Journal of Financial Economics 153 (2024) 103791

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Charting by machines

Scott Murray *, Yusen Xia, Houping Xiao

Georgia State University, Robinson College of Business, 35 Broad Street, Atlanta, GA, 30303, United States

A R T I C L E I N F O

Dataset link: https://dx.doi.org/10.17632/x63r376783.2

JEL classification:
G11
G12

Keywords:
Efficient market hypothesis
Machine learning
Deep learning
Charting
Technical analysis
Cross-section of stock returns

A B S T R A C T

We test the efficient market hypothesis by using machine learning to forecast stock returns from historical performance. These forecasts strongly predict the cross-section of future stock returns. The predictive power holds in most subperiods and is strong among the largest 500 stocks. The forecasting function has important nonlinearities and interactions, is remarkably stable through time, and captures effects distinct from momentum, reversal, and extant technical signals. These findings question the efficient market hypothesis and indicate that technical analysis and charting have merit. We also demonstrate that machine learning models that perform well in optimization continue to perform well out-of-sample.

1. Introduction

The weak form of the efficient market hypothesis (EMH hereafter) stipulates that the construction of a profitable portfolio based only on information discernable from plots depicting the historical performance of stocks (price plots hereafter) should not be possible. As such, technical analysis, or charting, should be a fruitless investment technique. Academic research on technical analysis has broadly supported this prediction. Despite the broad dismissal of technical analysis in the academic literature, it remains widely used by investment managers. The continued widespread use of technical analysis suggests that its merit may not be fully discovered in academic research, and that further investigation is warranted.

In this paper, we test the weak form of the EMH by examining whether forecasts produced by machine learning (ML hereafter) predict the cross-section of future stock returns. The forecasts are based only on data that are easily discernable from historical price plots, specifically the cumulative stock returns over the past 12 months. We find strong evidence that the ML-based forecasts have economically important and highly statistically significant predictive power. This predictive power prevails in most subperiods of our focal 196307-202212 test period, including the most recent 201501-202212 subperiod, and remains strong among the largest 500 stocks. The forecasting function is remarkably stable through time and highly complex, with substantial nonlinear and interaction components that are important for prediction. Finally, the predictive power of our ML-based forecasts is not explained by the well-known momentum (Jegadeesh and Titman (1993)) and reversal (Jegadeesh (1990)) effects, nor by previously studied technical or ML-based signals.

Execution of the ML process requires us to make several implementation decisions. To alleviate concerns related to data mining (Harvey et al. (2016)) or out-of-sample forecasting power (McLean and Pontiff (2016), Green et al. (2017)), we use data from 192701-196306, which we refer to as the "optimization period", to select our ML model.[1] Our analyses lead us to use a convolutional neural network with long short-term memory as the ML architecture, mean-squared error as the loss function, a weighting scheme that assigns the same total weight to observations from each month and equal weight to each stock within a

* Corresponding author.
E-mail addresses: smurray19@gsu.edu (S. Murray), ysxia@gsu.edu (Y. Xia), hxiao@gsu.edu (H. Xiao).

[1] Lo and MacKinlay (1990) and Fama (1991) also raise concerns about falsely significant relations between predictive variables and the cross-section of future stock returns. Giglio et al. (2021) propose a methodology to address the issue of multiple testing, raised by Harvey et al. (2016), in the context of linear asset pricing models. Schwert (2003) shows that the magnitudes of the size and value effects decrease after the period examined by the initial research on these effects.

https://doi.org/10.1016/j.jfineco.2024.103791
Received 27 May 2022; Received in revised form 12 January 2024; Accepted 13 January 2024
Available online 24 January 2024

month, and a normalized measure of future return as the dependent variable.[2] Our use of data from only the optimization period to make these ML model decisions ensures that our main test results truly reflect out-of-sample predictive power.

We use the optimized ML model to generate stock return forecasts during the 196307-202212 period, which we refer to as the "test period". Our main analyses examine the performance of portfolios formed by sorting stocks on the ML-based forecasts, which we denote $MLER$. We find that $MLER$ is a strong predictor of the cross-section of future stock returns. The average excess returns of value-weighted decile portfolios constructed using breakpoints calculated from only NYSE-listed stocks increase from −0.14% per month for the decile one portfolio to 0.93% per month for the decile 10 portfolio. The 1.08% per month generated by the portfolio that is long the decile 10 portfolio and short the decile one portfolio is not only economically large, but highly statistically significant, with a t-statistic of 5.51. While our methodology is designed to ensure that our results reflect out-of-sample performance, this and other t-statistics from our asset pricing tests far surpass the benchmark t-statistics proposed by Harvey et al. (2016). We find no evidence that variation in the average returns of the $MLER$-sorted portfolios is attributable to risk. Alphas of the decile portfolios with respect to several established factor models exhibit patterns similar to those of the average excess returns. Risk metrics such as volatility, skewness, value at risk, and expected shortfall do not support a risk-based explanation for the patterns in average returns.

