Recurrent Neural Networks Meet Asset Pricing

Gerrit Liedtke[1]

Abstract

This paper proposes the conditional recurrent neural network (C-RNN), a flexible and interpretable recurrent factor model for capturing the dynamic structure of asset returns. The C-RNN conditions latent factors and factor loadings on observable variables and allows both to follow vector autoregressive dynamics. Using U.S. stock data, the C-RNN outperforms existing models in return prediction, and a simple zero-investment strategy achieves an annualized Sharpe ratio of 1.30 versus 0.82 for instrumented principal component analysis—the state-of-the-art benchmark model. These results highlight the value of combining temporal dependencies with conditioning information and show that few factors but many characteristics are needed to model returns.

Keywords: asset pricing, machine learning, recurrent neural network, factor model, cross section, time series

JEL classification: C33, G11, G12, G14, G17

1. Introduction

Machine learning has firmly arrived in empirical asset pricing, offering new tools to uncover complex patterns, exploit high-dimensional information, and model nonlinear economic relations that are difficult to capture with traditional econometric methods. Following the influential work of Gu et al. (2020), researchers have demonstrated that machine learning methods can effectively exploit the information embedded in asset characteristics to forecast returns (e.g., Bianchi et al., 2021; Bali et al., 2023; Cakici et al., 2023). Kelly et al. (2019a) introduce instrumented principal component analysis (IPCA), which imposes a characteristic-based factor structure on returns, while Gu et al. (2021) extend this framework by allowing for nonlinearities and interactions in firm characteristics. More recently, Chen et al. (2024) propose a deep neural network model that incorporates a no-arbitrage constraint. Although the list of machine learning applications in asset pricing continues

[1]Gerrit Liedtke: University of Bremen, Bremen, Germany, gliedtke@uni-bremen.de.

to grow, the role of recurrent neural networks (RNNs) remains largely underexplored, despite their ability to capture temporal dependencies that are central to understanding the dynamics of asset returns.

This paper studies RNN-type factor models in empirical asset pricing. Specifically, we propose the conditional recurrent neural network (C-RNN), a generalization of the recurrent double-conditional factor model (RDCFM) introduced by Fieberg et al. (2025a). The C-RNN models latent factors (i.e., hidden states in a classical RNN) as a function of observable conditioning variables.[2] Building on Avramov and Chordia (2006) and Kelly et al. (2019b), among others, the factor loadings are likewise conditioned on observable variables, namely, firm characteristics, allowing betas to cross-sectionally vary with observed firm characteristics. In contrast to models with static or memoryless conditioning, the C-RNN specifies that both the latent factors and the factor loadings evolve according to vector autoregressive dynamics, enabling the model to capture persistence, lagged dependencies, and richer interactions between conditioning variables. Finally, due to the recurrent structure, the model provides a more reliable estimate of the ex-ante factor risk premia and respective loadings.

The C-RNN combines several features that were found to improve pricing performance in previous studies. Conditioning the latent factors on observable variables reduces the dimensionality of a potentially large set of variables that drive the comovement in returns to a smaller number of latent factors (Ferson, 1990; Ferson and Harvey, 1993; Maio, 2013). Allowing these factors to be recurrent ensures that past information persists in the model, yielding more reliable ex-ante estimates of factor premia. "Dynamic factor models", in which the factors follow a vector autoregressive process, are common in fields such as macroeconomics (Kose et al., 2003; Del Negro and Otrok, 2008; Jungbacker et al., 2011; Bai and Wang, 2015), but do not find wide application in financial economics due to computational infeasibility in large cross sections. Introducing the conditioning of the factor loadings is key for reducing the dimensionality of the estimation problem, because the loadings are inferred from a small set of characteristic coefficients rather than being estimated directly through a large asset-by-factor output matrix. This approach also allows betas to reflect moderating effects of characteristics (e.g., small stocks with high book-to-market equity ratios have larger betas on a distress factor (Fama and French, 1993)) and to vary over time as these characteristics evolve. The

[2] This model feature is inspired by research on conditional factors (e.g., Ferson and Harvey, 1991, 1993; Maio, 2013).

outperformance of conditional models over unconditional ones is well documented—for example, Kelly et al. (2019a), Büchner and Kelly (2022), Kelly et al. (2023), Fieberg et al. (2025c), and Goyal and Saretto (2025) show that IPCA improves upon unconditional benchmarks, making it the state-of-the-art linear asset pricing model. Finally, some studies document the advantage of autoregressive loadings (Del Negro and Otrok, 2008; Bali et al., 2009, 2017) allowing betas to capture persistent movements in exposures that static or memoryless models cannot.

We apply the C-RNN to U.S. equity data comprising 18,258 stocks observed from July 1962 to December 2022. The latent factors are conditioned on fourteen Fama and French (1993)-style factor-mimicking portfolios, while the factor loadings are conditioned on 33 firm characteristics commonly used in empirical asset pricing. The empirical results indicate that a five-factor C-RNN attains an out-of-sample predictive R^2 of 0.81%, compared to 0.52% achieved by IPCA. This improvement in statistical performance translates into economically meaningful gains, as illustrated in Figure 1. A simple long–short strategy that buys the 10% of stocks with the highest expected returns and shorts the 10% with the lowest expected returns yields an average monthly return of 2.34% for the C-RNN, corresponding to an annualized Sharpe ratio of 1.30. By contrast, the RDCFM of Fieberg et al. (2025a) and IPCA produce substantially lower Sharpe ratios of 0.98 and 0.82, respectively. A double-conditional factor model (DCFM) that conditions both factors and loadings on observable variables—but restricts them to be static rather than recurrent—performs on par with IPCA. This contrast highlights the key insight of the paper: introducing recurrent dynamics into both factors and factor loadings improves the model's ability to estimate ex-ante risk premia and, in turn, translates to stronger economic performance.

[Insert Figure 1 about here]

In further analyses, we find that the performance of the C-RNN is not driven by small or illiquid stocks, nor does it stem primarily from periods in which limits-to-arbitrage are high. The model yields lower pricing errors for 165 double-sorted zero-investment portfolios and leaves fewer returns unexplained than either an instrumented version of the capital asset pricing model (CAPM) or IPCA. Finally, we shed light on the importance of the conditioning variables by examining how macroeconomic and firm-level characteristics shape both the latent factors and the cross-sectional variation in factor loadings.

This paper makes three contributions to the asset pricing and machine learning literature. First, it adds to the growing body of research that applies machine learning to asset pricing. Recently, machine learning methods have been used to tackle the high dimensionality of the factor zoo and to uncover complex nonlinear patterns in firm characteristics (Gu et al., 2020; Freyberger et al., 2020; Bianchi et al., 2021; Jensen et al., 2022; Bali et al., 2023; Cakici et al., 2023; Jiang et al., 2023; Fieberg et al., 2023, 2025b; Bryzgalova et al., 2025). Another strand of work embeds economic structure into machine learning algorithms by imposing a latent factor representation (Kelly et al., 2019a; Gu et al., 2021; Büchner and Kelly, 2022; Kelly et al., 2023; Chen et al., 2024; Fieberg et al., 2025c; Goyal and Saretto, 2025). The C-RNN builds on this conditional factor modeling framework by incorporating recurrent dynamics into the latent factors. While the general architecture was introduced in Fieberg et al. (2025a), the authors did not conduct comprehensive asset pricing tests. We fill this gap and further extend the framework by additionally allowing for recurrent dynamics in the factor loadings, enabling both the factors and their exposures to evolve over time. This richer specification captures persistent temporal dependence that static or feed-forward conditional models cannot represent.

Second, the paper extends sequence-learning methods to make them applicable to large cross sections. In financial economics, recurrent architectures have been employed to predict returns for small subsets of stocks (Fischer and Krauss, 2018; Murray et al., 2024) or to extract latent macroeconomic state variables from time-series data (Chen et al., 2024). However, applying RNNs to large cross sections is generally considered infeasible. The challenge arises from the high dimensionality of the output layer: a standard RNN would require mapping the hidden state to thousands of individual stock returns, resulting in an output matrix with ten thousands of parameters that cannot be estimated reliably in typical asset-pricing settings. Fieberg et al. (2025a) address this scalability problem by conditioning the output matrix—namely, the factor loadings—on observable firm characteristics. This approach replaces the infeasible hidden-to-returns mapping with a parsimonious characteristic-based representation, dramatically reducing the number of parameters while incorporating more information. The C-RNN extends this insight by allowing the factor loadings to evolve according to recurrent dynamics, while remaining conditional on current characteristic. This architecture combines the tractability of conditional factor models with the time-series learning capabilities of RNNs, making sequence learning feasible in large-scale cross-sectional settings.

Third, the C-RNN contributes to the literature on interpretable AI in finance by preserving the economic transparency of conditional factor models while allowing for flexible dynamics. The model's parametrization enables a clear decomposition of how factor-mimicking portfolios shape latent factors and how firm characteristics determine risk exposures, offering white-box interpretability that is typically absent in machine learning methods. Our out-of-sample importance analysis shows that only a small subset of factor-mimicking portfolios, most prominently the market and size factors, accounts for the majority of return comovement. In contrast, the cross-sectional structure of factor loadings is inherently high-dimensional: many firm characteristics, including sales-to-assets, operating leverage, size, beta, assets-to-market, idiosyncratic volatility, reversal, etc., contribute incrementally to cross-sectional variation. This analysis thus puts a new perspective on the discussion of how many anomalies are relevant (see, for example, Green et al., 2017; Kozak et al., 2020; Bryzgalova et al., 2023; Swade et al., 2024; Didisheim et al., 2025, for perspectives on that topic): these results indicate that, while the dynamics of latent factors are driven by a few key sources, capturing the heterogeneity in factor exposures requires conditioning on a broad set of firm-level characteristics. In other words, we need relatively few factors but many characteristics to adequately model the respective loadings.

The remainder of the paper is organized as follows. Section 2 introduces the C-RNN framework and briefly discusses its estimation. Section 3 describes the empirical setup, including the dataset, benchmark models, and performance evaluation metrics. Section 4 presents the empirical results. Section 5 concludes.

2. Method

2.1. The Conditional Recurrent Neural Network

This section introduces the C-RNN, a generalization of the RDCFM proposed by Fieberg et al. (2025a). To revise their model and show how the C-RNN embeds and extends their model, we begin with the standard multifactor asset pricing framework and then show how the C-RNN extends the RDCFM by allowing for dynamic factor loadings.

Let $r_{i,t}$ denote the excess return of asset $i = 1, \ldots, N$ in month $t = 1, \ldots, T$. Following theories of multifactor asset pricing (Merton, 1973; Ross, 1976; Cochrane, 2001), asset returns are assumed to follow a low-dimensional factor structure:

$$r_{i,t} = f_{1,t}\beta_{i,1,t} + \cdots + f_{K,t}\beta_{i,K,t} + e_{i,t}, \tag{1}$$

where $f_{j,t}$ denotes the realization of factor j at time t, $\beta_{i,j,t}$ is the contemporaneous factor loading of asset i to the j-th factor at t, and $e_{i,t}$ represents the idiosyncratic return that is not captured by the factor structure for which it is assumed that it is uncorrelated with the factors, i.e., $\text{COV}\left(e_{i,t}, f_{j,t}\right) = 0 \;\; \forall i, j, t$. In matrix notation, equation (1) can be rewritten as follows:

$$\boldsymbol{r}_t = \boldsymbol{f}_t \boldsymbol{B}_t^{'} + \boldsymbol{e}_t \tag{2}$$

with $\boldsymbol{r}_t$ denoting the $1 \times N$ vector of asset excess returns, $\boldsymbol{f}_t$ is a $1 \times K$ vector of risk factor realizations, and $\boldsymbol{B}_t$ denotes the $N \times K$ matrix of factor loadings at time t. $\boldsymbol{e}_t$ is a $1 \times N$ vector of residual returns.

