INSTITUTE
OF ECONOMIC STUDIES
Faculty of Social Sciences
Charles University

AUTOENCODER ASSET PRICING MODELS AND ECONOMIC RESTRICTIONS - INTERNATIONAL EVIDENCE

Lenka Nechvátalová

IES Working Paper 26/2024

Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Bibliographic information:
Nechvátalová L. (2024): " Autoencoder Asset Pricing Models and Economic Restrictions - International Evidence" IES Working Papers 26/2024. IES FSV. Charles University.

Autoencoder Asset Pricing Models and Economic Restrictions – International Evidence

Lenka Nechvátalová

Institute of Economic Studies, Faculty of Social Sciences, Charles University &
Institute of Information Theory and Automation, Czech Academy of Sciences
E-mail: lenka.nechvatalova@fsv.cuni.cz

August 2024

Abstract:
We evaluate the performance of the Conditional Autoencoder (CAE) model by Gu et al. (2021) in an international context and under economic constraints, such as the exclusion of microcap and illiquid firms, and accounting for transaction costs. The CAE model leverages latent factors and factor exposures dependent on asset characteristics, modelled as a flexible nonlinear function while adhering to the noarbitrage condition. The original study showed significant reductions in out-ofsample pricing errors from both statistical and economic perspectives in the U.S. context. We replicate these results on the U.S. dataset and extend the analysis to international data with a different set of firm characteristics, achieving consistent outcomes that demonstrate the model's robustness. However, the economic benefits after accounting for transaction costs are limited, even after the exclusion of illiquid firms, highlighting the importance of considering these constraints.

JEL: G11, G12, G15, C55
Keywords: Machine learning, asset pricing, economic restrictions, anomalies

Acknowledgement: The financial support was provided by the Grant Agency of Charles University (grant number 316521) and the Czech Science Foundation under the 19-28231X (EXPRO) project.

1 Introduction

Understanding the underlying factors that drive asset returns is a pivotal challenge in both academic research and practical finance. While traditional factor models are well-established tools for capturing these dynamics, the Conditional Autoencoder (CAE) proposed by Gu et al. (2021) represents a significant advancement. The CAE uses neural networks from the autoencoder family to perform dimension reduction, estimating latent factors and capturing nonlinear factor exposures based on asset characteristics. It ensures no-arbitrage and significantly reduces out-of-sample pricing errors compared to traditional linear models.

In this study, we study how well the CAE model generalizes to data and scenarios not covered in the original work. Our contribution is twofold: first, we assess the model under economic constraints; second, we apply it to an international context. Specifically, we investigate the effects of the exclusion of microcaps and illiquid firms as well as the inclusion of transaction costs. We find that while excluding illiquid firms decreases portfolio profitability, accounting for transaction costs actually increases net returns on the liquid sample compared to the full sample. Notably, in the U.S., the profitability of the short side of the portfolio is heavily concentrated in illiquid and small firms, a pattern that is also observed internationally but to a lesser extent. By applying the model to international markets, we observe results consistent with the U.S. results, further validating the model's applicability.

The CAE model leverages neural networks from the autoencoder family to assist in dimension reduction, capturing the intricate dynamics of asset returns through a flexible nonlinear function of covariates while adhering to the no-arbitrage condition. This model builds on the earlier Instrumental Principal Component Analysis (IPCA) model by Kelly et al. (2019), which also incorporates firm characteristics but under linear assumptions. Relaxing this assumption in the CAE model allows for a more nuanced understanding of the factors driving asset returns. The original study demonstrated superior results in terms of statistical criteria, achieving higher total and predictive R^2, and economic performance, evidenced by higher Sharpe ratios of portfolios, compared to simpler models.

Recent literature has explored various machine-learning approaches to asset pricing, highlighting the broader potential of these techniques. To name some influential examples, Bryzgalova et al. (2019) and Gu et al. (2020) provide insights into using advanced machine learning methods, demonstrating significant improvements in predictive accuracy. Building on these advancements, Yang et al. (2024) further refined the CAE model by introducing the Conditional Quantile Variational Autoencoder (CQVAE), which enhances the accuracy of mean return estimates by focusing on conditional quantiles.

Even though the application of machine learning has significantly improved the predictability of stock returns in recent years, it is essential to consider liquidity constraints for the practical applications of asset pricing models. Hou et al. (2020) examine individual anomalies and their replicability under various liquidity constraints, emphasizing the importance of market conditions. Similarly, Avramov et al. (2023) investigate the impact of economic constraints on the performance of machine learning models in the U.S. market, finding that excluding microcaps, distressed stocks, or periods of high market volatility significantly reduces profitability. Furthermore, literature focused on transaction costs underscores the need to account for trading expenses in asset pricing applications. DeMiguel et al. (2020) examine the transaction costs of multiple anomalies, while Novy-Marx and Velikov (2019) and Nechvátalová (2024) consider various cost mitigating techniques and their performance.