The predictive power of $MLER$ is strong throughout most of our 196307-202212 sample period. The portfolio that is long (short) stocks with high (low) values of $MLER$ generates a large and highly-significant average excess return of more than 1% per month during most subperiods. The exception is the 200501-201412 subperiod, during which the average excess return is close to zero due to a small number of very large monthly losses in 2009. These losses are attributable to the portfolio's unusually large momentum tilt, combined with large momentum strategy losses, in 200904 and 200908. During the most recent 201501-202212 subperiod, the portfolio generates 1.20% per month with an associated t-statistic of 2.13. The predictive power of $MLER$ remains remarkably strong when the sample is restricted to only large stocks. Most notably, when using only the top 500 stocks by market capitalization, the average excess return of the long-short portfolio is 0.72% per month (t-statistic = 4.37).

The objective of our next tests is to characterize the predictive power of the ML-based forecasts. The main challenge in this is that the ML process is a black box methodology and the forecasting function generated by the learning process is difficult to interpret.

The first such tests examine whether the relation between past price patterns and future stock returns is stable through time. We find that this is indeed the case. Forecasts based on fits of subsets of the data that are separated by long periods of time are highly-correlated, and the associated long-short portfolios have substantial common holdings and high return correlation. Furthermore, the performance of the ML-based forecasts is slightly better when the ML model is fit to all available past data compared to when the fitting process is applied to data from rolling-window subperiods.

We then investigate the importance of nonlinearities and interaction terms in the forecasting function, and find them to be responsible for a substantial portion of the variation in forecasts and important for prediction. Nearly two thirds of variation in the forecasts is attributable to nonlinearities and interactions in the forecasting function, and nearly half of this variation is driven by interactions. Regressions of future returns on $MLER$, historical cumulative returns, and terms capturing nonlinearities in the forecasts, indicate that the predictive power of $MLER$ remains strong after controlling for these terms.

Our next tests examine whether the predictive power of the ML-based forecasts is subsumed by previously-documented relations between variables calculated from historical market data and future stock returns. First, we find that while the ML-based forecasts do include components related to the momentum (Jegadeesh and Titman (1993)) and reversal (Jegadeesh (1990), Lehmann (1990)) effects, a large portion of the ML-based forecasts' predictive power is unrelated to these phenomena. We next show that our ML-based forecasts have a positive but relatively weak relation with the image-based ML forecasts of Jiang et al. (2022) and that the predictive power of our forecast remains strong after controlling for the image-based forecasts. We also find that none of the 14 technical signals examined by Neely et al. (2014) or 14 trading friction variables examined by Freyberger et al. (2020) explain the relation between our ML-based forecasts and future stock returns. Finally, we visually examine plots associated with high and low future returns and find that even among stocks with similar values of momentum and reversal, there are differences in such charts that can easily be discerned by a human chart reader.

Our last tests examine the effectiveness of our optimization procedure. Specifically, we investigate whether ML models that performed well during the 192701-196306 optimization period continued to perform well during the 196307-202212 test period. To assess whether this is the case, we construct long-short portfolios by sorting stocks based on forecasts generated from each of the candidate ML models considered in our optimization exercise. We find that the correlation between values of the metric used to assess model performance in our optimization procedure and the associated Sharpe ratios of the long-short portfolios during the test period is 0.69. This strong positive correlation indicates that the optimization process successfully identified which ML models would produce better out-of-sample forecasts.

Our work contributes to three broad strands of research. First, we add to the literature examining whether past returns contain information useful for predicting future returns, which is tantamount to the literature on the weak form of the EMH. The most prominent findings in this literature are the aforementioned momentum and reversal effects. A subset of this literature explicitly investigates the validity of technical analysis and charting. Several papers examine the ability of technical signals to predict the performance of broad market indices or diversified portfolios. Brock et al. (1992) find that simple technical signals predict the future returns of the Dow Jones Index, but subsequent work attributes this finding to data snooping (Sullivan et al. (1999), Ready (2002), and Bajgrowicz and Scaillet (2012)) and nonsynchronous trading (Bessembinder and Chan (1998)). Allen and Karjalainen (1999) use genetic algorithms (a form of ML) to identify technical rules for trading the S&P 500 index and find that they do not work out of sample. More recently, Zhu and Zhou (2009) find that technical analysis can be useful for making asset allocation decisions, Moskowitz et al. (2012) find time-series momentum in a large number of asset classes, Neely et al. (2014) show that technical indicators are useful for predicting the market risk premium, and Han et al. (2013) find evidence that moving average strategies work well for timing investment in volatility-sorted portfolios.[3]

Despite the wide-spread use of technical analysis in practice (Menkhoff (2010), Lo and Hasanhodzic (2010)), research on the predictive power of technical signals is sparse. Most recently, Jiang et al.