Based on research on conditional factor models (Ferson, 1990; Ferson and Harvey, 1993; Maio, 2013; Fieberg et al., 2025a) we estimate the factors from exogenous instruments. Let $\boldsymbol{z}_t$ denote a $1 \times L$ vector of stationary variables with only a time-series dimension that capture information about the latent factors $\boldsymbol{f}_t$. In the empirical analysis in Section 4, $\boldsymbol{z}_t$ corresponds to a vector of factor-mimicking portfolios at time t, although in principle $\boldsymbol{z}_t$ may include any macroeconomic variable or lag polynomial thereof. Rather than treating each element of $\boldsymbol{z}_t$ as an individual factor in equation (1), a smaller set of conditional factors is extracted to summarize the relevant common time-series variation in returns. These conditional factors are constructed as linear combinations of the instruments $\boldsymbol{z}_t$, as shown in equation (3):

$$\boldsymbol{f}_t = \boldsymbol{z}_t \boldsymbol{\Omega}, \tag{3}$$

where $\boldsymbol{f}_t$ is the $1 \times K$ vector of conditional factors at time t, and $\boldsymbol{\Omega}$ is an $L \times K$ matrix of time-invariant coefficients mapping the factor instruments to the latent factors.

Next, the static conditional factor specification in equation (3) is combined with recurrent dynamics to obtain conditional factors that depend on their own past. The resulting formulation is inspired by RNNs (Elman, 1990; Rumelhart et al., 1986) and given by

$$\boldsymbol{f}_t = \tanh(\boldsymbol{z}_t \boldsymbol{\Omega} + \boldsymbol{f}_{t-1} \boldsymbol{AR}_f), \tag{4}$$

where $\boldsymbol{z}_t \boldsymbol{\Omega}$ captures the contemporaneous contribution of the conditioning instruments and $\boldsymbol{f}_{t-1} \boldsymbol{AR}_f$ introduces a recurrent autoregressive component through a $K \times K$ coefficient matrix $\boldsymbol{AR}_f$. The tanh(·) activation stabilizes the dynamics and bounds the hidden state, yielding a specification equivalent to the hidden-state update in a classical Elman-type RNN.[3]

A large body of evidence shows that characteristic-based loadings improve pricing performance (Avramov and Chordia, 2006; Kelly et al., 2019a, 2023; Fieberg et al., 2025a,c; Goyal and Saretto, 2025), therefore, instead of modeling unconditional betas, the C-RNN assumes that factor loadings are conditional on observable asset characteristics. Let $\boldsymbol{C}_t$ denote an $N \times M$ matrix of firm characteristics. Inspired by Kelly et al. (2019a,b), among others, the conditional factor loadings are constructed as linear combinations of these characteristics through an $M \times K$ coefficient matrix $\boldsymbol{\Gamma}_\beta$. The conditional loading matrix is thus given by

$$\boldsymbol{B}_t = \boldsymbol{C}_t \boldsymbol{\Gamma}_\beta. \tag{5}$$

Conditioning loadings on characteristics substantially reduces the dimensionality of the problem when $M < N$. Furthermore, if characteristics change over time, the factor loadings also change; thus, conditional betas capture cross-sectional information through the characteristic rank and time-series information through the characteristic's variation over time.

Taken together, equations (4) and (5) yield the RDCFM as proposed in Fieberg et al. (2025a). The authors show that allowing the factors to "remember" past information—through recurrent dynamics—substantially reduces prediction errors for future stock returns, underscoring the importance of temporal dependencies in modeling the latent structure of returns. Motivated by this insight, it is natural to extend the same logic to factor loadings. If exposures reflect evolving firm fundamentals, changing risk characteristics, and shifting economic conditions, then modeling betas as static or memoryless functions of characteristics may discard relevant information. Introducing recurrent dynamics into the loading structure allows the C-RNN to capture persistence and lagged effects in exposures. Dynamic factor loadings are specified as follows:

$$\boldsymbol{B}_t = \boldsymbol{C}_t \boldsymbol{\Gamma}_\beta + \boldsymbol{B}_{t-1} \boldsymbol{A}\boldsymbol{R}_\beta, \tag{6}$$

[3]Fieberg et al. (2025a) apply the nonlinear transformation only to the past factor realization to remain closer to linear conditional factor models. In contrast, and in line with the machine learning literature, we adopt the formulation in equation 4, which reflects the standard construction of hidden states in recurrent neural networks. This makes the entire factor construction process nonlinear, however, in unreported results, we find that empirical results are qualitatively unchanged.

where $\boldsymbol{C}_t\boldsymbol{\Gamma}_\beta$ captures the contemporaneous effect of firm characteristics and $\boldsymbol{B}_{t-1}\boldsymbol{AR}_\beta$ introduces an autoregressive component through a $K \times K$ coefficient matrix $\boldsymbol{AR}_\beta$. Unlike the factor dynamics in (4), no nonlinear activation function is applied here, so that the resulting factor pricing relation remains linear in factors and loadings.

After deriving the construction of both the dynamic conditional factors and the conditional loadings, the C-RNN can be written as an extended version of equation (2):

$$\boldsymbol{r}_t = \underbrace{\Big(\tanh(\boldsymbol{z}_t\boldsymbol{\Omega} + \boldsymbol{f}_{t-1}\boldsymbol{AR}_f)\Big)}_{\boldsymbol{f}_t}\underbrace{\Big(\boldsymbol{C}_t\boldsymbol{\Gamma}_\beta + \boldsymbol{B}_{t-1}\boldsymbol{AR}_\beta\Big)'}_{\boldsymbol{B}_t'} + \boldsymbol{e}_t \tag{7}$$

Figure 2 illustrates how the C-RNN relates to both a standard RNN and the RDCFM. A standard RNN also produces "conditional factors" in the sense that the hidden state depends on observable "factor instruments" fed into the network and its own lagged hidden state, without incorporating firm-level characteristics. Moreover, the hidden–output matrix in a standard RNN is highly parameterized, making the architecture infeasible for large cross-sections. The RDCFM overcomes this limitation and extends this idea by introducing a characteristic layer that conditions the factor loadings on observable variables. The C-RNN further generalizes this framework by additionally allowing the factor loadings to depend on their own lagged realizations.

[Insert Figure 2 about here.]

The C-RNN is a flexible framework that nests several previously studied models as special cases, obtained by selectively enabling or disabling specific features. For instance, setting $\boldsymbol{AR}_\beta = 0$ reduces the specification to the RDCFM of Fieberg et al. (2025a). Further imposing $\boldsymbol{AR}_f = 0$ yields a double-conditional factor model (DCFM).[4] Although the C-RNN nests more special cases, we restrict attention to these three models in our empirical application because this allows examining whether recurrent dynamics in factors and/or loadings improve pricing performance.[5]

[4]The idea of the DCFM is closely related to the autoencoder asset pricing model with no hidden layer studied in Gu et al. (2021). Originally derived to extract latent factors from the matrix of individual stock returns, Gu et al. (2021) propose to introduce a preprocessing layer that first creates characteristic-managed portfolios and then extracts factors based on the return matrix of portfolio returns. Thus, using the characteristic-managed portfolios as factor instruments, the DCFM is equivalent to the autoencoder asset pricing model that involves this data preprocessing step.

[5]Alternatively, the factor loadings may be estimated unconditionally, for example, via OLS to obtain the recurrent

2.2. Parameter Estimation

Due to the connection of conditional factor models with RNNs, the C-RNN itself can be seen as a neural network with economic restrictions to protect against overfitting (Israel et al., 2020; Nagel, 2021). Instead of estimating a highly parameterized matrix of factor loadings, which is of dimension $N \times K$, the C-RNN takes advantage of the findings in the financial economics literature that asset betas vary with their characteristics (Avramov and Chordia, 2006; Kelly et al., 2019a; Büchner and Kelly, 2022; Kelly et al., 2023). However, unlike other neural networks applied to asset pricing (e.g., Gu et al., 2020, 2021; Chen et al., 2024), the C-RNN is sparsely parameterized, increasing the interpretability and model stability. Therefore, there is no need for regularization of the network parameters, which increases the speed of model estimation.

Because the factor loadings follow a recurrent process in the C-RNN, the estimation approach of Fieberg et al. (2025a)—which relied on pooled ordinary least squares (POLS) to estimate $\mathbf{\Gamma}_\beta$—is no longer applicable. Consequently, all model parameters $\{\mathbf{\Omega}, \boldsymbol{AR}_f, \mathbf{\Gamma}_\beta, \boldsymbol{AR}_\beta, \boldsymbol{f}_0\}$ must be estimated simultaneously using nonlinear least squares.[6]

The parameter optimization is based on the modified backpropagation through time algorithm, described in Fieberg et al. (2025a). In Fieberg et al. (2025a), the authors use the well-known BFGS algorithm for parameter optimization. However, in unreported simulations, it is found that the resilient propagation (Rprob) (Riedmiller and Braun, 1992, 1993) algorithm produces results that are as good as—or even better than—those of BFGS. However, the Rprob algorithm has two advantages over BFGS. First, it is a first-order method and thus does not require the approximation of the Hessian matrix, reducing the computational costs. Second, as shown in Hochreiter and Schmidhuber (1997), RNNs may suffer from the exploding and vanishing gradient problem when trained over a long sequence. Rprob overcomes this issue by not adjusting the model parameters

conditional factor model (RCFM) or an RNN-like model, respectively. Further restricting the coefficients of the VAR process to be zero in an RCFM yields the conditional factor model (Ferson and Harvey, 1991, 1993), which is closely related—but not equivalent—to partial least squares (PLS). The key difference is that the CFM extracts latent factors that best summarize variation in the original returns, whereas PLS first maps returns into a transformed space and extracts factors that best explain variation in that transformed space.

[6]To ensure stable and reproducible optimization, the coefficients in $\mathbf{\Omega}$ are initialized using PLS applied to the characteristic-managed portfolios, which are defined later in equation (11). This provides economically meaningful starting values for the factor–instrument relationship. This initialization was found to considerably accelerate convergence relative to random initialization. The autoregressive matrices $\boldsymbol{AR}_f$ and $\boldsymbol{AR}_\beta$, as well as the initial states $\boldsymbol{f}_0$ are initialized at zero. The initial estimate for $\mathbf{\Gamma}_\beta$ is obtained by a POLS regression of asset returns on their characteristics interacted with factor realizations, which are obtained using the start values for the factor parameters. To keep the optimization problem feasible, the initial betas are not optimized and simply set to zero.

according to the magnitude of the gradient, but only in its direction. Furthermore, Rprob assumes a minimum and maximum step size, which ensures that training does not stop. An additional advantage of Rprob compared to other first-order methods established in the machine learning literature is that it does not require the specification of a learning rate, which might be crucial for the optimization success. Unreported simulations also show that the hyperparameters of the Rprob optimization algorithm do not substantially affect its performance. Therefore, all hyperparameters are kept at the default settings as suggested in Riedmiller and Braun (1992, 1993).