The majority of asset pricing research is conducted in the U.S. only (Andrew Karolyi, 2016). International evidence thus provides crucial out-of-sample validation, as results found in the U.S. do not necessarily translate to other markets. Building upon the methodology of Gu et al. (2020), Tobek and Hronec (2021) apply machine learning techniques to developed countries, while Hanauer and Kalsbach (2023) focus on emerging markets, demonstrating the broader applicability of these methods. Moreover, Jiang et al. (2023) show that price patterns can predict returns both in the U.S. and internationally, further validating the potential of machine learning in diverse market conditions.

While Gu et al. (2021) demonstrated the effectiveness of the CAE model using the U.S. equity dataset, our study replicates their results and extends the analysis to an international context, as well as introducing economic constraints critical for practical asset management. Additionally, we employ a different set of firm characteristics for the international analysis, further validating the model's robustness across diverse datasets.

Consistent with the findings of Avramov et al. (2023), we show that portfolio returns and profitability decrease when restricting our analysis to a liquid subset of stocks in the U.S. market. However, after accounting for transaction costs, the net returns and profitability actually increase in the liquid sample compared to the full sample. This indicates that focusing on more liquid stocks can enhance overall portfolio performance once transaction costs are considered. Using predictions from the CAE model to create portfolios offers limited economic benefit in the U.S., where the profitability of the short side is concentrated in illiquid firms. Consequently, only the long-only decile portfolio on the liquid sample outperforms the S&P index after transaction costs, achieving a Sharpe ratio of 0.75 (equal-weighted) and 0.81 (value-weighted).

Extending our analysis to international markets, we find results that align with the U.S. outcomes in both statistical performance, measured by total and predictive R^2, and economic performance, reflected in portfolio returns. In contrast to the U.S. results, for international portfolios, the long-short strategy on the liquid sample is more effective, with a Sharpe ratio of 0.91 (equal-weighted) and 0.63 (value-weighted), both outperforming the market index. This demonstrates the CAE model's robustness and reliability across different datasets and firm characteristics, providing valuable insights into the practical applicability of advanced machine learning techniques in asset pricing. It highlights the importance of considering economic constraints in practical asset management, as these factors significantly influence portfolio performance.

The rest of this paper is organized as follows: Section 2 provides methodology and dataset description. Section 3 presents empirical results for the U.S. dataset and internationally, and finally Section 4 concludes the paper.

2 Methodology and Data

2.1 Conditional Autoencoder model

In this section, we describe the CAE model used in our study. Introduced by Gu et al. (2021), it incorporates firm characteristics and characteristic-managed portfolios to estimate latent factors and conditional betas. The main equation, which captures the essence of the model's structure, is:

$$r_{i,t} = \beta'_{i,t-1} f_t + u_{i,t}, \tag{1}$$

where $r_{i,t}$ represents the excess return of asset i at time t, $\beta_{i,t-1}$ denotes the factor loadings, f_t represents the latent factors, and $u_{i,t}$ is the error term.

Figure 1 shows the overall structure of the conditional autoencoder model. The left part of the network is modelling conditional betas as a nonlinear function of firm characteristics. The right side shows how we model the latent factors as a linear function of a set of characteristic-managed portfolios x_t. The portfolios x_t are calculated as follows:

$$x_t = \left(Z'_{t-1} Z_{t-1}\right)^{-1} Z'_{t-1} r_t, \tag{2}$$

where Z_{t-1} is the matrix of firm characteristics at time $t-1$, and r_t is the vector of returns at time t. Compared to simply using r_t as input to obtain factor loadings, using x_t reduces the dimensionality dramatically from the number of firms to the number of firm characteristics. It also allows for an incomplete panel dataset, using only non-missing observations to obtain x_t.