[2] We rely on the machine learning literature (see e.g. Goodfellow et al. (2016)) to motivate our choices of hyperparameters, such as the number of layers in the neural network, the number of epochs used for early stopping to prevent overfitting, etc. The details of the configurations of our neutral networks are described in Section I and Figure A1 of the Internet Appendix.

[3] Sweeney (1986), Levich and Thomas (1993), Neely et al. (1997), Chang and Osler (1999), and Gehrig and Menkhoff (2006) find evidence that investment strategies based on technical analysis are profitable in foreign exchange markets, a finding that Osler (2003) attributes to the clustering of stop-loss and take-profit orders. LeBaron (1999) argues that the profitability of such strategies is due to central bank intervention in currency markets, but Neely (2002) concludes the opposite.

(2022) use ML to predict whether a stock will generate a positive or negative future return based on image representations of past price and volume data.[4] Their forecasts differ from ours in many ways. First, the input to their ML-based forecasts is an image of a chart that depicts daily open, close, high, and low prices, as well as trading volume and moving average prices over a past period of between five days and 3 months. Our approach is simpler in that we use only 12 past monthly cumulative return observations as the basis for our forecasts. Furthermore, Jiang et al. (2022) scale their images to reflect only the shape, but not the magnitude, of patterns in historical price and volume plots. We do not scale the data upon which our forecasts are based, thereby allowing our ML process to learn patterns based on both the shape and magnitude of past returns. Not surprisingly given the difference in our approaches, the predictive power of our ML-based forecasts is not subsumed by that of the forecasts generated by Jiang et al. (2022). Additionally, while Jiang et al. (2022) find that the predictive patterns are context-independent, meaning that the same patterns hold in different markets and using different time scales, we find that the predictive patterns we detect are persistent through time, suggesting that it is possible for a chartist to learn the patterns over a relatively long period. Our paper is most similar to Lo et al. (2000), who use smoothing estimators to extract nonlinear relations between historical price patterns and future stock returns. This finding has been the subject of much scrutiny, most notably by Jegadeesh (2000), who argues that Lo et al. (2000)'s findings are not robust to variation in the bandwidth parameter used for smoothing, a subjective empirical decision, and that the profitability of trading strategies based on the patterns detected by Lo et al. (2000) is close to zero. Our paper is similar in spirit to Lo et al. (2000), except we use a more modern technology (ML) than they do (smoothing estimators) to detect the relations between past price patterns and future returns. Our paper overcomes the robustness critique of Jegadeesh (2000) by implementing a systematic procedure to optimize our ML model using only data from the period prior to our main test period. Our use of portfolio analyses as our main empirical methodology addresses Jegadeesh (2000)'s profitability critique. The main contribution of our paper, therefore, is to demonstrate the merit of technical analysis, and thus provide evidence contradicting the EMH, in a manner that overcomes the critiques of previous work in this area.

Second, we add to the growing literature using ML to understand the cross-section of stock returns.[5] Messmer (2017), Messmer and Audrino (2020), and Freyberger et al. (2020) use ML to identify relations between stock-level characteristics and expected stock returns. Kelly et al. (2019), Gu et al. (2020, 2021), Kozak et al. (2020), Lettau and Pelger (2020), Bryzgalova et al. (2023), Chen et al. (2023), and Feng et al. (2023) use ML to extract latent factors, factor exposures, and risk premia from characteristics. Guijarro-Ordonez et al. (2023) use deep learning to construct optimal arbitrage portfolios.[6] A consistent theme in this work is the use of ML to synthesize the information in a broad set of stock-level variables that previous work has already found to be related to expected stock returns. Our work differs in that our forecasts are based only on 12 past monthly cumulative returns. The main benefit of ML in our setting is its ability to detect highly complex relations between past and expected future returns.[7] This contrasts with previous work that harnesses ML's ability to generate meaningful forecasts from a large number of predictive variables. Unlike the work that identifies factors and arbitrage portfolios, which uses ML to identify optimal tradeoffs between risk and expected return, our objective of assessing the efficacy of technical analysis for predicting future stock returns leads us to use ML to forecast only the expected return.