In line with previous asset pricing models, factors are sorted from highest to lowest variance and estimated coefficients are re-scaled such that the factor mean is positive (e.g., Kelly et al., 2019a; Lettau and Pelger, 2020).[7]

3. Empirical Setup

3.1. Data

In this study, we examine the cross-section of U.S. stock returns over the period July 1962 to December 2022. Stock returns and market capitalizations are obtained from the Center for Research in Security Prices (CRSP), while accounting-based information is sourced from Compustat.

The instruments used to condition the factors are fourteen popular empirical asset pricing factors widely documented to capture common variation in stock returns. These include the six factors from the Fama–French six-factor model obtained from Kenneth French's data library: the market (MKT) factor, the size-based small-minus-big (SMB) factor, the book-to-market-based high-minus-low (HML) factor, the profitability-based robust-minus-weak (RMW) factor, the investment-based

[7]Specifically, to ensure positive factor means, the $\hat{\Omega}$ coefficients are re-scaled according to:

$$\hat{\boldsymbol{\Omega}}^{+} = \hat{\boldsymbol{\Omega}} \odot \operatorname{sign}\left(\bar{\boldsymbol{f}}\right)$$

with $\hat{\boldsymbol{\Omega}}$ denoting the estimate of $\boldsymbol{\Omega}$ and $\bar{\boldsymbol{f}}$ is the $1 \times K$ vector of the means of the estimated factors. Element-wise multiplication is indicated by $\odot$. For both $\hat{\boldsymbol{AR}}_f$ and $\hat{\boldsymbol{AR}}_\beta$, the coefficients are rescaled to positive mean according to:

$$\hat{\boldsymbol{AR}}^{+} = \hat{\boldsymbol{AR}} \odot \left(\operatorname{sign}\left(\bar{\boldsymbol{f}}\right) \odot \operatorname{sign}\left(\bar{\boldsymbol{f}}\right)'\right)$$

and for the initialization of the factor process:

$$\hat{\boldsymbol{f}}_0^{+} = \hat{\boldsymbol{f}}_0 \odot \operatorname{sign}\left(\bar{\boldsymbol{f}}\right)$$

Likewise, the estimated $\boldsymbol{\Gamma}_\beta$ coefficients are adjusted according to:

$$\hat{\boldsymbol{\Gamma}}_\beta^{+} = \hat{\boldsymbol{\Gamma}}_\beta \odot \operatorname{sign}\left(\bar{\boldsymbol{f}}\right)$$

conservative-minus-aggressive (*CMA*) factor, and the momentum-based winner-minus-loser (*WML*) factor (Fama and French, 2015, 2018; Carhart, 1997). This set is extended by a short-term reversal (*STREV*) factor, a long-term reversal (*LTREV*) factor, and the expected investment growth (*EG*) factor of Hou et al. (2021).[8] In addition, we include the two behavioral factors of Daniel et al. (2020)—post-earnings-announcement drift (*PEAD*) and financing (*FIN*)—and add the liquidity-based (*LIQ*) factor of Pástor and Stambaugh (2003).[9] Finally, the factor set is completed with the betting-against-beta (*BAB*) factor (Frazzini and Pedersen, 2014) and the quality-minus-junk (*QMJ*) factor (Asness et al., 2019), both obtained from the AQR data library. Missing observations are replaced with zeros.

To ensure comparability with Kelly et al. (2019b), the factor loadings are conditioned on a constant characteristic and 33 market- and accounting-based firm characteristics that are fully replicable using the datasets of Chen and Zimmermann (2022) and Jensen et al. (2023). Specifically, the characteristics include market beta (*beta*); bid-ask spread (*bidask*); assets-to-market (*a2me*); total assets (*assets*); sales-to-assets (*ato*); book-to-market (*bm*); cash-to-short-term-investment (*c*); capital turnover (*cto*); capital intensity (*d2a*); the ratio of the change in property, plant, and equipment to the change in total assets (*dpi2a*); earnings-to-price (*e2p*); cash-flow-to-book (*freecf*); idiosyncratic volatility with respect to the Fama–French three-factor model (*idiovol*); investment (*investment*); leverage (*lev*); market capitalization (*mktcap*); turnover (*turn*); net operating assets (*noa*); operating accruals (*oa*); operating leverage (*ol*); price-to-cost margin (*pcm*); profit margin (*pm*); gross profitability (*prof*); the ratio of price to its 52-week high (*w52h*); return on net operating assets (*rna*); return on assets (*roa*); return on equity (*roe*); 12-month momentum (*mom*); intermediate momentum (*intmom*); short-term reversal (*strev*); long-term reversal (*ltrev*); sales-to-price (*s2p*); and the ratio of selling, general, and administrative expenses to sales (*sga2s*). Each characteristic is lagged by one month and cross-sectionally standardized using a rank transformation mapped into the interval $[-0.5, , 0.5]$. This transformation ensures stationarity over time and mitigates the impact of outliers.

For a stock to be included in the dataset, we require a small set of core characteristics—*mktcap*, *bm*, *mom*, *investment*, and *prof*—to be nonmissing. Any remaining missing characteristic values

[8] We do not include their return-on-equity (*ROE*) profitability factor or investment-based (*IA*) factor to avoid multicollinearity concerns. The EG factor is obtained from https://global-q.org/factors.html.

[9] The PEAD and FIN factors are available from Kent Daniel's website and the LIQ factor is available from Robert Stambaugh's website.

are replaced with the post-standardized cross-sectional median, i.e., 0 (Gu et al., 2020, 2021). The final dataset thus covers 18,258 individuals stocks over a period of sixty years.

3.2. Benchmark Models

As discussed in Section 2, the C-RNN is a generalization of the RDCFM proposed in Fieberg et al. (2025a). Fieberg et al. (2025a) show that the RDCFM delivers strong performance in predicting stock returns, but the model has not yet been rigorously evaluated in a full asset pricing context, particularly with respect to pricing errors, portfolio implications, or its ability to explain cross-sectional anomalies. To provide such an assessment, and to understand how the additional flexibility introduced by the C-RNN affects empirical performance, we estimate the RDCFM and its static counterpart, the DCFM, within the same empirical framework.

Conceptually, the three models differ in the dynamics assigned to factors and loadings: (i) the C-RNN features instrumented and dynamic factors together with dynamic, characteristic-based loadings, (ii) the RDCFM combines instrumented and dynamic factors with static, characteristic-based loadings, and (iii) the DCFM assumes both instrumented factors and characteristic-based loadings but imposes static dynamics on both components. Evaluating all three models side by side allows us to isolate the empirical relevance of dynamic factors and factor loadings.

We compare the performance with various linear pricing models from the literature, beginning with the standard empirical factor models. Specifically, we include the CAPM, the Fama–French three-factor model (FF3) (Fama and French, 1993), the Fama–French–Carhart four-factor model (FFC4) (Carhart, 1997), the Fama–French five-factor model (FF5) (Fama and French, 2015), and the Fama–French–Carhart six-factor model (FFC6) (Fama and French, 2020). For these models, we estimate unconditional variants with static OLS betas as well as instrumented betas using the same set of characteristics described above. We further include IPCA (Kelly et al., 2019b), which imposes a linear factor structure with characteristic-based loadings. These benchmark models provide a natural reference point against which to assess the empirical value added by conditioning on instruments, introducing recurrent dynamics, and allowing factor loadings to evolve over time.[10]

[10]Note that only linear benchmark asset pricing models are considered in this study to allow for a fair comparison. Gu et al. (2021) show that modeling conditional betas non-linearly improves the performance of the models; however, at the cost of interpretability. However, the C-RNN may also be extended to account for such nonlinearities in future research.

3.3. Model Estimation and Evaluation Metrics

All models are estimated and evaluated on a strictly out-of-sample basis using an expanding-window scheme. Beginning in month $t > 120$, the model parameters are estimated using all information available up to month t. Based on these parameters, returns for month $t+1$ are estimated. To reduce computational complexity, the estimated parameters are held fixed for the subsequent year before the next update of the expanding window takes place.

The statistical performance evaluation of the models throughout this paper follows Kelly et al. (2019a). Specifically, two measures of R^2s quantify the fraction of return variation explained by the factor structure relative to a naive zero-return benchmark. The first measure is the total R^2, which captures the overall fraction of return variation explained by contemporaneous factor realizations. It is defined as

$$\text{Total } R^2 = 1 - \frac{\sum_t^T \sum_i^{N_t} \left(r_{i,t} - \left(\hat{\boldsymbol{f}}_t \hat{\boldsymbol{\beta}}_{i,t}\right)\right)^2}{\sum_t^T \sum_i^{N_t} r_{i,t}^2}, \tag{8}$$

where N_t denotes the number of assets in month t, and $\boldsymbol{\beta}_{i,t}$ is the estimated factor loading of asset i in month t.

As the total R^2 measures how well an asset pricing model explains the contemporaneous variation in returns—using information that is only known ex post—it does not capture the model's usefulness for ex-ante decision making. To evaluate a model's ability to describe expected returns or risk premia, a second measure is used: the predictive R^2, which measures how much of the variation in returns is explained solely by the model's estimate of factor premia. It is obtained by replacing the realized factor value $\hat{\boldsymbol{f}}_t$ with the model's estimate of factor premia $\hat{\boldsymbol{\lambda}}_t$ based on information up to t. Thus, the predictive R^2 is defined as:

$$\text{Pred } R^2 = 1 - \frac{\sum_t^T \sum_i^{N_t} \left(r_{i,t} - \left(\hat{\boldsymbol{\lambda}}_t \hat{\boldsymbol{\beta}}_{i,t}\right)\right)^2}{\sum_t^T \sum_i^{N_t} r_{i,t}^2}, \tag{9}$$

For the benchmark models—namely, IPCA and the empirical factor models—it is standard to use the historical mean of each factor as the estimate of its corresponding risk premium (e.g., Kelly et al., 2019a). For models with conditional factors, this requires an estimate of the conditional factor premia. Because the factor instruments are not lagged—consistent with the common practice that time t factor-mimicking portfolios capture contemporaneous variation in returns—their in-sample means are used to obtain the conditional factor estimate, as described in as described in equation (10):

$$\hat{\boldsymbol{\lambda}}_t = \tanh\left(\bar{\boldsymbol{z}}\hat{\boldsymbol{\Omega}} + \hat{\boldsymbol{f}}_{t-1}\boldsymbol{A}\hat{\boldsymbol{R}}_f\right) \tag{10}$$

with $\bar{\boldsymbol{z}}$ denoting the $1 \times L$ vector of the in-sample means of the factor instruments.[11]

4. Empirical Results

4.1. Statistical Performance and Predictive Accuracy

Table 1 reports the out-of-sample R^2s for individual stock returns over the period from July 1972 to December 2022. All models are estimated with $K = 1, \ldots, 6$ factors. Before turning to the recurrent models, it is useful to revisit a well-established empirical finding in the literature: models with conditional factor loadings systematically outperform their unconditional counterparts (Kelly et al., 2019a; Büchner and Kelly, 2022; Kelly et al., 2023; Fieberg et al., 2025a,c). Panels E and F clearly confirm this pattern. Unconditional empirical factor models tend to overfit with respect to the factor loadings, resulting in substantially lower total and predictive R^2s compared to their instrumented versions. In our results, unconditional models explain only between 3.84% and 9.77% of contemporaneous return variation, and at most 0.20% when ex-ante factor premia are used. In contrast, instrumented empirical factor models capture up to 14.8% of total variation, with predictive R^2s that are roughly twice as large as those of the unconditional specifications.