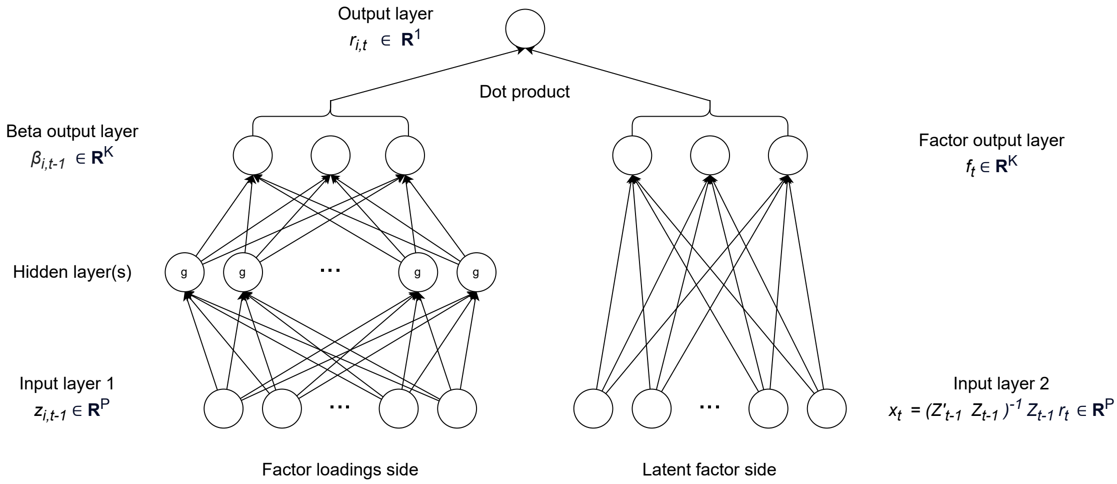

Figure 1: Conditional Autoencoder Model

The left-hand side of the figure illustrates the process of obtaining factor loadings $\beta_{i,t-1}$ using firm characteristics $z_{i,t-1}$ as input. It involves one or more hidden layers with a nonlinear activation function g. The right-hand side shows how P characteristic-managed portfolios x_t are linearly transformed to derive the latent factor f_t.

The conditional betas $\beta_{i,t-1}$ (left side of Figure 1) are derived through a series of transformations of the firm characteristics $z_{i,t-1}$:

$$z_{i,t-1}^{(0)} = z_{i,t-1}, \tag{3}$$

$$z_{i,t-1}^{(l)} = g\left(b^{(l-1)} + W^{(l-1)} z_{i,t-1}^{(l-1)}\right), \quad l = 1, \ldots, L_\beta, \tag{4}$$

$$\beta_{i,t-1} = b^{(L_\beta)} + W^{(L_\beta)} z_{i,t-1}^{(L_\beta)}. \tag{5}$$

Here, $z_{i,t-1}^{(l)}$ represents the firm characteristics transformed through l-th layer of the neural network with the nonlinear activation function g. In the empirical part, we use Rectified Linear Unit (ReLU) as the nonlinear activation function g and one to three hidden layers.

The latent factor f_t (right side of Figure 1) is obtained from the characteristic-managed portfolios x_t:

$$x_t^{(0)} = x_t, \tag{6}$$

$$f_t = \widetilde{b}^{(0)} + \widetilde{W}^{(0)} x_t^{(0)}. \tag{7}$$

Here, $x_t^{(0)}$ is the initial input which is the characteristic-managed portfolios, and Equation 7 shows the linear transformation used to transform x_t into the latent factor f_t.

The final output from the CAE model is obtained by taking the dot product of the conditional betas $\beta_{i,t-1}$ and the latent factors f_t.

2.2 Data

U.S. Equity Dataset

For the U.S., we use individual stock returns from the CRSP database, with the three-month Treasury bill rate serving as a proxy for the risk-free rate to calculate excess returns. The dataset spans from 1957 to 2018.

We incorporate 94 firm characteristics as used and provided by Gu et al. (2021, 2020), comprising 61 annual, 13 quarterly, and 20 monthly updated characteristics. To avoid forward-looking bias, the characteristics are appropriately lagged. For a comprehensive list of characteristics and additional details, refer to Gu et al. (2020).

International Equity Dataset

The international dataset includes the U.S. and 22 other developed countries[1]. These countries are further divided into four regions: U.S., Europe, Japan, and Asia Pacific. The international dataset is obtained from Datastream by Refinitiv. The data preprocessing procedure follows Nechvátalová (2024).

For the firm characteristics in the international dataset, we utilize a set of 153 anomalies used by Tobek and Hronec (2021), comprising 93 fundamental characteristics, 43 market friction characteristics, and 11 I/B/E/S characteristics. The fundamental characteristics are based solely on annual data. To ensure consistency between the U.S. and international datasets, we exclude quarterly fundamental characteristics due to their limited availability internationally. Refer to Tobek and Hronec (2021) for comprehensive implementation details.

Liquidity filters

For the U.S. full sample, we do not apply additional filtering to maintain consistency with the dataset used by Gu et al. (2021). For the international full sample, we exclude observations with a price lower than \$1 (\$0.10 for Asia Pacific) at the end of the previous month.

The liquid sample excludes microcap stocks, discarding illiquid and small firms that would be costly or impossible to trade. First, we sort firms by market capitalization and exclude the lowest market-cap firms each month. Specifically, we remove the least capitalized firms until the excluded firms' total market capitalization equals 5% of the region's total market capitalization. We also apply a similar filter based on trading volume over the last 12 months. Firms with low trading volumes are excluded until the total trading volume of the excluded firms equals 5% of the region's total traded volume. If a firm lacks trading volume data, it is excluded if it falls within the lowest 10% based on market capitalization. For non-U.S. stocks, we require a market capitalization exceeding the lowest decile of NYSE market cap for the given month, ensuring non-U.S. firms have capitalization levels comparable to U.S. stocks. As in the full sample, firms must have a stock price exceeding one dollar at the end of the previous month, or \$0.10 for firms in the Asia Pacific region.

1. Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, Japan, Australia, New Zealand, Hong Kong, and Singapore.

Missing firm characteristics are replaced by the cross-sectional median of that characteristic within the region. To stabilize learning and mitigate the influence of outliers, we cross-sectionally rank normalize all characteristics to be within the (-1, 1) interval for each region. For excess returns, we clip observations below -3 and above 3 to avoid the effect of severe outliers.

Table 1 presents descriptive statistics for both datasets for the full sample and the liquid subsample. In the U.S., the number of firms per month is on average over 5000 for the full sample and around 1000 for the liquid subsample. Internationally, we have on average almost 17000 firms in the full dataset each month, which reduces to 3000 firms in the liquid sample.

Table 1: Descriptive statistics for U.S. and international datasets

Panel A presents descriptive statistics for the U.S. dataset spanning the period from 1957 to 2018. Panel B covers the international dataset, including the U.S. and 22 developed countries, from 1990 to 2018. We report the monthly mean, standard deviation, and the 25%, 50%, and 75% quantiles for excess returns (in %), market capitalization (in millions of dollars), and the number of firms per month. The statistics are provided for both the full sample (left) and the liquid subsample (right).

	Full Sample			Liquid Sample		
Statistic	Excess Return	MC	# Firms	Excess Return	MC	# Firms
Panel A: U.S. dataset						
Mean	-2.79	1504.49	5242	-3.10	6008.72	1097
Std	17.72	10318.30	2267	11.72	22232.25	249
25%	-10.35	21.13	2594	-9.24	348.30	945
50%	-3.45	90.70	5704	-3.10	1152.09	1041
75%	3.27	445.87	6787	2.95	3746.23	1248
Panel B: international dataset						
Mean	-1.30	1665.98	16918	-1.81	7946.65	2998
Std	20.72	10377.43	1195	11.92	22755.74	431
25%	-8.98	25.40	16494	-7.91	957.37	2655
50%	-2.39	101.42	17102	-1.87	2290.75	3009
75%	4.24	489.65	17812	4.08	6034.42	3223

2.3 Model estimation

To estimate our model, we divide our dataset into training, validation and testing datasets while maintaining the temporal ordering of the data. We train the model with various hyperparameter configurations on the training set and evaluate them on the validation set. The best set of hyperparameters is based on mean square error over the validation sample. The testing sample is used to obtain out-of-sample predictions on previously unseen data.

To avoid overfitting, we follow the Gu et al. (2021) approach, using l_1 LASSO penalization and taking an ensemble of ten best models based on the validation sample. The objective function to optimize the parameters of the model is:

$$\mathcal{L}(\theta; \cdot) = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \left\| r_{i,t} - \beta'_{i,t-1} f_t \right\|^2 + \lambda \sum_{j} |\theta_j| \quad (8)$$

where the second part of the objective function is the regularization term, applying the LASSO penalty to the network's weights. The regularization strength parameter λ is determined using the validation sample during hyperparameter tuning. We implement early stopping to terminate training if the validation loss increases, preventing overfitting. The Adam algorithm (Kingma and Ba, 2014), an extension of stochastic gradient descent with adaptive learning rates, is used for optimization. Additionally, batch normalization (Ioffe and Szegedy, 2015) is applied to each hidden layer to stabilize and accelerate training.

In our analysis, we implement an expanding window approach, sequentially training and evaluating our models over various time periods. This method allows us to obtain more precise estimates by incorporating more recent data. For the training of our first model, we divide the dataset into an 18-year training period (1957–1974), a 12-year validation period (1975–1986), and a 31-year out-of-sample testing period (1987–2018). To save computational costs, we refit the model every five years, extending the training sample by five years each time while keeping the validation sample size fixed.

In our analysis, we keep the number of latent factors $K = 3$. We perform a hyperparameter search to select a suitable model. The number of hidden layers on the factor loadings side is either one, two, or three, with 32 neurons in the first hidden layer, 16 in the second, and 8 in the third. The learning rates checked were 0.01, 0.001, 0.0001, and 0.00001. The strength of the LASSO regularization λ can be 0, 0.0001, 0.00001 or 0.000001. The batch size is either 64 or 128.

3 Empirical results

3.1 Statistical performance

To evaluate the out-of-sample statistical performance of the model, we use the total and predictive R^2 metrics proposed by Kelly et al. (2019).