Finally, we contribute to the broader literature that uses ML for asset pricing applications by demonstrating the effectiveness of model selection via ex-ante optimization. Most research using ML to predict returns uses some form of model optimization procedure to select model parameters. Gu et al. (2020) compare the performance of different ML models in generating estimates of risk premia from a large number of stock-level characteristics but do not examine whether out-of-sample performance correlates with performance in the optimization process. Our finding of a strong correlation between optimization period and test period performance of different ML models shows that such optimization processes are effective.

The remainder of this paper is organized as follows. Section 2 describes the construction of our sample. Section 3 describes how we optimize our ML model and the calculation of the ML-based return forecasts. Section 4 presents our main evidence that the ML-based forecasts successfully predict the cross-section of future stock returns. Section 5 characterizes the nature of the predictive power of the ML-based forecasts. In Section 6 we conduct an evaluation of our optimization procedure. Section 7 concludes.

2. Sample and variables

In this section we describe our sample and the variables used in our tests.

2.1. Sample

The stock data used in our study come from CRSP. Our sample contains stock, month observations for months t from 192701-202212 (inclusive). In each month t, the sample contains all stocks that, on the last trading day of month $t-1$, are common shares of US-based firms listed on the NYSE, AMEX, or NASDAQ. To ensure that the construction of a one-year historical price plot for each stock in the sample would have been feasible, we require that each included stock has a non-missing return for each of months $t-12$ through $t-1$, inclusive. Finally, because our focal analyses weight stocks by market capitalization, we require that the market capitalization of each stock in the sample as of the end of month $t-1$, which we define as the number of shares outstanding times the price of each share, can be calculated. Our focal tests examine the ability of ML-based forecasts calculated as of the end of month $t-1$ to predict the cross-section of month t stock returns.

To ensure that the results of our focal tests reflect out-of-sample predictive power, we use sample months t from 192701-196306, inclusive, to optimize our ML model, and refer to this period as the "optimization period". The first month of the optimization period is 192701 because return data in CRSP begin in 192601, making 192701 the first month for which a full year of prior return data are available. Our focal asset pricing tests cover months t from 196307 through 202212, which we refer to as the "test period". We begin the test period in 196307 (and end the optimization period in 196306) to conform with the start date of the sample in Fama and French (1992, 1993) and several subsequent asset pricing studies.

[4] Moritz and Zimmermann (2016) apply tree-based conditional portfolio sorts to past return deciles and find that the associated forecasts predict the cross-section of future stock returns. Their use of past return deciles, instead of the actual cumulative returns, makes it difficult to reach a conclusion about the efficacy of traditional charting from their results.

[5] Earlier work includes Allen and Karjalainen (1999), who use genetic programming to search for profitable technical patterns in the S&P 500 index, and find none. In related work, Feng et al. (2018) and Rossi (2018) use ML to predict S&P 500 index returns, and Bianchi et al. (2021) use ML to predict bond returns.

[6] Rapach et al. (2013) use ML to generate forecasts of international stock returns based on past US market returns. Chinco et al. (2019) use ML to generate forecasts of one-minute-ahead future stock returns based on returns in the past three minutes. Bali et al. (2022) use ML to generate forecasts of corporate bond and stock returns.

[7] We find similar results when we using 60 months' past cumulative returns to generate the forecasts.

2.2. Variables

For each stock i in each month t, we calculate several variables. The focal dependent variable is the month t excess stock return, calculated as the delisting-adjusted stock return minus the return of the risk-free security in month t.[8] The independent variables, which are used to predict the cross-section of month t returns, are calculated from data available as of the end of month $t-1$. The focal independent variable is the ML-based return forecast, $MLER$. Section 3 describes in detail how we calculate $MLER$. For now, we simply note that the forecast of stock i's month t return is calculated by applying the function generated by the ML process to the cumulative returns of stock i over months $t-12$ through $t-1$, which we denote $CR_1, \ldots, CR_{12}$. Specifically, CR_k is the cumulative stock return over the k-month period covering months $t-12$ through $t-12+k-1$. This notation is motivated from the point of view of an investor looking at a one-year price plot created at the end of month $t-1$. CR_1 is the stock return in the first month that appears on the plot, and more generally CR_k is the cumulative return of the stock during the first k months that appear on the plot.

3. ML model and forecasts

In this section we describe how we select our ML model using data from the optimization period and how we use the selected ML model to generate forecasts of future stock returns.

3.1. Optimizing the ML model

Our objective in using ML is to generate a function f that, given values of input variables discernable from a historical price plot of a stock, will produce a forecast of the stock's future return. Very generally, the ML process takes a panel data set that includes one or more input variables, a dependent variable, and a weight variable, which combined we refer to as the fitting data, and attempts to "learn" the function f that describes the relation between the input variables and the expectation of the dependent variable. The ML process attempts to learn f by minimizing the value of a loss function $\mathcal{L}$. Once the function f has been learned, it can be applied to data not included in the fitting data to generate forecasts.

