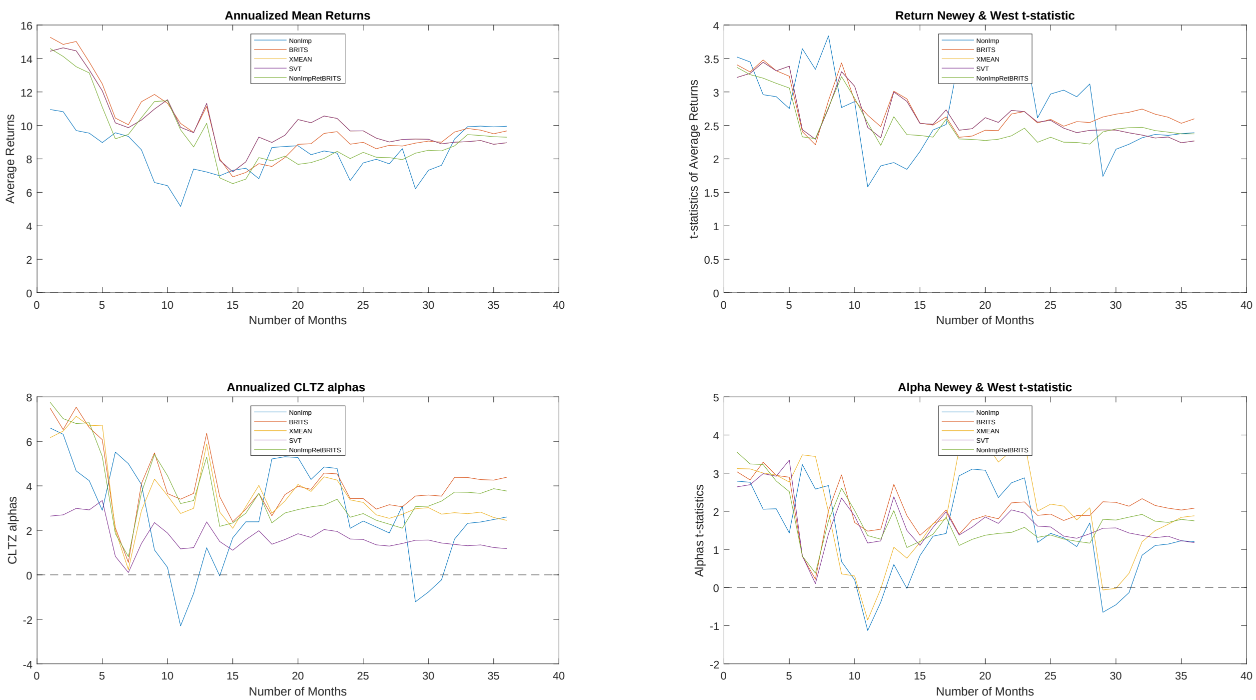

Figure A2. Equally-weighted Top Decile Portfolio Performance Persistence

The figure presents the out-of-sample performance persistence of the NN equally-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.

E. Additonal Empirical Results: LASSO-based Models

E.1. Linear-LASSO Models

E.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:

$$\hat{r}_{i,t+1}^{\text{Lasso}} = \boldsymbol{x}'_{i,t}\hat{\boldsymbol{\eta}}, \tag{7}$$

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:

$$\arg\min_{\boldsymbol{\eta}} \frac{1}{2(t-1)n}\sum_{i=1}^{n}\sum_{s=1}^{t-1}\left(r_{i,s+1} - \boldsymbol{x}'_{i,s}\boldsymbol{\eta}\right)^2 + \lambda\|\boldsymbol{\eta}\|_1, \tag{8}$$

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 $\{10^{-4}, 10^{-1}\}$.

E.1.2. Group LASSO Model

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

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

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 $\{10^{-4}, 10^{-1}\}$.

E.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:

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

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: $\alpha \in \{0, 0.25, 0.5, 0.75, 1\}$.

E.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:

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

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 A9 below.

Table A9: 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	$\lambda \in \{10^{-4}, 10^{-1}\}$ $\gamma \in \{0, 8, 1\}$

E.2. Out-of-sample Portfolio Performance

We provide a selection of the most important empirical findings measuring the informational value of the BRITS method against the benchmarks while using the Linear-Lasso models as our prediction method. We start with the performance report of the top-decile prediction-weighted portfolios constructed using LASSO ensemble predictions. Those have been generated by training the LASSO methods on information provided by each imputation specification and no imputation. We continue by presenting the cumulative returns of their corresponding decile portfolios.

Table A10 reports the relevant results of the prediction-weighted portfolios of top-performing funds as given by the LASSO predictions. Similar to the relevant results for the deep neural networks predictions, BRITS *informed* portfolios generate the highest performance among all the benchmark imputation methods and using no imputation. Interestingly, the second-best imputation method this time is the XMean. The BRITS *informed* portfolio, based on both fund returns' and predictors' recovered data, yields an annualized mean return of 25.11% and a Fung and Hsieh (2004) and a Chen et al. (2024) alpha of 14.32% and 18.15%, all statistically significant at least at 5%. The same portfolio reports an outstanding Sharpe ratio of 1.31.