The total R^2 measures how well the model explains the current factor realizations, evaluating its effectiveness in capturing individual stock riskiness:

$$R^2_{\text{total}} = 1 - \frac{\sum_{(i,t)\in \text{OOS}}(r_{i,t} - \hat{\beta}'_{i,t-1}\hat{f}_t)^2}{\sum_{(i,t)\in \text{OOS}} r^2_{i,t}}. \tag{9}$$

where OOS denotes the out-of-sample testing subsample for a given model, comprising data not used for training the model or selecting hyperparameters

The predictive R^2 evaluates how accurately the model forecasts future individual excess stock returns, indicating its capability to capture variations in risk compensation across the panel. It is defined as:

$$R^2_{\text{pred}} = 1 - \frac{\sum_{(i,t)\in \text{OOS}}(r_{i,t} - \hat{\beta}'_{i,t-1}\hat{\lambda}_{t-1})^2}{\sum_{(i,t)\in \text{OOS}} r^2_{i,t}}, \tag{10}$$

where $\hat{\lambda}_{t-1}$ is exponentially weighted moving average of the estimated latent factor $\hat{f}$ up to month $t-1$ with smoothing factor $\alpha = 0.5$.

In Table 2 is the total and predictive R^2 for both datasets. We replicate the results of Gu et al. (2021) using the U.S. dataset, achieving a total R^2 of 14.70% and a predictive R^2 of 0.32%. In comparison, Gu et al. (2021) report a total R^2 of 12.5% and a predictive R^2 of 0.52%. These differences may be attributed to the varying set of hyperparameters tested and a slightly different out-of-sample period, as their dataset ends two years earlier. Unlike Gu et al. (2021), who separate the results for models with one, two, or three hidden layers, we aggregate these models in our analysis. This likely does not affect our results, as they report very similar R^2 values across these models.

For the international dataset, we observe a total R^2 of 9.06%, and the predictive R^2 is 0.52%. These results demonstrate the model's effectiveness in capturing the complexity of asset returns across different markets, highlighting its robustness and applicability beyond the U.S. context.

Table 2: Out-of-sample R^2 values

Out-of-sample R^2_{total} and R^2_{pred} for individual stocks are reported. The U.S. dataset covers the period from 1987 to 2018 and the international period from 1995 to 2018. Values are in percentages.

Dataset	R^2_{total}	R^2_{pred}
U.S.	14.70	0.32
International	9.06	0.52

3.2 Evidence from the U.S. equity dataset

To evaluate the economic performance of our models we construct long-short decile portfolios based on the out-of-sample predicted returns. Each month, we sort the return forecasts and buy the top 10% while selling the bottom 10% of the firms.

To estimate transaction costs, we utilize the closing quoted spread as described by Chung and Zhang (2014). For missing observations, we fill them using the volatility over volume method proposed by Fong et al. (2018). Any remaining gaps are assumed to have a transaction cost rate of 5%. Portfolio returns with transaction costs are calculated iteratively to account for these estimated firm-month-specific transaction costs (Nechvátalová, 2024).

The portfolio turnover is the sum of absolute values of trade sizes divided by the gross exposure. A turnover of 200% in a given month indicates that all currently held positions were liquidated and new firms were added to the portfolio on both the long and short sides.

Table 3 reports monthly mean returns, standard deviation, annualized Sharpe ratio, and monthly portfolio turnover. We report equal-weighted as well as value-weighted portfolios, trained on the full dataset. We show performance over the full sample and liquid subsample, with and without transaction costs.

The results for the equal-weighted and value-weighted portfolios differ significantly when evaluated on the full sample. For the liquid samples, the performance of value-weighted and equal-weighted portfolios is comparable. As we restrict our analysis to a liquid subsample, the returns are reduced by approximately half. The decrease in the Sharpe ratio is more pronounced for the equal-weighted portfolio due to the greater impact of small firms.

Table 3: Performance of portfolios in the U.S.

This table presents monthly mean returns, standard deviations, annualized Sharpe ratios, and turnover rates for U.S. equal-weighted (Panels A, B, C) and value-weighted portfolios (Panels D, E, F) from 1987 to 2018. The results are shown for long-short, long-only, and short-only decile portfolios, both with and without transaction costs. Models are trained on the full sample and evaluated on either the full sample or a liquid subsample. All values are expressed as percentages, except for the Sharpe ratio.

Equal-weight		Without Transaction Costs			With Transaction Costs			
Training	Testing	Mean	Std	Sharpe	Mean	Std	Sharpe	Turnover
				Panel A: Long-Short				
Full	Full	3.69	5.04	2.54	-1.68	5.02	-1.16	126
Full	Liquid	1.37	6.53	0.73	0.75	6.50	0.40	116
				Panel B: Long side				
Full	Full	2.82	8.78	1.11	-0.29	8.44	-0.12	124
Full	Liquid	1.51	5.45	0.96	1.18	5.42	0.75	122
				Panel C: Short side				
Full	Full	0.87	8.16	0.37	-1.39	7.98	-0.60	128
Full	Liquid	-0.14	8.78	-0.06	-0.42	8.76	-0.17	109
Value-weight		Without Transaction Costs			With Transaction Costs			
Training	Testing	Mean	Std	Sharpe	Mean	Std	Sharpe	Turnover
				Panel D: Long-Short				
Full	Full	2.14	6.68	1.11	-0.86	6.68	-0.45	137
Full	Liquid	1.28	5.94	0.75	0.84	5.91	0.49	129
				Panel E: Long side				
Full	Full	1.93	7.67	0.87	-0.10	7.58	-0.04	149
Full	Liquid	1.47	5.18	0.98	1.21	5.17	0.81	141
				Panel F: Short side				
Full	Full	0.21	7.71	0.09	-0.77	7.65	-0.35	125
Full	Liquid	-0.19	7.82	-0.08	-0.37	7.80	-0.17	118