ML is a broad term that encapsulates several different empirical methods, including linear dimension reduction techniques such as principal components regression and partial least squares, nonlinear methods such as group LASSO that both perform dimension reduction and allow the forecast to be nonlinear in the inputs, nonparametric methods such as boosted regression trees and random forests that account for interactions between input variables, and neural networks, which accommodate most features of all other models and are therefore the most complex but most general and flexible learning tool.[9] We choose to use neural networks to generate our ML-based forecasts for two reasons. First, the learning process underlying neural networks is intended to mimic that of the human brain. We want our empirical design to be well-suited for capturing any predictive power discernable by the type of human chartist described in Lo et al. (2000), who learns complex patterns in stock returns from observation of historical price plots. Second, since our forecasts are based on only 12 input variables, the main benefit of ML in our context is its ability to account for both nonlinearities and interactions in the forecasting function. Methods whose main benefit is dimension reduction or variable selection, therefore, are likely to be less useful in our context.

As input variables, we use the monthly cumulative stock returns over the 12 months prior to the month whose return is to be forecasted, $CR_1, \ldots, CR_{12}$. While return predictability based on past returns of any frequency or horizon violates the EMH, we use one year of past monthly cumulative returns for several reasons. Our use of one year of past data follows previous technical analysis research, which typically uses at most one year of data to generate signals.[10] The use of a year's worth of past returns also ensures that our input variables contain the same information as the standard measure of momentum, which is the cumulative return over months $t-12$ through $t-2$. We use monthly (instead of daily or some other shorter frequency) cumulative returns to ensure that any patterns detected by our ML process could plausibly be detected by a human chartist, who may only be able to discern relatively coarse data from a one-year price plot. This decision follows previous ML research that uses data measured at the monthly frequency when calculating most historical return-based variables (see e.g. Neely et al. (2014), Gu et al. (2020), and Freyberger et al. (2020)). The use of monthly data also alleviates the need to address challenges associated with different years having different numbers of days when implementing our ML model. In Section III and Table A1 of the Internet Appendix, we show that the results of tests that use ML-based forecasts based on 60 months of past return data are qualitatively similar to those of our main measure. While our focal tests use cum dividend returns to generate forecasts of future returns, traditional chartists frequently work with price histories that do not account for the impact of dividends on prices. Therefore, in Section IV and Table A2 of the Internet Appendix, we show that our results hold when using price-based cumulative returns, instead of cum dividend cumulative returns, to forecast future returns.

The remaining decisions related to selection of our ML model are based on analysis of data from the optimization period, which covers sample months t from 192701-196306. We aim to find the combination of neural network architecture, loss function, loss function weight variable, and dependent variable that generates the best forecasts of future stock returns.

Following LeCun et al. (2015) and Goodfellow et al. (2016), we consider four neural network architectures: a feed-forward neural network (FNN), a convolutional neural network (CNN), a long short-term memory network (LSTM), and a convolutional neural network with long short-term memory (CNNLSTM). To avoid the curse of dimensionality, we rely on the ML literature to motivate our choices of hyperparameters, such as the number of layers in the neural network, etc. The neutral networks' configuration details, along with graphical depictions of each neural network architecture, are provided in Section I and Figure A1 of the Internet Appendix. All four architectures are widely used representation-based deep learning methods designed to learn complex relations in the data. Here, we give a very brief comparison of the different architectures. We refer the reader to LeCun et al. (2015) and Goodfellow et al. (2016) for more technical discussions.

FNN is the most general of the four architectures, but it does not explicitly consider the grid topology or variable dependency that frequently characterize sequential and time-series data. The CNN and LSTM architectures are designed to overcome this shortfall of FNN. CNN is a specialized type of neural network designed to process data with a grid-like topology, such as our cumulative return data, which can be interpreted as a one-dimensional grid. Previous work (Hoseinzade and Haratizadeh (2019)) has shown CNNs to be highly successful at time-series prediction. LSTM is a form of recurrent neural network (RNN). RNNs are explicitly designed to process one-dimensional sequential data. However, conventional RNNs have been shown to be

[8] The month t delisting-adjusted stock return is calculated following Shumway (1997). Details of the calculation are in Section II of the Internet Appendix. Monthly risk-free security returns are from Ken French's website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[9] Gu et al. (2020) provide an excellent overview of the different machine learning models.