Turning to the recurrent and double-conditional models in Panels A–C, several patterns emerge that highlight the empirical relevance of introducing dynamic factors and dynamic loadings. First, the C-RNN in Panel A achieves the highest overall performance among the three related specifications. The total R^2 increases monotonically with the number of factors, reaching 15.4% for $K = 6$ while the predictive R^2 peaks at 0.81% for $K = 5$. These values are consistently above those of the static DCFM and generally at least as large as, or slightly higher than, those of the RDCFM. This suggests that allowing both the latent factors and the factor loadings to follow recurrent dynamics produces measurable gains in explaining and predicting the cross section of returns.

The RDCFM in Panel B performs similarly strongly. Total R^2 values closely track those of the C-RNN across all values of K and predictive values differ only modestly—for example, 0.75%

[11] Since the factor instruments are empirical factors, the conditioning information for the factor premia is equivalent to the historical premia of the factor-mimicking portfolios, adjusted by previous factor realizations if the factors follow a recurrent dynamic.

versus 0.81% at $K = 5$. By contrast, the static DCFM in Panel C exhibits noticeably weaker performance. Total R^2 plateaus at approximately 14.7% for five factors and the predictive R^2s are also significantly lower, ranging between 0.31% and 0.56%, well below those of both the C-RNN and RDCFM. This highlights the importance of dynamic factors—and, in the case of the C-RNN, dynamic loadings—for capturing the time-series behavior of expected returns.

Taken together, Panels A–C show that while the DCFM provides a useful baseline outperforming empirical factor models in Panel E and F, the introduction of recurrent dynamics substantially improves performance, confirming the results in Fieberg et al. (2025a). The RDCFM already delivers strong results, and the C-RNN extends these gains by allowing both factors and loadings to evolve through a recurrent structure.

Panel D reports the performance of IPCA, which generally achieves the highest total R^2s among all benchmark models. Consistent with the findings in Gu et al. (2021)—who document that IPCA captures most contemporaneous return variation even when compared with nonlinear autoencoder-based asset pricing models—IPCA again performs remarkably well in our setting. Specifically, the total R^2 increases steadily from 13.8% at $K = 1$ to 17.8% with six factors, exceeding the values obtained by both linear empirical factor models and the conditional models in Panels A–C.

However, the predictive performance of IPCA is substantially weaker. Predictive R^2s remain flat across the number of factors, ranging only from 0.29% to 0.52%. In fact, its predictive performance closely matches that of the DCFM, which shows slightly higher values across all specifications. This similarity reflects that both models rely on static factor structures and therefore extract comparable information about expected returns.

Once recurrent dynamics are introduced, however, predictive accuracy improves noticeably. The RDCFM, which augments the DCFM with recurrently updated factors, achieves predictive R^2s between 0.72% and 0.78%. Allowing both factors and loadings to evolve through recurrent updates—as in the C-RNN—further enhances predictive performance, with predictive R^2 reaching up to 0.81% with five factors. These results indicate that while IPCA provides a strong linear benchmark for contemporaneous fit, incorporating recurrent dynamics into conditional factor models is key for obtaining reliable estimates of ex-ante risk premia resulting to higher predictability for stock returns.

[Insert Table 1 about here]

Table 2 reports the out-of-sample R^2 values for characteristic-managed portfolios (CMPs) defined as:

$$\boldsymbol{x}_t = \frac{\boldsymbol{r}_t \boldsymbol{C}_{t-1}}{N_t} \tag{11}$$

with $\boldsymbol{x}_t$ being the portfolio return in t, $\boldsymbol{r}_t$ is the vector of asset returns, $\boldsymbol{C}_{t-1}$ is the matrix of centered characteristic ranks in $t-1$, and N_t is the number of available assets in t. Because the characteristic ranks are centered, the resulting portfolio is a zero-investment strategy whose weights place larger magnitude exposures on stocks with more extreme characteristic values (Kelly et al., 2019a; Kozak et al., 2020; Gu et al., 2021).

The qualitative patterns closely mirror those observed for individual stocks. IPCA continues to excel in explaining contemporaneous return variation, achieving total R^2s above 98% for six factors. This strong fit is expected because IPCA is tied to CMPs and thus designed to perfectly fit these test assets. In terms of predictive performance, however, IPCA performs similarly to the DCFM. Predictive R^2s for IPCA range from 1.32% to 1.57%, while the DCFM yields values between 1.47% and 1.70%.

The gains from introducing recurrent dynamics are again substantial. The RDCFM achieves predictive R^2s between 2.74% and 3.11%, almost doubling the predictive accuracy of both IPCA and the DCFM. The C-RNN performs similar. These improvements underscore the value of allowing factors to incorporate past information. Recurrent dynamics enable the models to better capture time-variation in expected CMP returns, leading to markedly higher ex-ante explanatory power compared to benchmarks.

[Insert Table 2 about here]

4.2. Economic Performance and Portfolio Implications

The preceding results indicate that incorporating recurrent dynamics into a factor structure materially enhances a model's ability to capture variation in expected returns. While statistical measures quantify improvements in explanatory and predictive power, it is equally informative to examine whether these gains translate into meaningful differences in return-sorted portfolio performance.

To this end, we form expected-return–sorted portfolios using each model's one-month-ahead forecasts. At the end of each month, all stocks are sorted into deciles based on their predicted returns. Stocks are ranked in ascending order and assigned to ten value-weighted portfolios, each containing 10% of the cross section. The "low" portfolio comprises stocks with the lowest predicted returns, while the "high" portfolio contains those with the highest predicted returns. Portfolios are rebalanced monthly, and a zero-investment portfolio (H–L) taking a long position in the high portfolio and a short position in the low portfolio is created.

Table 3 shows the performance of the expected-return–sorted portfolios. Panel A reports the results for the five-factor C-RNN. The portfolio returns increase monotonically across the deciles, with the average monthly return rising from -0.31% in the lowest decile to 2.03% in the highest decile. This strong monotonicity indicates that the model's predicted returns contain substantial cross-sectional information for realized returns. The high-minus-low (H-L) portfolio earns an average monthly return of 2.34%, corresponding to an annualized Sharpe ratio of 1.30. Moreover, the six-factor model alpha is 2.16% per month with a t-statistic of 8.11, demonstrating that the return spread cannot be explained by standard factors.

Panel B compares the H–L performance across models.[12,13] Although these benchmark models generate economically meaningful spreads, their risk-adjusted performance is consistently lower than that of the C-RNN: The RDCFM generates an average monthly return of 1.92% with an annualized Sharpe ratio of 0.98, thus also outperforming both the DCFM and IPCA which achieve average returns of only 1.21% and IPCA at 1.24%, respectively. Compared to the RDCFM, the C-RNN delivers only marginal improvements in statistical performance, as shown in Table 1, yet it achieves noticeably higher average returns and risk-adjusted performance, underscoring the economic value of additionally allowing the loadings to follow recurrent structures. Empirical factor models produce considerably smaller spreads and, in the case of the unconditional model, statistically insignificant performance.

[Insert Table 3 about here]

Figure 3 plots the cumulative returns of the H-L portfolios for all models over the full out-of-sample period, providing a visual comparison of the performance differentials documented in

[12] Note that we report the results for the models that yield the hightest spreads, namely, the five-factor RDCFM and DCFM, the four-factor IPCA, instrumented Fama and French (2015) five-factor model and the unconditional Fama and French (1993) three-factor model.

[13] Test results for assessing whether one model statistically significantly achieves higher average returns are shown in Table A.1 in the Online Appendix.

Table 3. The figure highlights the extent to which the C-RNN and RDCFM strategies dominate other benchmarks over time.

[Insert Figure 3 about here]

Since factor loadings in conditional factor models may be a function of fast changing firm characteristics (e.g., short-term reversal, momentum), high portfolio turnover may overstate the performance of spread strategies. Therefore, to analyze the practical implementability of those strategies, the last column of Table 3 reports the breakeven transaction cost (BETC) rate at which the strategy becomes unprofitable. Specifically, a strategy's net return is defined as:

$$r_{p,t}^{net} = \sum_{i=1}^{N} w_{i,t} r_{i,t} - c \sum_{i=1}^{N} |w_{i,t} - w_{i,t-1}| \tag{12}$$

where c denotes the transaction cost rate per trade. The BETC rate c^* is obtained by setting equation (12) to zero and solving for c, i.e.,:

$$c^* = \frac{\sum_{t=1}^{T} \sum_{i=1}^{N_t} w_{i,t} r_{i,t}}{\sum_{t=1}^{T} \sum_{i=1}^{N_t} |w_{i,t} - w_{i,t-1}|} \tag{13}$$

The BETC estimate of 0.72% for the C-RNN indicates that the return spread remains economically meaningful even under moderate trading frictions. By comparison, the IPCA strategy exhibits a lower BETC of 0.39%. Importantly, applying the IPCA breakeven cost rate of 0.39% to the C-RNN portfolio does not eliminate its profitability. When subtracting this cost rate from the monthly gross H–L returns of the C-RNN, the strategy still yields an average net return of 1.08% per month with a t-statistic of 4.20. Thus, even under the more stringent cost conditions implied by IPCA, the C-RNN strategy continues to deliver statistically and economically significant returns, underscoring the robustness of the economic value embedded in its forecasts.

Overall, the results in Table 3 and Figure 3 show that models incorporating recurrent dynamics, and especially the C-RNN, produce the strongest economic signals in the cross section of expected returns.

4.3. Explaining Asset Pricing Anomalies

The preceding analyses demonstrate that recurrent models provide substantial improvements in both statistical and economic performance. However, expected return-sorted portfolios are not designed to evaluate how well a model prices known cross-sectional anomalies. CMPs also provide limited power for this purpose because IPCA is effectively tied to CMPs by design, so its strong performance in explaining their returns is not surprising.

To conduct a more stringent evaluation, we construct a new set of test assets that are not mechanically linked to the factor-instrument design of any method. Specifically, we follow a long tradition in the empirical asset pricing literature and form 160 zero-investment portfolios by independently double-sorting stocks into six portfolios based on size (median split) and each of the 32 characteristics (tertile split).[14] For each characteristic, we collect the resulting portfolios serve as a broad and heterogeneous collection of test assets designed to assess whether the models generate small conditional pricing errors out of sample.