When transaction costs are included, the full sample portfolio strategies are not profitable. However, strategies trading on liquid samples remain profitable but with much lower Sharpe ratios, at 0.4 for the equal-weighted and 0.5 for the value-weighted portfolios. For comparison, investing in a buy-and-hold S&P would yield a mean monthly return of 0.64% and a Sharpe ratio of 0.53.

Examining the long-only portfolio metrics, the long side is more profitable in the liquid subsample than in the long-short portfolio. As expected, both the short-leg and long-leg of the portfolio are less profitable without transaction costs in the liquid sample compared to the full sample, as we exclude microcap and illiquid firms. For both sides of the portfolio, after accounting for transaction costs, the liquid subsample has higher returns and Sharpe ratios compared to the full sample.

The results evaluated on the full sample are consistent with Gu et al. (2021) for both equal-weighted and value-weighted portfolios. Avramov et al. (2023) report results for value-weighted portfolio returns of CAE with two hidden layers and five latent factors across various subsets; however, their full sample metrics are significantly worse than those reported by us or by Gu et al. (2021). Their reported Sharpe ratio is 0.78 compared to 1.45 in the study by Gu et al. (2021) for a comparable model. They also have mean returns one percentage point lower compared to our model with only three latent factors. Our portfolio's performance on the liquid sample is very similar to their 'nonmicrocaps' sample, even though the liquid sample includes only half the number of firms, with their 'nonmicrocaps' sample having over 13,000 firms. This suggests that their models might be suboptimal and underperforming. Nevertheless, they also observe the common pattern of decreasing profitability in more liquid samples without transaction costs.

The turnover of the long-short strategy is 137% and 129% for the value-weighted portfolio on the full and liquid sample, respectively, which is comparable to Avramov et al. (2023) results.

As the current model was trained on the full sample dataset, the performance of the portfolios is likely to be improved by training specifically on the liquid subsample.

3.3 International evidence

In this section, we report results for the international dataset, which includes the U.S. and 22 other developed countries. The out-of-sample period spans from 1995 to 2018 due to data availability.

Table 4 presents performance metrics for the equal-weighted and value-weighted long-short decile portfolios. Values are shown for both the full and liquid samples, as well as before and after accounting for transaction costs. For the full sample without transaction costs, Sharpe ratios are higher compared to the U.S. results, with Sharpe ratios of 4.49 and 1.40 for the equal-weighted and value-weighted portfolios, respectively. This difference mostly disappears after considering transaction costs, but the value-weighted global portfolio remains the only one with positive returns (0.45%) on the full sample, albeit with a Sharpe ratio of only 0.35. For comparison, investing in a buy-and-hold strategy of the MSCI World Index would yield a monthly mean return of 0.45% with a Sharpe ratio of 0.37 for the same period.

When we restrict the investing universe to the liquid sample, we observe the same pattern as with the U.S. portfolios. Before including transaction costs, the returns and Sharpe ratios are lower for the liquid subsample, but this reverses after including transaction costs, where the liquid subsample becomes superior. The equal-weighted portfolio has a Sharpe ratio of 0.91 for the liquid subsample after transaction costs (0.63 for the value-weighted). For the international portfolios, there is a greater difference between the equal-weighted and value-weighted portfolios, even in the liquid subsample after transaction costs. This could be due to better investment options with a larger number of firms of different sizes across regions. As we include firm-specific transaction costs, these estimates should be reasonably reliable, especially for the liquid sample.

The turnover of portfolios is slightly higher but similar to that of the U.S. portfolios. Compared to the U.S. case, the long-short portfolio performs better than the long-only portfolio, as the short side of international portfolios is more effective.

Overall, the results for the international portfolios are very similar to those for the U.S. portfolios, suggesting that the conditional autoencoder model is robust across different datasets and firm characteristics. However, we find limited economic benefit after accounting for transaction costs or when restricting to the liquid subsample.

Table 4: Performance of Portfolios Internationally

This table presents monthly mean returns, standard deviations, annualized Sharpe ratios, and turnover metrics for international portfolios from 1995 to 2018. It includes results for equal-weighted portfolios (Panels A, B, C) and value-weighted portfolios (Panels D, E, F). The results are shown for long-short, long-only, and short-only decile portfolios, both with and without transaction costs. Models are trained on the full sample and evaluated on either the full sample or a liquid subsample. All values are expressed as percentages, except for the Sharpe ratio.