[10] Brock et al. (1992), Zhu and Zhou (2009), and Han et al. (2013) consider moving average prices of up to 200 days, whereas Neely et al. (2014) use data from the past year to calculate many technical signals. Allen and Karjalainen (1999) use a 250 day moving average to scale S&P 500 index levels.

ineffective at discerning relations at lags greater than five (Goodfellow et al. (2016)), a drawback that LSTM overcomes. The CNNLSTM architecture combines the features of both the CNN and LSTM architectures.

The neural network learning process aims to minimize a loss function $\mathcal{L}$ that aggregates forecast errors through iterative, gradient-based optimizers (Goodfellow et al. (2016)). The two loss functions we consider are mean squared error (MSE, $\mathcal{L} = \sum w_j \epsilon_j^2$) and mean absolute error (MAE, $\mathcal{L} = \sum w_j |\epsilon_j|$), where w_j is the weight assigned to observation j in the training data, ϵ_j is the error, defined as the difference between the forecast value and the actual observed value of the dependent variable, and the summation is taken over all observations in the training data. Qualitatively, the main difference between MSE and MAE is that MSE is more sensitive to outliers than MAE.

The ML process uses panel data to determine the function f that best describes the relation between the input and dependent variables. However, our objective is to forecast the *cross-section* of future stock returns. As shown in Section V and Figure A2 of the Internet Appendix, the number of stocks in our sample varies substantially across months. Thus, giving the same weight in the loss function to each observation would cause months with more stocks to carry more total weight. To address this concern, in addition to the equal-weighting (EW) approach, we consider two alternative weighting methodologies. The first alternative is to equally-weight each month, and within each month, equally-weight each stock (EWPM for equal weight per month). The second alternative is to equally-weight each month, but consistent with our intent to use value-weighted portfolios as our primary test methodology, weight each stock within a given month according to its market capitalization at the end of the previous month (EWPMVW for equal weight per month value-weighted).

The use of panel data may also cause the learned function f to be determined, in part, by time-series relations between historical and future returns of the aggregate market. To illustrate, consider the price plots shown in Fig. 1, which depict the cumulative return of the market portfolio during the 12-month periods prior to 193304 (Panel A) and 193109 (Panel B). These are the one-year periods leading up to the highest (38.95% in 193304) and lowest (−29.10% in 193109), respectively, monthly market portfolio returns during our sample period. Given that there is strong covaration in returns of individual stocks, many stocks likely exhibited price patterns similar to those in the plots during the corresponding periods. The high (low) market portfolio return in the next month suggests that most stocks had a high (low) return during this month.[11] The ML algorithm is therefore likely to associate price patterns similar to that in Panel A (Panel B) with a high (low) future return.

To ensure that the ML process does not generate the function f based on this sort of aggregate time-series pattern, in addition to the excess stock return, r, we consider three transformations of r as potential dependent variables. The first is the standardized excess return, r_{Std}, calculated by taking the excess return, subtracting the mean excess return, and then dividing this difference by the standard deviation of the excess stock returns, with the mean and standard deviation calculated from the excess returns of all stocks in the given month. Using r_{Std} guarantees that the mean and standard deviation of the dependent variable in any given month are zero and one, respectively, thereby making it unlikely that the function generated by the ML process is based largely on aggregate time-series patterns.

A potential concern with both r and r_{Std}, however, is that the cross-sectional distribution of stock returns is known to be highly leptokurtic (see e.g. Bali et al. (2016, Table 7.2)), making it plausible that the function f generated by the ML process is largely a manifestation of extreme observations. Furthermore, our aim is to identify stocks that are likely