Table A10. Prediction-Weighted Top-Decile Portfolio Performance of LASSO Predictions

The table reports the out-of-sample performance of the prediction-weighted top-decile portfolio of hedge funds based on the LASSO ensemble predictions across all imputation approaches. Specifically, we show the performance of the top-decile portfolio of realised fund returns based on the predictions of LASSO ensemble estimations. Each machine learning model is estimated on the imputed dataset of the "BRITS" and each benchmark imputation method along with the non-imputed dataset. We report the annualized mean return, the Fung and Hsieh (2004) (FH) and the Chen et al. (2024) (CLTZ) factor model annualized alphas, the maximum drawdown (MaxxDD) the Sharpe ratio, the upside-potential ratio (UPR) and the corresponding information ratios based on the Fung and Hsieh (2004) (FH IR) and the Chen et al. (2024) (CLTZ IR) factor model alphas. We use Newey and West (1987) standard errors with three lags to measure the statistical significance of mean returns and alphas. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively

	Mean (%)	FH alpha	CLTZ alpha	MaxxDD	SR	UPR	FH IR	CLTZ IR
NonImp	22.82***	11.31*	15.48**	25.72	1.26	0.79	0.56	0.73
BRITS	25.11***	14.32**	18.15***	23.02	1.31	0.88	0.68	0.88
XMEAN	23.98***	13.61*	19.75***	23.85	1.29	0.82	0.62	0.89
SVT	22.32***	12.20**	16.69***	20.06	1.22	0.74	0.51	0.70
NonImpRetBRITS	26.94***	15.03**	18.92***	25.02	1.27	0.91	0.68	0.87

Figure A3 presents the corresponding cumulative returns of the decile portfolios under each imputation method and no imputation. BRITS and SVT can provide information that can construct long-short portfolios with the highest dispersion compared to the rest of the imputation methods. Such a finding justifies the ability of BRITS to provide a *complete* dataset used for predicting and accurately distinguishing between the best and worst-performing funds. The BRITS method helps invest in the best-performing funds, yielding a cumulative return of 224%, while avoiding the worst-performing funds providing a return of 27%. The corresponding returns for the SVT and XMean 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%.

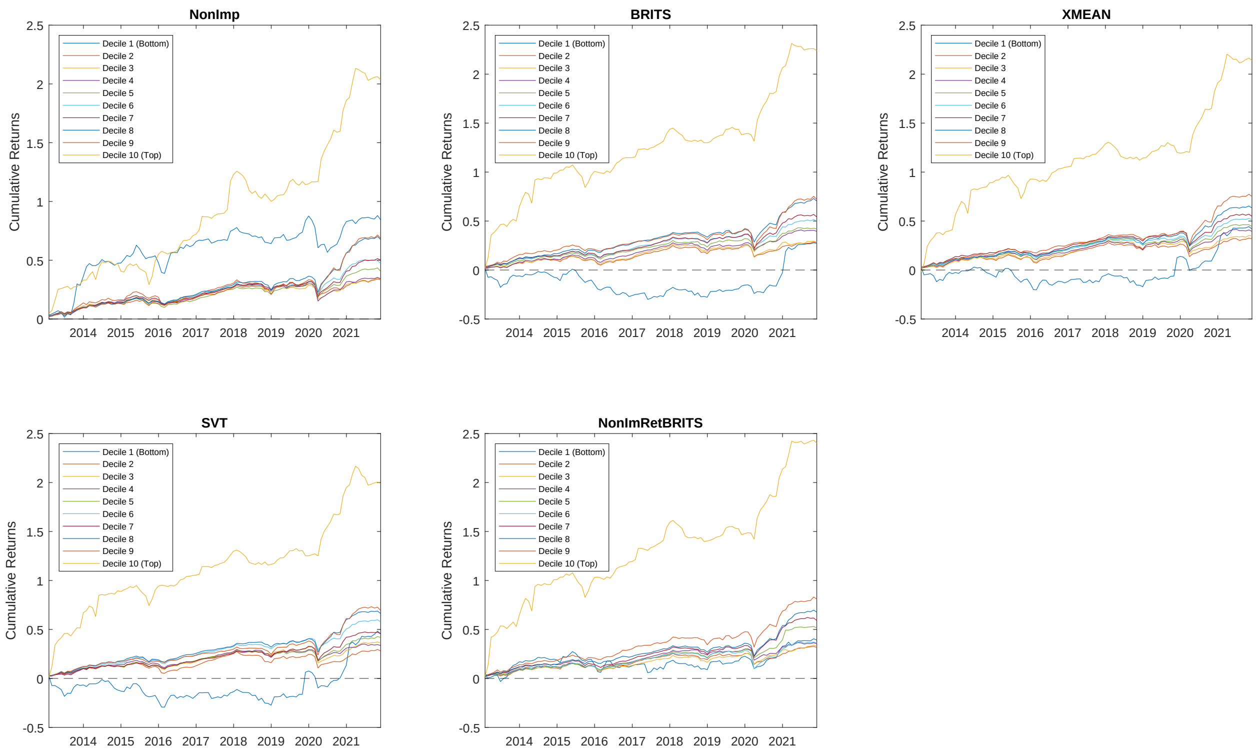

Figure A3. Prediction-weighted Decile Portfolio Cumulative Returns

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