Equal-weight		Without Transaction Costs			With Transaction Costs			
Training	Testing	Mean	Std	Sharpe	Mean	Std	Sharpe	Turnover
				Panel A: Long-Short				
Full	Full	5.16	3.98	4.49	-1.40	3.69	-1.31	126
Full	Liquid	1.61	3.66	1.52	0.96	3.64	0.91	128
				Panel B: Long side				
Full	Full	3.97	7.61	1.81	-0.05	7.22	-0.02	125
Full	Liquid	1.21	5.93	0.71	0.89	5.91	0.52	134
				Panel C: Short side				
Full	Full	1.18	6.01	0.68	-1.35	5.75	-0.81	127
Full	Liquid	0.40	6.07	0.23	0.06	6.03	0.04	123
Value-weight		Without Transaction Costs			With Transaction Costs			
Training	Testing	Mean	Std	Sharpe	Mean	Std	Sharpe	Turnover
				Panel D: Long-Short				
Full	Full	1.80	4.45	1.40	0.45	4.43	0.35	147
Full	Liquid	1.24	4.13	1.04	0.75	4.12	0.63	136
				Panel E: Long side				
Full	Full	1.41	6.67	0.73	0.69	6.67	0.36	158
Full	Liquid	1.07	5.47	0.68	0.84	5.47	0.53	141
				Panel F: Short side				
Full	Full	0.39	6.07	0.22	-0.23	6.01	-0.13	136
Full	Liquid	0.17	5.70	0.11	-0.09	5.68	-0.06	132

4 Conclusion

In this paper, we extended the application of the CAE model, demonstrating its ability to capture complex nonlinear relationships in asset returns. We evaluated the model under economic constraints and applied it to international markets. Our findings provide a comprehensive assessment of the CAE model's robustness and practical applicability, showing its effectiveness across different datasets and highlighting the importance of incorporating liquidity and transaction costs in asset pricing models.

Our findings reveal that excluding microcap and illiquid firms decreases portfolio profitability, a result documented in the U.S. context by Avramov et al. (2023). However, when accounting for transaction costs, net returns actually increase in the liquid sample compared to the full sample. This suggests that focusing on more liquid stocks can enhance overall portfolio performance once transaction costs are considered.

Internationally, the CAE model's performance aligns with U.S. outcomes in both statistical and economic metrics. While the model offers limited economic benefit in the U.S. due to the concentration of short-side profitability in illiquid firms, the long-only decile portfolio on the liquid sample still outperforms the S&P index after transaction costs. More promising results are observed for international portfolios, where the long-short strategy on the liquid sample achieves solid Sharpe ratios, outperforming the market index.

Our work underscores the potential of advanced machine learning techniques, such as the CAE model, in the field of asset pricing. It highlights both the strengths and limitations of these approaches, providing valuable insights for future research and practical asset management. We demonstrate that incorporating factors such as liquidity and transaction costs is crucial for a nuanced understanding of asset returns, as these elements can significantly alter the final conclusions. Additionally, by including international evidence, we show that the model is generalizable beyond the U.S. dataset.

References

Andrew Karolyi, G. 2016. "Home bias, an academic puzzle." *Review of Finance* 20 (6): 2049–2078.

Avramov, Doron, Si Cheng, and Lior Metzker. 2023. "Machine learning vs. economic restrictions: Evidence from stock return predictability." *Management Science* 69 (5): 2587–2619.

Bryzgalova, Svetlana, Markus Pelger, and Jason Zhu. 2019. "Forest through the trees: Building cross-sections of stock returns." *Available at SSRN 3493458.*

Chung, Kee H, and Hao Zhang. 2014. "A simple approximation of intraday spreads using daily data." *Journal of Financial Markets* 17:94–120.

DeMiguel, Victor, Alberto Martin-Utrera, Francisco J Nogales, and Raman Uppal. 2020. "A transaction-cost perspective on the multitude of firm characteristics." *The Review of Financial Studies* 33 (5): 2180–2222.

Fong, Kingsley YL, Craig W Holden, and Ondřej Tobek. 2018. "Are volatility over volume liquidity proxies useful for global or US research?" *Kelley School of Business Research Paper,* nos. 17-49.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2020. "Empirical asset pricing via machine learning." *The Review of Financial Studies* 33 (5): 2223–2273.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2021. "Autoencoder asset pricing models." *Journal of Econometrics* 222 (1): 429–450.

Hanauer, Matthias X, and Tobias Kalsbach. 2023. "Machine learning and the cross-section of emerging market stock returns." *Emerging Markets Review* 55:101022.

Hou, Kewei, Chen Xue, and Lu Zhang. 2020. "Replicating anomalies." *The Review of financial studies* 33 (5): 2019–2133.

Ioffe, Sergey, and Christian Szegedy. 2015. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *arXiv preprint arXiv:1502.03167.*

Jiang, Jingwen, Bryan Kelly, and Dacheng Xiu. 2023. "(Re-) Imag (in) ing price trends." *The Journal of Finance* 78 (6): 3193–3249.

