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

Ilias Filippou (Florida State University); Ioannis Psaradellis (University of Edinburgh); David E. Rapach (Federal Reserve Bank of Atlanta); Lazaros Zografopoulos (University of St. Andrews)
Contact: ifilippou@fsu.edu

Introduction

- Missing values are a notorious issue for the empirical analysis of hedge funds.
- The voluntary nature of hedge fund disclosure gives rise to pervasive missing values in hedge fund returns and fund characteristics.
- The surfeit of missing data can adversely affect the empirical analysis of the important topics when it comes to hedge funds, such as return prediction and fund selection from an asset allocation perspective.
- The diverse strategies and proprietary approaches employed by hedge funds make it inherently difficult to identify true managerial skill, exacerbating the challenges posed by missing data.

Contribution

- Introduce BRITS to finance for missing data
- Improve hedge fund return forecasting and fund selection
- Demonstrate economic value of imputation in portfolio performance

Dataset

- 3,800 hedge funds (1994–2021)
- 23 fund characteristics + interactions with 4 macro variables (23 + 4 × 23 = 115 predictors)
- Imputation methods: BRITS, Cross-sectional mean, and SVT
- Forecast models: Deep Neural Networks (NN3, NN4, NN5) and ensemble learning

Methodology

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

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

- Due to the missing values, the data matrix X is incomplete. The goal is to complete the matrix by filling in the missing values.

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

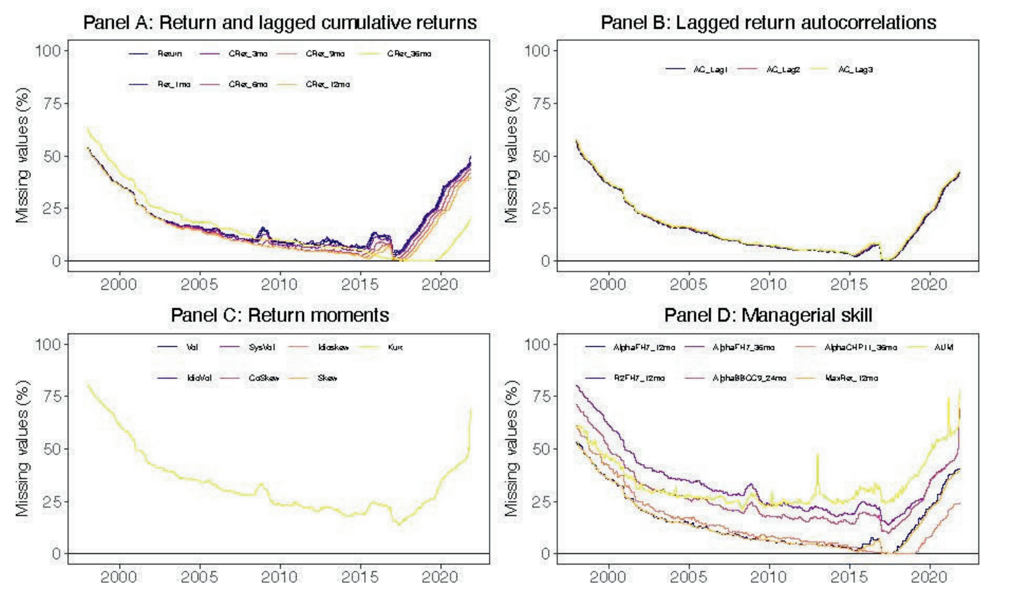
$$
\begin{array}{c|c|c|c|c|c|c|c|c|}
i=1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\
i=2 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
i=3 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\
i=4 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\
 & \boldsymbol{m}_1 & \boldsymbol{m}_2 & \boldsymbol{m}_3 & \boldsymbol{m}_4 & \boldsymbol{m}_5 & \boldsymbol{m}_6 & \boldsymbol{m}_7 & \boldsymbol{m}_8
\end{array}
$$