To quantify the magnitude of pricing errors, we compute out-of-sample conditional alphas for each of the 160 portfolios. In line with Kelly et al. (2019a) and Gu et al. (2021), the conditional alpha of portfolio m is defined as the time-series average of its out-of-sample residuals:

$$\alpha_m = \frac{1}{T}\sum_{t=1}^{T}\Big(x_{m,t} - \boldsymbol{f}_t\boldsymbol{\beta}^{\prime}_{m,t-1}\Big), \tag{14}$$

where $x_{m,t}$ denotes the return of portfolio m, $\boldsymbol{f}_t$ is the vector of out-of-sample factor realizations at time t, and $\boldsymbol{\beta}_{m,t-1}$ is the corresponding vector of conditional factor loadings estimated using only information available prior to t. Note that we do not re-estimate the models to test their generalizability fore test assets that were not part of the model estimation.

Figure 4 plots the out-of-sample conditional alphas against the portfolios' average returns for three representative models: the instrumented CAPM, a four-factor IPCA specification, and the five-factor C–RNN. In the left panel, the instrumented CAPM exhibits a clear linear relation between alphas and average returns. The pricing errors closely align with the 45-degree line, indicating that portfolios with higher average returns systematically produce larger alphas. This pattern reveals the model's inability to account for the return structures embedded in the test assets. Quantitatively, the instrumented CAPM produces an average absolute alpha of 0.26%, an average absolute t-statistic of 1.77, and leaves 60 portfolios significantly mispriced assuming a significance level of 5%.

[14]Note that we do not use *mktcap* in these double-sorts, leaving 32 characteristics.

In contrast, both IPCA and the C–RNN largely eliminate the strong linear relation between alphas and average returns. In the IPCA panel, alphas no longer display a systematic upward trend with returns; their dispersion is noticeably smaller, with the average absolute alpha falling to 0.18%. Despite this improvement, 50 portfolios remain significantly mispriced, indicating that the static factor structure still struggles to reconcile all cross-sectional patterns in the anomaly set.

The C–RNN produces the smallest and least structured pricing errors among the three models. The linear relationship between alphas and returns is virtually absent, and the cloud of points is substantially more concentrated around zero. Average absolute alphas drop to 0.14%, and only 35 portfolios exhibit statistically significant abnormal returns. This represents the strongest performance across the three models and suggests that recurrent conditional factor and loading dynamics capture a greater portion of the economic forces driving cross-sectional anomalies.

[Insert Figure 4 about here]

4.4. C-RNN vs. Economic Restrictions

Recent evidence suggests that strong statistical or portfolio performance alone is insufficient to assess the practical relevance of return-forecasting models. Avramov et al. (2023) show that many machine learning–based strategies extract their apparent profitability from stocks that are difficult to arbitrage, rely heavily on extreme short positions, earn disproportionately more in periods when limits to arbitrage are high, and fail to survive trading costs. In light of these concerns, it is crucial to establish whether the performance of the C–RNN and related conditional factor models arises from economically fragile sources or whether it reflects robust patterns in expected returns.

Table 3 already indicates that the profitability of the expected return-sorted strategy is not driven by extreme shorting but rather by the long leg of the portfolio, and the BETC rate of 0.72% is substantial. This suggests that the strategy remains economically viable after accounting for moderate trading frictions, contrasting with many machine learning approaches documented in Avramov et al. (2023).

To further assess the economic robustness of the model-implied forecasts, we examine two additional dimensions emphasized in the literature: (i) the dependence of performance on small, potentially illiquid stocks and (ii) the sensitivity of profitability to market conditions. Both dimensions help determine whether a model extracts returns from segments of the market where risk compensation is hard to distinguish from temporary mispricing.

Panel A of Table 4 examines whether the performance of the C–RNN expected return–sorted strategy is driven by small, difficult-to-arbitrage stocks. For each month, the strategy is implemented using only the largest 70%, 50%, and 30% of firms, respectively. If the model's profitability were concentrated among small and potentially difficult-to-trade stocks, we would expect performance to deteriorate sharply as smaller firms are excluded.

[Insert Table 4 about here]

Overall, the decline in performance is gradual and far from dramatic. Total R^2s naturally increase as the sample is restricted to larger firms, since returns of large-cap stocks are more strongly driven by common factor variation and contain less idiosyncratic noise. When restricting the universe to the largest 70% of stocks, the predictive R^2 is still 0.28% and the strategy achieves an FF6 alpha of 1.39%, being statistically significant at the 1% level. Narrowing the universe further to the largest 50% of firms, the predictive R^2 becomes negative, however, the strategy still yields an remarkable alpha of 1.24%. Finally, when restricting the sample to only the largest 30% of stocks, the alpha remains sizable at 0.95% (t-stat = 5.20), accompanied by a Sharpe ratio of 0.73, indicating that the strategy continues to perform well even when the influence of small, potentially harder-to-arbitrage stocks is minimized.

To examine whether refitting the C-RNN exclusively on the subset of large stocks improves its performance, Panel A also reports results from re-estimating the model on each restricted universe. As expected, R^2s increase as smaller and noisier firms are removed from training, allowing the model to better capture patterns across larger stocks. However, portfolio performance remains remarkably stable and unaffected from re-fitting. These results indicate that the C-RNN estimated on the full sample generalizes well to the large-stock universe and that its economic performance is not improved by re-estimation on restricted sets of firms.

The literature documents that the profitability of many stock-market anomalies varies with market conditions, often strengthening when limits to arbitrage are high—such as during periods of elevated volatility, uncertainty, or negative sentiment (Stambaugh et al., 2012; Nagel, 2012; Avramov et al., 2016). Recent evidence in Avramov et al. (2023) further shows that the profitability of most machine-learning-based return predictors is concentrated in states characterized by high sentiment-driven mispricing and tight arbitrage constraints. To assess whether the C-RNN exhibits similar sensitivities, we examine performance across a range of subperiods and market environments.

Specifically, the sample is partitioned into: (1) two equally long subsamples (first and second halves), (2) bear and bull markets based on the 12-month cumulative market return, (3) low and high market volatility using past 36-month realized volatility, (4) low and high economic policy uncertainty (EPU), and (5) low and high macroeconomic uncertainty (UNC) as measured by the Bekaert et al. (2022) uncertainty index. Following Avramov et al. (2023), a month is classified as "low" ("high") if the respective $t-1$ state variable falls below (above) its full-sample median.

Overall, the C-RNN delivers significant abnormal returns across all market states, indicating that its performance is not confined to periods associated with higher mispricing pressure. While raw H–L returns tend to be somewhat higher in the first half of the sample, risk-adjusted performance (FF6 alphas) is actually stronger in the second half, suggesting that the model's effectiveness has remained stable or even improved over time. The strategy shows no substantial dependence on past market performance, generating comparable alphas in both bear and bull markets. Consistent with theories of time-varying limits to arbitrage, performance improves when market volatility is high, yet returns remain economically and statistically significant even in low-volatility environments. Taken together, these results indicate that the C-RNN captures a persistent and robust return signal that is not restricted to specific market regimes.

4.5. Opening the Black Box

Having established that the C-RNN delivers strong statistical and economic performance—and that these results are robust across firm size subsamples and different market states—the natural next question is: what can we learn from the model? A central question in the recent asset pricing literature concerns how much (and what kind of) conditioning information is actually needed to explain the cross-section of expected returns. A growing body of work argues that only a small number of characteristics or latent factors are essential, while others suggest that broader sets of signals capture incremental structure in returns (see, for example, Green et al., 2017; Kozak et al., 2020; Bryzgalova et al., 2023; Swade et al., 2024; Didisheim et al., 2025, for a discussion on that question). Existing empirical frameworks, however, differ sharply in how they aggregate information. For example, the IPCA allows researchers to evaluate the combined contribution of characteristics to the factor structure, but it does not disentangle whether predictability originates from the conditioning of factors, the conditioning of loadings, or both.

By contrast, the C-RNN separates these two channels: (1) factor instruments that shape the latent factors and (2) characteristic instruments that shape the factor loadings. This separation allows us to ask a more refined question: which specific empirical factors and which firm characteristics matter most for describing return comovement and expected returns—out of sample and in a fully dynamic setting?

To address this question on the factor side, we examine the out-of-sample importance of each factor instrument. Importance is measured as the decline in the model's total R^2 when all coefficients in $\mathbf{\Omega}$ associated with a given instrument are set to zero (Kelly et al., 2019a; Gu et al., 2021). Unlike prior in-sample analyses such as Fieberg et al. (2025a), the importance measures here are computed entirely out-of-sample following the same recursive estimation scheme used throughout the empirical analysis.

The results in Table 5 reveal a striking pattern: only a handful of empirical factors meaningfully contribute to the C-RNN's ability to explain return comovement. The market factor stands out as by far the most important instrument, and removing it reduces the out-of-sample total R^2 by 7.54 percentage points. The next most important factor is SMB, whose exclusion lowers the fit by an additional 2.63 percentage points points. Together, these two factors alone account for the vast majority of the conditioning information used by the model. Beyond these core instruments, the marginal contributions drop sharply. WML and QMJ exert moderate influence, each reducing the total R^2 by roughly half a percentage point or less. HML and RMW contribute even less, and most of the remaining instruments—including reversal, liquidity, financing, and expected growth factors—have near-zero importance.

[Insert Table 5 about here]

Turning to the factor loading instruments, a very different pattern emerges. Table 6 reports the out-of-sample decline in total R^2 when the coefficients in $\mathbf{\Gamma}_\beta$ associated with a given characteristic are set to zero. In contrast to the sparse structure observed for factor instruments, a broad set of characteristics contributes meaningfully to the model's performance: asset turnover (ato), operating leverage (ol), total assets (assets), market capitalization (mktcap), and market beta (beta) all produce reductions between 0.40 and 0.90 percentage points when removed. Measures of distress or frictions—such as idiosyncratic volatility (idiovol), leverage (lev), and short-term reversal (strev) also exhibit nontrivial contributions. Importantly, investment- and profitability-related characteristics (such as a2me, pm, roa, and roe), which the literature consistently identifies as central drivers of expected returns, also appear as meaningful instruments in shaping loadings, though their marginal out-of-sample impact is somewhat smaller than that of the leading characteristics.

Although many individual characteristics contribute only modestly when considered in isolation, the set of influential characteristics is broad rather than narrow. This indicates that the C-RNN draws on a rich cross-sectional information set to model factor exposures. This stands in sharp contrast to the factor instrument results, where only a few empirical factors matter substantially.

[Insert Table 6 about here]

Together, these findings point to a clear conclusion: few factor-mimicking portfolios suffice to describe the dynamics of the latent factors, but the heterogeneity in factor loadings is inherently high-dimensional and requires conditioning on many firm characteristics. Conditioning both factors and factor loadings on different sets of variables reconciles competing views in the literature: while return comovement is driven by a small number of core economic sources, capturing the cross-sectional variation in risk exposures demands a much richer set of firm-level instruments.

5. Conclusion

Machine learning has now become an integral part of empirical asset pricing, reshaping how researchers extract latent structures, forecast returns, and evaluate competing theories of market efficiency and mispricing. This paper contributes to this growing literature by proposing the C-RNN, a flexible and interpretable recurrent factor model that jointly conditions latent factors and factor loadings on observable information while allowing both to follow dynamic autoregressive processes.