Kelly, Bryan T, Seth Pruitt, and Yinan Su. 2019. "Characteristics are covariances: A unified model of risk and return." *Journal of Financial Economics* 134 (3): 501–524.

Kingma, Diederik P, and Jimmy Ba. 2014. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980.*

Nechvátalová, Lenka. 2024. “Multi-Horizon Equity Returns Predictability via Machine Learning.” *Czech Journal of Economics and Finance (Finance a úvěr)* 74 (2): 142–190.

Novy-Marx, Robert, and Mihail Velikov. 2019. “Comparing cost-mitigation techniques.” *Financial Analysts Journal* 75 (1): 85–102.

Tobek, Ondřej, and Martin Hronec. 2021. “Does it pay to follow anomalies research? machine learning approach with international evidence.” *Journal of Financial Markets* 56:100588.

Yang, Xuanling, Zhoufan Zhu, Dong Li, and Ke Zhu. 2024. “Asset pricing via the conditional quantile variational autoencoder.” *Journal of Business & Economic Statistics* 42 (2): 681–694.

IES Working Paper Series

2024

1. Nino Buliskeria, Jaromir Baxa, Tomáš Šestořád: *Uncertain Trends in Economic Policy Uncertainty*
2. Martina Lušková: *The Effect of Face Masks on Covid Transmission: A Meta-Analysis*
3. Jaromir Baxa, Tomáš Šestořád: *How Different are the Alternative Economic Policy Uncertainty Indices? The Case of European Countries.*
4. Sophie Ghvanidze, Soo K. Kang, Milan Ščasný, Jon Henrich Hanf: *Profiling Cannabis Consumption Motivation and Situations as Casual Leisure*
5. Lorena Skufi, Meri Papavangjeli, Adam Gersl: *Migration, Remittances, and Wage-Inflation Spillovers: The Case of Albania*
6. Katarina Gomoryova: *Female Leadership and Financial Performance: A Meta-Analysis*
7. Fisnik Bajrami: *Macroprudential Policies and Dollarisation: Implications for the Financial System and a Cross-Exchange Rate Regime Analysis*
8. Josef Simpart: *Military Expenditure and Economic Growth: A Meta-Analysis*
9. Anna Alberini, Milan Ščasný: *Climate Change, Large Risks, Small Risks, and the Value per Statistical Life*
10. Josef Bajzík: *Does Shareholder Activism Have a Long-Lasting Impact on Company Value? A Meta-Analysis*
11. Martin Gregor, Beatrice Michaeli: *Board Bias, Information, and Investment Efficiency*
12. Martin Gregor, Beatrice Michaeli: *Board Compensation and Investment Efficiency*
13. Lenka Šlegerová: *The Accessibility of Primary Care and Paediatric Hospitalisations for Ambulatory Care Sensitive Conditions in Czechia*
14. Kseniya Bortnikova, Tomas Havranek, Zuzana Irsova: *Beauty and Professional Success: A Meta-Analysis*
15. Fan Yang, Tomas Havranek, Zuzana Irsova, Jiri Novak: *Where Have All the Alphas Gone? A Meta-Analysis of Hedge Fund Performance*
16. Martina Lušková, Kseniya Bortnikova: *Cost-Effectiveness of Women's Vaccination Against HPV: Results for the Czech Republic*
17. Tersoo David Iorngurum: *Interest Rate Pass-Through Asymmetry: A Meta-Analytical Approach*
18. Inaki Veruete Villegas, Milan Ščasný*: Input-Output Modeling Amidst Crisis: Tracing Natural Gas Pathways in the Czech Republic During the War-Induced Energy Turmoil*
19. Theodor Petřík: *Distribution Strategy Planning: A Comprehensive Probabilistic Approach for Unpredictable Environment*
20. Meri Papavangjeli, Adam Geršl*: Monetary Policy, Macro-Financial Vulnerabilities, and Macroeconomic Outcomes*

21. Attila Sarkany, Lukáš Janásek, Jozef Baruník: *Quantile Preferences in Portfolio Choice: A Q-DRL Approach to Dynamic Diversification*
22. Jiri Kukacka, Erik Zila: *Unraveling Timing Uncertainty of Event-driven Connectedness among Oil-Based Energy Commodities*
23. Samuel Fiifi Eshun, Evžen Kočenda: *Money Talks, Green Walks: Does Financial Inclusion Promote Green Sustainability in Africa?*
24. Mathieu Petit, Karel Janda: *The Optimal Investment Size in the Electricity Sector in EU Countries*
25. Alessandro Chiari: *Do Tax Havens Affect Financial Management? The Case of U.S. Multinational Companies*
26. Lenka Nechvátalová: *Autoencoder Asset Pricing Models and Economic Restrictions – International Evidence*

All papers can be downloaded at: http://ies.fsv.cuni.cz

Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail : ies@fsv.cuni.cz http://ies.fsv.cuni.cz