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

$$
\begin{array}{c|c|c|c|c|c|c|c|c|}
i=1 & 0 & 1 & 1 & 2 & 1 & 2 & 1 & 1 \\
i=2 & 0 & 1 & 2 & 3 & 4 & 1 & 1 & 1 \\
i=3 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 2 \\
i=4 & 0 & 1 & 1 & 2 & 3 & 1 & 2 & 1 \\
 & \boldsymbol{\delta}_1 & \boldsymbol{\delta}_2 & \boldsymbol{\delta}_3 & \boldsymbol{\delta}_4 & \boldsymbol{\delta}_5 & \boldsymbol{\delta}_6 & \boldsymbol{\delta}_7 & \boldsymbol{\delta}_8
\end{array}
$$

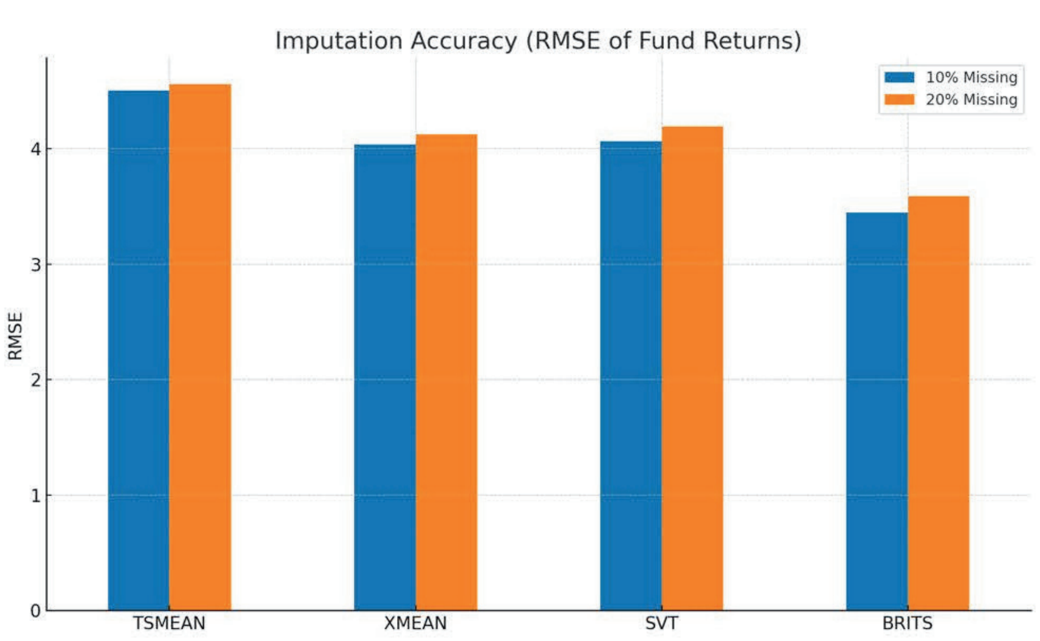
- We construct the matrix triplet $\{X, M, \Delta\}$ for the hedge fund returns as well as each of the 23 fund characteristics in our panel of data.

- The matrix triplet is subsequently used in the BRITS algorithm to fill in the missing values and thereby complete the data matrix $\boldsymbol{X}$ for each of the 24 variables.



Imputation Accuracy

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



Prediction Model

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

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

where $r_{i,t}$ is the month-t excess return for hedge fund i, $\boldsymbol{x}_{i,t} = [x_{i,1,t} \ldots x_{i,K,T}]'$ is a K-dimensional vector of predictor variables, $f(\boldsymbol{x}_{i,t}; \boldsymbol{\theta})$ is the conditional expectation (i.e., prediction) function that depends on the parameter vector $\boldsymbol{\theta}$, and $\varepsilon_{i,t+1}$ is a mean-zero and serially uncorrelated additive disturbance term.

- The out-of-sample excess return forecast is given by

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

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

- Our core model for hedge fund return prediction is a feedforward neural network.
- A neural network consists of multiple layers of interconnected neurons.
    - The first layer consists of the predictors (which total 115 in our models, corresponding to the 23 fund characteristics and their interactions with the four economic variables).
    - One or more hidden layers follow, while the final layer produces the prediction. Each hidden layer consists of P_l neurons, where the mth neuron in layer l computes a nonlinear transformation of the inputs from the previous layer:

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

where $h_j^{(0)} = x_{j,i,t}$ corresponds to the predictors, $\omega_{m,0}^{(l)}$ and $\omega_{m,j}^{(l)}$ are network weights, and $g(.)$ is a nonlinear activation function.