Using more than sixty years of U.S. equity data, we show that the C-RNN consistently outperforms existing linear and nonlinear benchmark models in both statistical and economic dimensions. The model achieves substantially higher predictive R^2 for individual stock returns and produces large and statistically significant high-minus-low portfolio returns that persist across firm-size subsamples, market regimes, and periods of varying uncertainty. These results highlight the importance of combining conditioning information with recurrent dynamics when estimating risk premia.

Compared to other black-box machine learning approaches, the C-RNN is highly interpretable and reveals an informative structural pattern. Only a small number of factor-mimicking portfolios meaningfully drive stock return comovement, whereas a broad set of firm characteristics drives differences in factor loadings. This asymmetry suggests that while common return variation is driven by a limited number of underlying forces, the heterogeneity in risk exposures across firms is inherently high-dimensional—reconciling recent debates about the dimensionality of the pricing kernel.

Overall, the C-RNN provides a unified, economically interpretable, and highly effective framework for modeling the dynamic structure of asset returns. By integrating conditional information with recurrent dynamics, it offers a promising direction for future work on factor models, anomaly detection, and machine learning-based asset pricing theory. Moreover, the architecture can naturally be extended in future research to capture nonlinear return–characteristic interactions in the spirit of Gu et al. (2021). It can also be augmented with long-term memory components such as long short-term memory (LSTM) units (Hochreiter and Schmidhuber, 1997), enabling the model to capture more persistent temporal dependencies in the evolution of factors and factor loadings.

References

Asness, C.S., Frazzini, A., Pedersen, L.H., 2019. Quality minus junk. Review of Accounting Studies 24, 34–112. doi:10.1007/s11142-018-9470-2.

Avramov, D., Cheng, S., Hameed, A., 2016. Time-varying liquidity and momentum profits. Journal of Financial and Quantitative Analysis 51, 1897–1923. doi:10.1017/S0022109016000764.

Avramov, D., Cheng, S., Metzker, L., 2023. Machine learning vs. economic restrictions: Evidence from stock return predictability. Management Science 69, 2587–2619. doi:10.1287/mnsc.2022.4449.

Avramov, D., Chordia, T., 2006. Asset pricing models and financial market anomalies. Review of Financial Studies 19, 1001–1040. doi:10.1093/rfs/hhj025.

Bai, J., Wang, P., 2015. Identification and bayesian estimation of dynamic factor models. Journal of Business & Economic Statistics 33, 221–240. doi:10.1080/07350015.2014.941467.

Bali, T.G., Beckmeyer, H., Mörke, M., Weigert, F., 2023. Option return predictability with machine learning and big data. Review of Financial Studies 36, 3548–3602. doi:10.1093/rfs/hhad017.

Bali, T.G., Cakici, N., Tang, Y., 2009. The conditional beta and the cross–section of expected returns. Financial Management 38, 103–137. doi:10.1111/j.1755-053X.2009.01030.x.

Bali, T.G., Engle, R.F., Tang, Y., 2017. Dynamic conditional beta is alive and well in the cross section of daily stock returns. Management Science 63, 3760–3779. doi:10.1287/mnsc.2016.2536.

Bekaert, G., Engstrom, E.C., Xu, N.R., 2022. The time variation in risk appetite and uncertainty. Management Science 68, 3975–4004. doi:10.1287/mnsc.2021.4068.

Bianchi, D., Büchner, M., Tamoni, A., 2021. Bond risk premiums with machine learning. Review of Financial Studies 34, 1046–1089.

Bryzgalova, S., Huang, J., Julliard, C., 2023. Bayesian solutions for the factor zoo: We just ran two quadrillion models. Journal of Finance 78, 487–557. doi:10.1111/jofi.13197.

Bryzgalova, S., Pelger, M., Zhu, J., 2025. Forest through the trees: Building cross-sections of stock returns. The Journal of Finance 80, 2447–2506. doi:10.1111/jofi.13477.

Büchner, M., Kelly, B., 2022. A factor model for option returns. Journal of Financial Economics 143, 1140–1161. doi:10.1016/j.jfineco.2021.12.007.

Cakici, N., Fieberg, C., Metko, D., Zaremba, A., 2023. Machine learning goes global: Cross-sectional return predictability in international stock markets. Journal of Economic Dynamics and Control 155, 104725. doi:10.1016/j.jedc.2023.104725.

Carhart, M.M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57–82. doi:10.1111/j.1540-6261.1997.tb03808.x.

Chen, A.Y., Zimmermann, T., 2022. Open source cross-sectional asset pricing. Critical Finance Review 27, 207–264.

Chen, L., Pelger, M., Zhu, J., 2024. Deep learning in asset pricing. Management Science 70, 714–750. doi:10.1287/mnsc.2023.4695.

Cochrane, J.H., 2001. Asset Pricing. Princeton University Press, Princeton.

Daniel, K., Hirshleifer, D., Sun, L., 2020. Short- and long-horizon behavioral factors. Review of Financial Studies 33, 1673–1736. doi:10.1093/rfs/hhz069.

Del Negro, M., Otrok, C.M., 2008. Dynamic factor models with time-varying parameters: Measuring changes in international business cycles. Working Paper. doi:10.2139/ssrn.1136163.

Didisheim, A., Ke, S., Kelly, B.T., Malamud, S., 2025. APT or "AIPT"? the surprising dominance of large factor models. Working Paper.

Elman, J., 1990. Finding structure in time. Cognitive Science 14, 179–211. doi:10.1016/0364-0213(90)90002-E.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56. doi:10.1016/0304-405X(93)90023-5.

Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22. doi:10.1016/j.jfineco.2014.10.010.

Fama, E.F., French, K.R., 2018. Choosing factors. Journal of Financial Economics 128, 234–252. doi:10.1016/j.jfineco.2018.02.012.

Fama, E.F., French, K.R., 2020. Comparing cross-section and time-series factor models. Review of Financial Studies 33, 1891–1926. doi:10.1093/rfs/hhz089.

Ferson, W.E., 1990. Are the latent variables in time-varying expected returns compensation for consumption risk? Journal of Finance 45, 397–429. doi:10.1111/j.1540-6261.1990.tb03696.x.

Ferson, W.E., Harvey, C.R., 1991. The variation of economic risk premiums. Journal of Political Economy 99, 385–415.

Ferson, W.E., Harvey, C.R., 1993. The risk and predictability of international equity returns. Review of Financial Studies 6, 527–566.

Fieberg, C., Liedtke, G., Poddig, T., 2025a. Recurrent double-conditional factor model. OR Spectrum 47, 205–254.

Fieberg, C., Liedtke, G., Poddig, T., Walker, T., Zaremba, A., 2025b. A trend factor for the cross-section of cryptocurrency returns. Journal of Financial and Quantitative Analysis 60, 3116–3153. doi:10.1017/S0022109024000747.

Fieberg, C., Liedtke, G., Zaremba, A., Cakici, N., 2025c. A factor model for the cross-section of country equity risk premia. Journal of Banking & Finance 171, 107373. doi:10.1016/j.jbankfin.2024.107373.

Fieberg, C., Metko, D., Poddig, T., Loy, T., 2023. Machine learning techniques for cross-sectional equity returns' prediction. OR Spectrum 45, 289–323. doi:10.1007/s00291-022-00693-w.

Fischer, T., Krauss, C., 2018. Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research 270, 654–669. doi:10.1016/j.ejor.2017.11.054.

Frazzini, A., Pedersen, L.H., 2014. Betting against beta. Journal of Financial Economics 111, 1–25. doi:10.1016/j.jfineco.2013.10.005.

Freyberger, J., Neuhierl, A., Weber, M., 2020. Dissecting characteristics nonparametrically. Review of Financial Studies 33, 2326–2377. doi:10.1093/rfs/hhz123.

Goyal, A., Saretto, A., 2025. Can equity option returns be explained by a factor model? IPCA says yes. Review of Financial Studies 38, 1783–1821. doi:10.1093/rfs/hhae087.

Green, J., Hand, J.R.M., Zhang, X.F., 2017. The characteristics that provide independent information about average U.S. monthly stock returns. Review of Financial Studies 30, 4389–4436. doi:10.1093/rfs/hhx019.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. Review of Financial Studies 33, 2223–2273. doi:10.1093/rfs/hhaa009.

Gu, S., Kelly, B., Xiu, D., 2021. Autoencoder asset pricing models. Journal of Econometrics 222, 429–450. doi:10.1016/j.jeconom.2020.07.009.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Computation 9, 1735–1780. doi:10.1162/neco.1997.9.8.1735.

Hou, K., Mo, H., Xue, C., Zhang, L., 2021. An augmented q -factor model with expected growth. Review of Finance 25, 1–41. doi:10.1093/rof/rfaa004.

Israel, R., Kelly, B.T., Moskowitz, T., 2020. Can machines learn finance? Journal of Investment Management 18, 23–36.

Jensen, T.I., Kelly, B.T., Malamud, S., Pedersen, L.H., 2022. Machine learning and the implementable efficient frontier. Review of Financial Studies (Forthcoming). URL: https://doi.org/10.2139/ssrn.4187217.

Jensen, T.I., Kelly, B.T., Pedersen, L.H., 2023. Is there a replication crisis in finance? Journal of Finance 78, 2465–2518. doi:10.1111/jofi.13249.

Jiang, J., Kelly, B., Xiu, D., 2023. (Re-)Imag(in)ing Price Trends. Journal of Finance 78, 3193–3249. doi:10.1111/jofi.13268.

Jungbacker, B., Koopman, S.J., van der Wel, M., 2011. Maximum likelihood estimation for dynamic factor models with missing data. Journal of Economic Dynamics and Control 35, 1358–1368. doi:10.1016/j.jedc.2011.03.009.

Kelly, B.T., Palhares, D., Pruitt, S., 2023. Modeling corporate bond returns. Journal of Finance 78, 1967–2008. doi:10.1111/jofi.13233.

Kelly, B.T., Pruitt, S., Su, Y., 2019a. Characteristics are covariances: A unified model of risk and return. Journal of Financial Economics 134, 501–524. doi:10.1016/j.jfineco.2019.05.001.

Kelly, B.T., Pruitt, S., Su, Y., 2019b. Instrumented principal component analysis. Working Paper.

Kose, M.A., Otrok, C., Whiteman, C.H., 2003. International business cycles: World, region, and country-specific factors. American Economic Review 93, 1216–1239. doi:10.1257/000282803769206278.

Kozak, S., Nagel, S., Santosh, S., 2020. Shrinking the cross-section. Journal of Financial Economics 135, 271–292. doi:10.1016/j.jfineco.2019.06.008.

Lettau, M., Pelger, M., 2020. Factors that fit the time series and cross-section of stock returns. Review of Financial Studies 33, 2274–2325. doi:10.1093/rfs/hhaa020.

Maio, P., 2013. Intertemporal capm with conditioning variables. Management Science 59, 122–141. doi:10.1287/mnsc.1120.1557.

Merton, R.C., 1973. An intertemporal capital asset pricing model. Econometrica 41, 867–887. doi:10.2307/1913811.

Murray, S., Xia, Y., Xiao, H., 2024. Charting by machines. Journal of Financial Economics 153, 103791. doi:10.1016/j.jfineco.2024.103791.

Nagel, S., 2012. Evaporating liquidity. Review of Financial Studies 25, 2005–2039. doi:10.1093/rfs/hhs066.