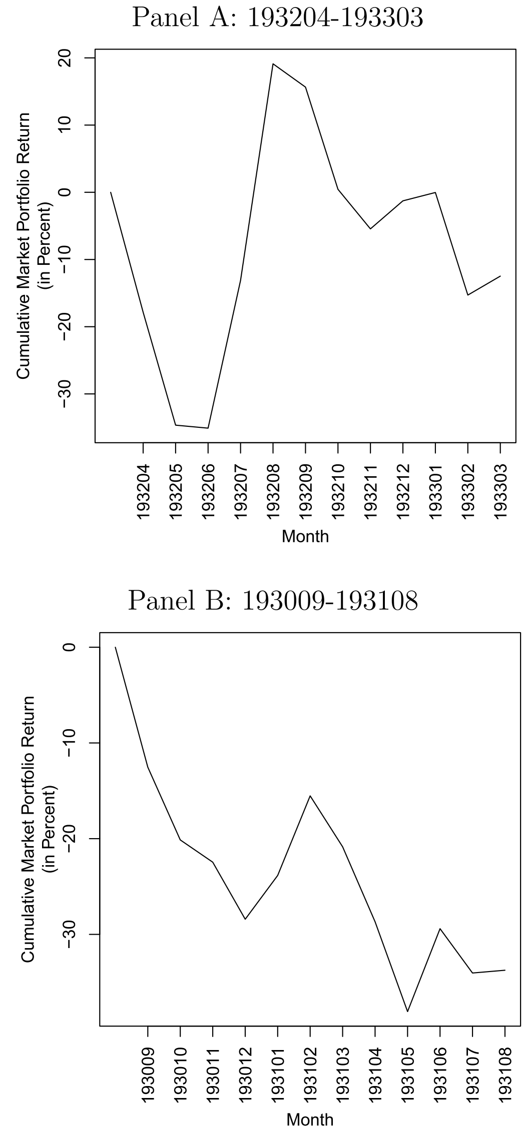

Fig. 1. Market Portfolio Price Plots. This figure plots the monthly cumulative returns of the market portfolio between 193204 and 193303 (Panel A) and between 193009 and 193108 (Panel B), inclusive.

to perform better (or worse) than other stocks, but not necessarily forecast the magnitude of the difference in returns. We therefore consider two additional transformations of the excess stock return. The first is the normalized excess return, which is constructed to ensure that values of the dependent variable in any given month have a standard normal distribution. Specifically, we define $r_{Norm,i,t} = \Theta^{-1}[rank_{i,t}/(N_t + 1)]$ where $\Theta^{-1}[\cdot]$ is the inverse standard normal cumulative distribution function, N_t is the number of stocks in our sample in month t, and $rank_{i,t}$ is the rank of stock i's month t return among the month t returns of all stocks in our sample, with the stock that has the lowest return having $rank_{i,t} = 1$ and the stock with the highest return having $rank_{i,t} = N_t$. The final candidate for the dependent variable is simply the percentile of the given stock's return among all stocks' returns in the given month, calculated as $r_{Pctl} = rank_{i,t}/(N_t + 1)$. Values of r_{Pctl} in any given month have a uniform distribution on the interval from zero to one.

While we consider several different transformations of the excess stock return as the dependent variable, we choose to not consider analogous transformations of the past cumulative returns used as input variables in our ML model. We make this choice to ensure that our empirical setup is consistent with the way a human chartist, as characterized in Lo et al. (2000), would make investment decisions.

[11] The equal-weighted average return of the stocks in our sample in 193304 (193109) is 52.48% (−31.47%), and 598 out of 631 (618 out of 627) such stocks had a positive (negative) return.

Table 1

ML Model Optimization. This table presents the results of tests used to determine the optimal ML model. The tests are run using only data from the 192701-196306 optimization period. We examine all combinations of four ML architectures, two loss functions, three loss function weighting methodologies, and four dependent variables. The ML architectures are feed-forward neural network (FNN), convolutional neural network (CNN), long short-term memory (LSTM), and convolutional neural network with long short-term memory (CNNLSTM). The loss functions are the mean squared error (MSE) and the mean absolute error (MAE). The loss function weighting methodologies are to equal-weight each observation (EW), to equal-weight each month and within each month give equal weight to each observation (EWPM), and to equal-weight each month and within each month weight each observation according to its market capitalization (EWPMVW). The four dependent variables we examine are the excess stock return (r), the standardized excess stock return (r_{Std}), the normalized excess stock return (r_{Norm}), and the percentile of the stock return (r_{Pctl}). Using each of the 96 possible combinations of implementation choices, we train each model 30 times using data from the even months in even years and odd months in odd years, to generate 30 forecasting functions. We then apply each of the forecasting functions to each observation in odd months in even years and even months in odd years, and for each such observation, take the average of the 30 resulting values to be the forecast. The table shows the time-series averages of the monthly cross-sectional Spearman rank correlations between the forecasts and the actual excess return. The column labeled "Dependent Variable" indicates the dependent variable. The column labeled "Weighting Methodology" indicates the weighting methodology. The remaining column headers indicate the ML architecture and loss function. The Spearman rank correlations are shown in percent.