- We use the Gaussian error linear unit (GELU) activation function (Hendrycks and Gimpel 2023): $g(x) = x\Phi(x)$, where $\Phi(x)$ is the standard Gaussian cumulative distribution function. GELU often offers performance gains relative to the popular rectified linear unit (ReLU) activation function. The final layer transforms the neurons in the last hidden layer into a prediction:

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

where $\omega_0^{(L+1)}$ and $\omega_j^{(L+1)}$ are additional weights. We consider deep neural networks with $L = 3, 4, 5$ hidden layers.

- Each network is trained by estimating the weights using stochastic gradient descent (SGD) with the objective of minimizing the MSE over the training sample. We implement SGD via the Adam optimizer.
- We consider an ensemble prediction that is the average of the three predictions (NN-Ensemble)
- We generate a sequence of out-of-sample neural network predictions using a rolling estimation window of 15 years (180 months).

Statistical Accuracy

- We use a pooled version of the popular out-of-sample R^2 statistic to measure the out-of-sample accuracy of hedge fund return forecasts based on a neural network and a data imputation method relative to a benchmark return forecast. The pooled version of out-of-sample R^2 statistic for comparing a neural network forecast to a zero benchmark forecast is given by

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

where t_{in} $(T)$is the end of the initial in-sample (total sample) period. The $R_{OS,Pool}^2$ measures proportional reduction in out-of-sample MSE for the neural network vis-a-vis the benchmark forecast across all of the hedge funds.

(1) Model	(2) No data imputation	(3) CS-Mean	(4) SVT	(5) BRITS	(6) BRITS (no return imputation)
NN3	0.16	−0.09	1.68***	1.58***	1.26
NN4	0.01	0.05	1.43	1.59***	2.04***
NN5	0.19	0.58	2.76***	2.45***	2.68***
NN-Ensemble	0.16	0.29	2.28***	2.16***	2.25***

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).

Long-Short portfolio performance

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

Top-decile portfolio performance

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

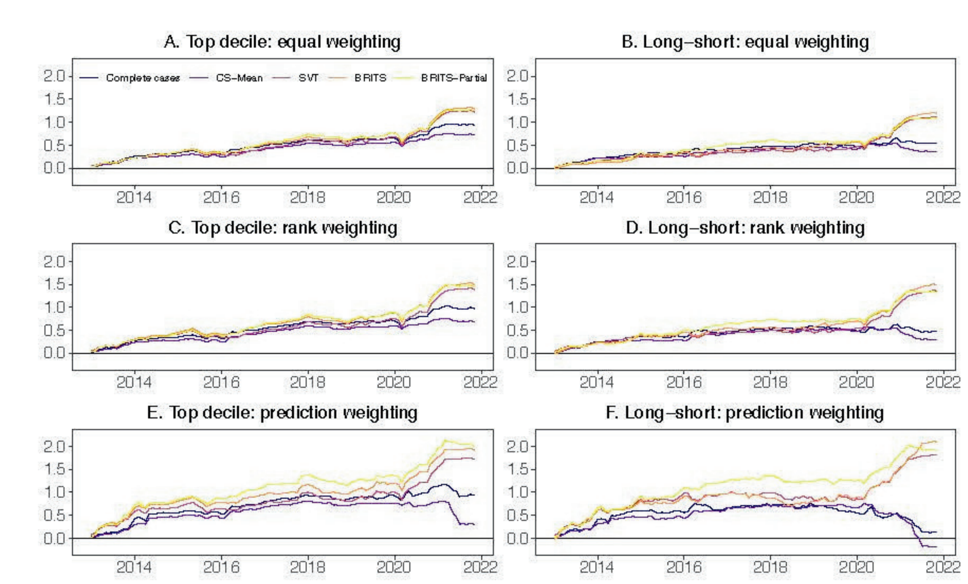


Figure 2. Log cumulative excess returns. Each panel depicts the log cumulative excess return for the portfolio (top decile or long-short) and weighting scheme (equal, rank, or prediction) in the panel heading.

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.

- 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.