Nagel, S., 2021. Machine Learning in Asset Pricing. Princeton Lectures in Finance, Princeton University Press, New Jersey.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Pástor, Ľ., Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642–685. doi:10.1086/374184.

Riedmiller, M., Braun, H., 1992. Rprop - a fast adaptive learning algorithm. Proceedings of the International Symposium on Computer and Information Science 7.

Riedmiller, M., Braun, H., 1993. A direct adaptive method for faster backpropagation learning: The Rprop algorithm. Proceedings of the IEEE International Conference on Neural Networks , 586–591doi:10.1109/ICNN.1993.298623.

Ross, S.A., 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341–360. doi:10.1016/0022-0531(76)90046-6.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations ed., MIT, Cambridge.

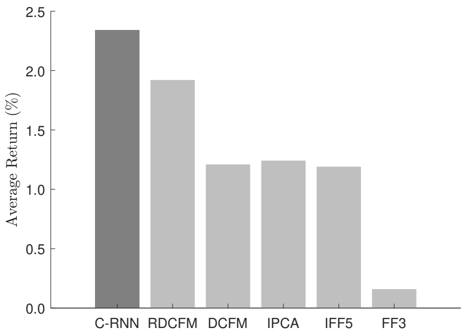
Stambaugh, R.F., Yu, J., Yuan, Y., 2012. The short of it: Investor sentiment and anomalies. Journal of Financial Economics 104, 288–302. doi:10.1016/j.jfineco.2011.12.001.

Swade, A., Hanauer, M.X., Lohre, H., Blitz, D., 2024. Factor zoo. Journal of Portfolio Management 50, 11–31. doi:10.3905/jpm.2023.1.561.

Figure 1: Comparison of Asset Pricing Models

The figure shows average returns (a) and annualized Sharpe ratios (b) for long-short strategies based on the C-RNN, RDCFM, DCFM, IPCA, instrumented Fama and French (2015) five-factor model, and the unconditional Fama and French (2015) three factor model. The sample spans the out-of-sample period from July 1972 to December 2022.

(a) Average Returns



(b) Sharpe Ratios

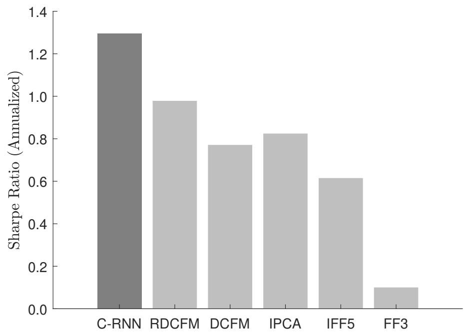


Figure 2: Conditional Recurrent Neural Network

This figure illustrates the architecture of the C–RNN over three consecutive time steps. The core RNN structure (black arrows) represents the evolution of latent factors and their mapping into asset returns. The blue arrows highlight how the RDCFM extends a standard RNN by incorporating an additional characteristics layer that conditions the factor loadings on observable firm observables. The red arrows show the further extension introduced by the C–RNN, which allows factor loadings themselves to evolve dynamically through a recurrent structure. Dashed arrows indicate that the recurrent relationships extend beyond the three periods shown, both into the past and the future.

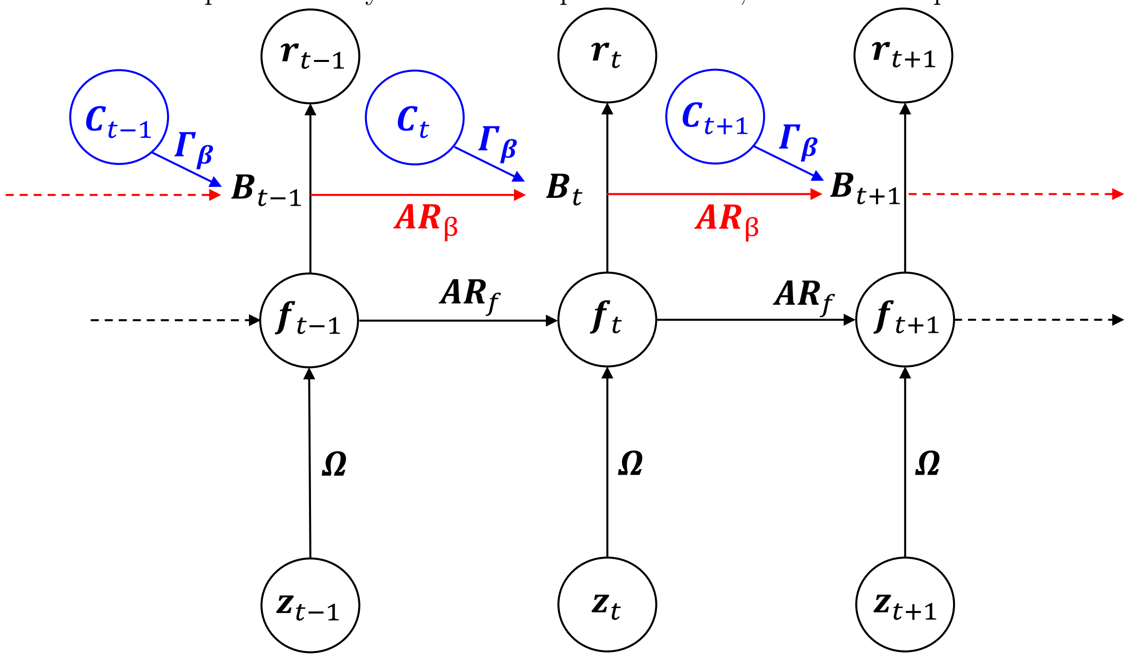


Figure 3: Cumulative Returns of Hedge Portfolios

The figure shows cumulative returns of zero-investment portfolios with a long (short) position in stocks with high (low) expected returns, derived from either the five-factor DRDCFM, RDCFM, DCFM, IPCA, or conditional and unconditional empirical factor models. All portfolios are value-weighted and rebalanced monthly. The sample period is from July 1972 to December 2022.

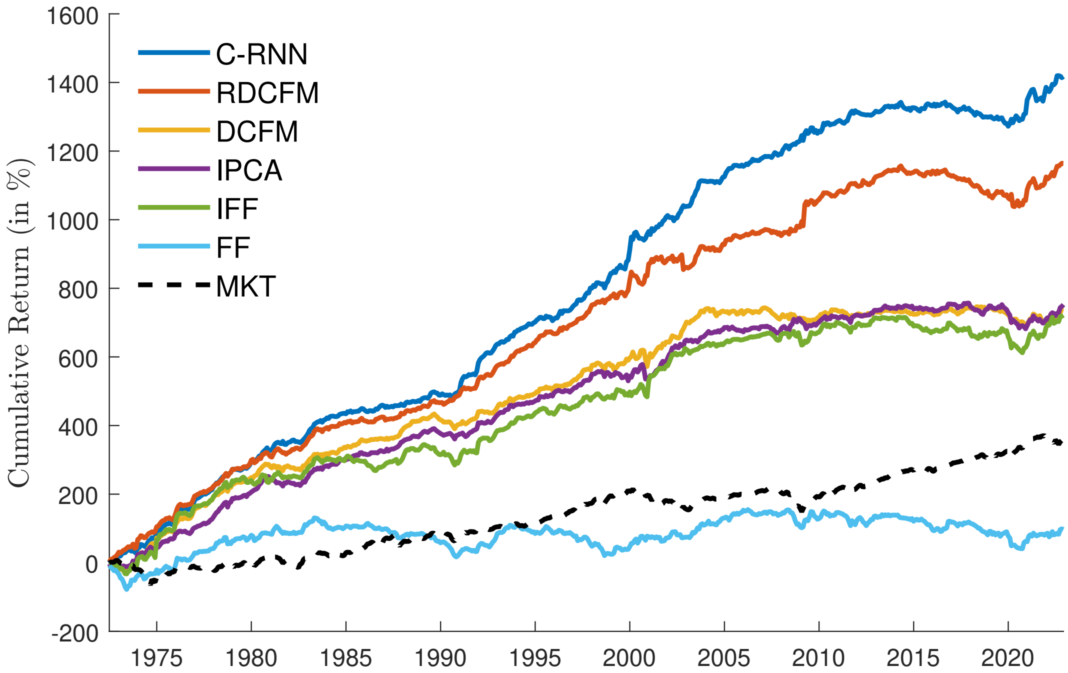


Figure 4: Conditional Alphas

This figure plots out-of-sample conditional alphas from the instrumented CAPM, the four-factor IPCA, and the five-factor C-RNN for 160 double-sorted zero-investment portfolios formed from 32 firm characteristics. The vertical axis reports alphas, while the horizontal axis shows the corresponding average returns (*Avg.*). All values are expressed in percentage terms. Black filled dots indicate alphas that are significantly different from zero at the 5% level. In the lower-right corner, we report the average absolute alpha (Avg $|\alpha|$), the average absolute t-statistic (Avg $|t|$), and the number of abnormal returns significant at the 5% level. Standard errors used to compute the t-statistics follow the Newey and West (1987) adjustment with four lags. The sample covers the out-of-sample period from July 1972 to December 2022.

(a) Instrumented CAPM **(b) IPCA ($K = 4$)** **(c) C-RNN ($K = 5$)**

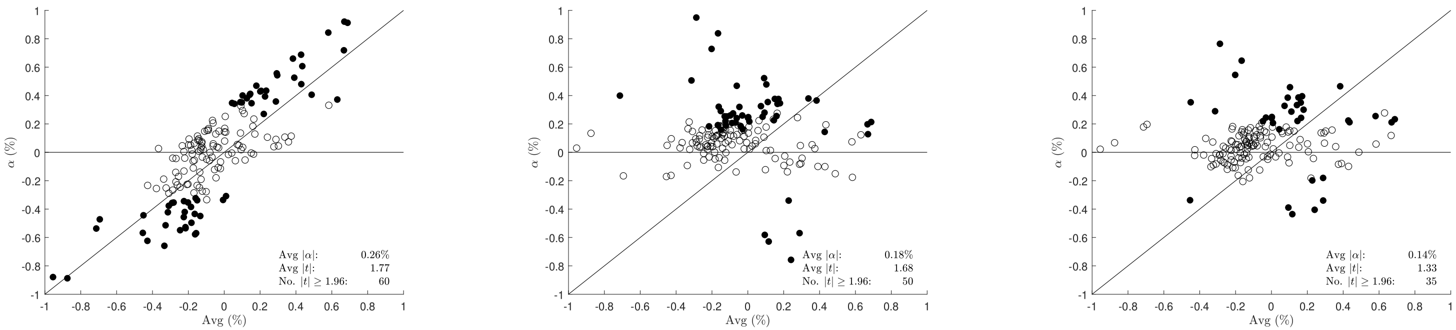


Table 1: Individual Stock Pricing Performance

The table reports the out-of-sample total R^2 and predictive R^2 (in %) for 18,258 individual stocks for the C-RNN, RDCFM, DCFM, IPCA, and empirical asset pricing models with unconditional and conditional factor loadings. All models are estimated assuming $K = 1, \ldots, 6$ factors. The out-of-sample period starts in July 1972 and ends in December 2022.

	K					
	1	2	3	4	5	6
Panel A: Conditional Recurrent Neural Net						
Total R^2	13.0	13.7	14.5	15.0	15.3	15.4
Pred R^2	0.51	0.56	0.74	0.72	0.81	0.78
Panel B: Recurrent Double-Conditional Factor Model						
Total R^2	13.0	13.7	14.1	15.0	15.2	15.3
Pred R^2	0.51	0.56	0.72	0.78	0.75	0.77
Panel C: Double-Conditional Factor Model						
Total R^2	13.0	13.7	14.1	14.4	14.7	13.3
Pred R^2	0.31	0.37	0.48	0.54	0.56	0.56
Panel D: Instrumented Principal Component Analysis						
Total R^2	13.8	15.5	16.5	17.0	17.5	17.8
Pred R^2	0.29	0.30	0.48	0.52	0.52	0.52
Panel E: Instrumented Empirical Factors						
Total R^2	10.4		13.9	14.7	14.0	14.8
Pred R^2	0.22		0.35	0.36	0.42	0.41
Panel F: Unconditional Empirical Factors						
Total R^2	9.07		9.77	7.45	6.36	3.84
Pred R^2	0.20		0.20	0.07	0.11	0.01

Table 2: Characteristic-Managed Portfolio Pricing Performance

The table reports the out-of-sample total R^2 and predictive R^2 (in %) for 34 characteristic-managed portfolios for the C-RNN, RDCFM, DCFM, IPCA, and empirical asset pricing models with unconditional and conditional factor loadings. All models are estimated assuming $K = 1, \ldots, 6$ factors. The out-of-sample period starts in July 1972 and ends in December 2022.

	K					
	1	2	3	4	5	6
Panel A: Conditional Recurrent Neural Net						
Total R^2	82.6	84.9	86.3	87.0	88.2	88.3
Pred R^2	2.74	2.72	3.04	2.46	2.78	2.43
Panel B: Recurrent Double-Conditional Factor Model						
Total R^2	82.7	85.1	85.6	88.3	88.7	88.8
Pred R^2	2.74	2.79	2.75	2.93	2.86	3.11
Panel C: Double-Conditional Factor Model						
Total R^2	82.3	84.6	85.6	85.4	87.0	84.6
Pred R^2	1.47	1.57	1.59	1.70	1.56	1.59
Panel D: Instrumented Principal Component Analysis						
Total R^2	88.4	94.4	96.6	97.3	98.3	98.8
Pred R^2	1.34	1.32	1.51	1.57	1.50	1.48
Panel E: Instrumented Empirical Factors						
Total R^2	64.3		84.5	86.7	85.1	87.4
Pred R^2	1.10		1.47	1.50	1.60	1.57
Panel F: Unconditional Empirical Factors						
Total R^2	63.2		82.0	84.4	82.9	85.7
Pred R^2	1.16		1.50	1.44	1.59	1.50

Table 3: Expected-Return Sorted Portfolios

This table summarizes portfolio performance for the five-factor C-RNN and several benchmark approaches. Panel A reports results for the ten value-weighted decile portfolios formed using predicted returns from the five-factor C-RNN, which are rebalanced at the end of each month. Reported statistics include average monthly return (in %) (Avg), t-statistics, standard deviation (in %), annualized Sharpe ratio, Fama and French (2018) six-factor model alpha (in %) and its t-statistic, and the breakeven transaction costs rate (BETC). Panel B presents the corresponding high-minus-low portfolio performance for benchmark models. Standard errors for computing the t-statistics are Newey and West (1987)-adjusted using four lags. All portfolios are evaluated over the sample period from July 1972 to December 2022.

	Avg	t(Avg)	Std	Shp	α^{FF6}	$t(\alpha^{FF6})$	BETC
Panel A: Performance of C-RNN							
Low	-0.31	-1.23	6.01	-0.18	-0.79	-6.14	0.00
2	0.43	1.93	5.30	0.28	-0.15	-1.40	0.25
3	0.55	2.58	5.23	0.36	-0.17	-1.56	0.31
4	0.59	2.85	5.04	0.40	-0.11	-1.27	0.33
5	0.73	3.39	5.09	0.49	0.08	0.84	0.41
6	0.71	3.39	5.17	0.48	-0.04	-0.46	0.40
7	0.89	3.69	5.69	0.54	0.21	1.88	0.49
8	1.23	4.59	5.98	0.72	0.58	4.80	0.68
9	1.44	4.71	6.46	0.77	0.72	4.40	0.80
High	2.03	6.04	7.26	0.97	1.37	6.79	1.20
H-L	2.34	8.33	6.26	1.30	2.16	8.11	0.72
Panel B: Comparison With Other Models							
RDCFM	1.92	6.81	6.80	0.98	1.58	4.74	0.59
DCFM	1.21	5.28	5.43	0.77	0.85	3.62	0.37
IPCA	1.24	5.38	5.22	0.82	0.62	3.16	0.39
IFF	1.19	4.14	6.71	0.62	0.61	2.22	0.51
FF	0.16	0.62	5.51	0.10	-0.42	-2.47	0.69

Table 4: Performance Depending on Small Stocks and Market States

The table shows the statistical and economic performance of the five-factor C-RNN depending on small stocks and market states. Panel A shows the metrics for subsample of stocks, where stocks are excluded based on their month $t-1$ market capitalization. The results are shown for the C-RNN estimated on the full sample as well as the respective subsamples. Panel B shows the metrics in the first and second half of the out-of-sample period, as well as in bear and bull market states and states characterized by low/high volatility, economic policy uncertainty (EPU), and aggregate macroeconomic uncertainty (UNC). Statistical pricing performance is evaluated based on the total and predictive R^2s. Economic performance is measured by the average monthly return (in %) of a zero-investment portfolio that takes a long (short) position in the 10% stocks with the highest (lowest) expected returns. The alphas (in %) against the Fama and French (2018) six-factor model and the annualized Sharpe ratio are also reported. t-statistics are reported in parentheses, and standard errors to compute these are adjusted using the Newey and West (1987) method assuming four lags. Statistical significance at the 5% level is indicated by bold font. All results are computed using out-of-sample estimates from July 1972 to December 2022.

	Total R^2	Pred R^2	Avg	α^{FF6}	Shp
Panel A: Excluding Small Stocks					
Full sample estimation:					
Largest 70%	23.22	0.28	**1.45**	**1.39**	0.91
			(6.00)	(6.35)	
Largest 50%	25.81	-0.02	**1.24**	**1.17**	0.86
			(5.94)	(5.79)	
Largest 30%	28.58	-0.32	**0.95**	**0.88**	0.73
			(5.20)	(4.87)	
Subsample estimation:					
Largest 70%	24.23	0.37	**1.21**	**1.35**	0.65
			(4.52)	(5.24)	
Largest 50%	27.01	0.32	**1.17**	**1.17**	0.68
			(5.11)	(4.95)	
Largest 30%	30.29	0.17	**0.98**	**0.81**	0.66
			(4.87)	(4.12)	
Panel B: Dependence on Market States					
First Half	15.6	1.09	**2.58**	**1.61**	1.91
			(8.71)	(5.27)	
Second Half	15.1	0.56	**2.10**	**2.39**	0.97
			(4.32)	(5.61)	
Bear Market	17.4	0.81	**2.36**	**2.51**	1.46
			(7.60)	(7.43)	
Bull Market	13.1	0.81	**2.32**	**2.04**	1.17
			(4.95)	(5.14)	

Table 4: Performance Depending on Small Stocks and Market States (continued)

	Total R^2	Pred R^2	Avg	α^{FF6}	Shp
Low Volatility	15.1	0.32	**1.74**	**1.47**	1.22
			(5.60)	(3.91)	
High Volatility	15.5	1.18	**2.93**	**2.68**	1.39
			(6.36)	(6.59)	
Low EPU	14.3	0.35	**2.29**	**2.25**	1.47
			(6.50)	(6.02)	
High EPU	16.1	1.17	**2.38**	**2.18**	1.18
			(5.37)	(5.30)	
Low UNC	9.64	0.86	**2.21**	**2.01**	1.08
			(4.12)	(4.01)	
High UNC	17.0	0.58	**2.23**	**2.45**	1.17
			(4.77)	(5.14)	

Table 5: Out-of-Sample Factor Instrument Importance

The table shows how much the model's out-of-sample total R^2 (in percentage points) declines when all coefficients in $\mathbf{\Omega}$ linked to a specific factor instrument are set to zero. A larger reduction indicates that the instrument plays a more important role in capturing the common variation in returns. All results are computed using out-of-sample estimates from July 1972 to December 2022.

	Δ Total R^2		Δ Total R^2		Δ Total R^2
MKT	-7.54	STREV	-0.12	LTREV	-0.03
SMB	-2.63	RMW	-0.09	CMA	-0.02
WML	-0.52	BAB	-0.07	LIQ	0.00
QMJ	-0.39	PEAD	-0.07	FIN	0.01
HML	-0.25	EG	-0.04		

Table 6: Out-of-Sample Factor Loading Instrument Importance

The table shows how much the model's out-of-sample total R^2 (in percentage points) declines when all coefficients in $\mathbf{\Gamma}_\beta$ linked to a specific characteristic are set to zero. A larger reduction indicates that the instrument plays a more important role in capturing the common variation in returns. All results are computed using out-of-sample estimates from July 1972 to December 2022.

	Δ Total R^2		Δ Total R^2		Δ Total R^2
ato	-0.90	w52h	-0.09	dpi2a	-0.01
ol	-0.62	cto	-0.07	roa	-0.01
assets	-0.54	pm	-0.05	pcm	-0.00
mktcap	-0.52	roe	-0.03	rna	-0.00
beta	-0.43	c	-0.02	intmom	-0.00
a2me	-0.43	freecf	-0.01	invest	0.00
idiovol	-0.36	ltrev	-0.01	turnover	0.00
strev	-0.34	sga2m	-0.01	d2a	0.00
lev	-0.27	oa	-0.01	e2p	0.01
mom	-0.15	bm	-0.01	prof	0.01
s2p	-0.14	noa	-0.01	bidask	0.03

Online Appendix for

"Recurrent Neural Networks Meet Asset Pricing"

Appendix A. Additional Results from the Study

Table A.1: Differences in Average H–L Returns Across Models

This table reports pairwise differences in average monthly H-L returns (in %) between the C-RNN, RDCFM, DCFM, IPCA, and unconditional and instrumented empirical factor models. Positive values indicate that the model shown in rows outperforms the model in the column. *t*-statistics are reported in parentheses, and standard errors to compute these are adjusted using the Newey and West (1987) method assuming four lags. Statistical significance at the 5% level is indicated by bold font. All results are computed using out-of-sample estimates from July 1972 to December 2022.

	RDCFM	DCFM	IPCA	IFF	FF
C-RNN	**0.42**	**1.13**	**1.10**	**1.15**	**2.18**
	(2.00)	(4.12)	(3.72)	(3.89)	(7.24)
RDCFM		**0.71**	**0.68**	**0.73**	**1.76**
		(2.37)	(2.18)	(2.51)	(5.85)
DCFM			-0.03	0.02	**1.05**
			(-0.20)	(0.08)	(4.25)
IPCA				0.05	**1.08**
				(0.23)	(4.08)
IFF					**1.03**
					(4.33)