Dependent Variable	Weighting Methodology	FNN MAE	FNN MSE	CNN MAE	CNN MSE	LSTM MAE	LSTM MSE	CNNLSTM MAE	CNNLSTM MSE
r	EW	4.6	1.6	4.0	1.0	5.5	−1.8	5.3	0.8
	EWPM	4.6	1.4	3.8	1.2	6.7	0.8	5.0	2.6
	EWPMVW	0.9	−0.5	0.5	−2.3	2.5	0.9	1.0	2.0
r_{Std}	EW	3.4	2.3	9.7	7.3	9.8	8.5	10.4	9.8
	EWPM	5.0	1.7	8.9	7.4	9.8	7.6	10.5	9.9
	EWPMVW	4.7	2.0	4.7	4.6	4.7	5.1	6.0	4.7
r_{Norm}	EW	7.5	1.5	9.7	8.8	10.3	10.2	10.4	10.7
	EWPM	6.9	1.4	9.1	8.0	10.1	10.2	10.6	10.8
	EWPMVW	3.7	0.4	4.3	4.3	5.2	6.3	6.9	7.5
r_{Pctl}	EW	−2.3	−1.6	7.7	−2.7	9.8	9.4	10.2	10.1
	EWPM	0.7	−3.0	7.8	−3.5	9.6	9.3	10.2	10.2
	EWPMVW	−0.1	−3.1	1.2	−1.7	4.9	5.7	7.3	7.8

Specifically, the chartist would observe a plot of past prices (not transformed prices) for different stocks and from those charts assess which stocks are likely to outperform. The commonality in individual stock returns described above may therefore cause the ML-based forecasts in any given month to be either predominantly high or predominantly low based on whether the market portfolio as a whole has a past return profile that leads to a high or low forecast, respectively. A benefit of our approach, however, is that it may help the learning process detect patterns that are distinct from traditional momentum and reversal signals, which are based on relative past performance. Since all of our tests are cross-sectional in nature, even in a month where all stocks have a high (or low) forecast future return, our tests will still assess whether the forecasts have the ability to discern which stocks are likely to have relatively high (or low) future returns.

In sum, we consider 96 different ML models, found by taking all combinations of four architectures (FNN, CNN, LSTM, CNNLSTM), two loss functions (MSE, MAE), three weighting schemes (EW, EWPM, EWPMVW), and four dependent variables (r, r_{Std}, r_{Norm}, r_{Pctl}).

To assess the effectiveness of each model, we first apply each model to a subset of the data in the 192701-196306 optimization period. Specifically, we fit each model 30 times on data from optimization period sample months t corresponding to even months from even years and odd months from odd years (fitting months hereafter). Each model is fit 30 times because the ML process is random. If we run the same model on the same data twice, the two resulting functions will *not* be the same. Each fitting uses 70% of the observations for training and the other 30% for validation. Because each fitting selects a different 30% of observations for validation, our methodology essentially uses cross validation. However, since the 30% of observations used for validation are randomly selected, they come from the same months as the observations used for training and may not be considered independent of the training month observations. For this reason, as described in the next paragraph, we evaluate the out-of-sample performance of each model using non-fitting month observations. Each of the 30 fittings produces a different forecasting function f. We then apply each of these 30 forecasting functions f to each non-fitting month observation in the optimization period. Finally, for each such observation, we take the average of those 30 forecasts to be the forecast based on the given ML model.[12] This approach of averaging several forecasts is referred to as "ensemble averaging."[13] In Section VI and Figure A3 of the Internet Appendix, we show that using an average of 30 forecasts removes almost all randomness from the ensemble forecast. Removing this randomness is important to ensure that our results can be replicated by future researchers. In the end, for each non-fitting month observation in the optimization period, we have 96 different ensemble forecasts, one corresponding to each different ML model.

We evaluate the models using the time-series average of the monthly cross-sectional Spearman rank correlations between forecasts and actual excess returns, calculated from non-fitting month observations. We use Spearman rank, instead of Pearson product-moment, correlations because our main tests are portfolio analyses that rely solely on the ordering of stocks with respect to the forecasts. Furthermore, because several of the forecasts are for transformed versions of the future excess stock return, the linearity assumption inherent in the Pearson correlation is unlikely to hold. Table 1 presents the average Spearman rank correlations for each model. The highest average correlation is generated by the CNNLSTM architecture using the MSE loss function with EWPM weights and r_{Norm} as the dependent variable. We therefore choose this model (CNNLSTM, MSE, EWPM, and r_{Norm}) to generate

[12] Krizhevsky et al. (2012), Sutskever et al. (2014), and Krogh and Vedelsby (1995) discuss the benefits of taking the average of multiple forecasts when using neural networks.

[13] Other papers that use ensemble averaging in asset pricing applications are Moritz and Zimmermann (2016), Gu et al. (2020, 2021), Bianchi et al. (2021), Bali et al. (2022), and Jiang et al. (2022).

Table 2
ML-Based Forecast Summary Statistics. This table presents summary statistics for the ML-based return forecasts, $MLER$. $MLER$ is a forecast of the normalized excess stock return. The column labeled "Period" indicates the period used to calculate the summary statistics in the given row. The columns labeled "Mean", "S.D.", "Min.", "P_1", "P_5", "P_{25}", "Median", "P_{75}", "P_{95}", "P_{99}", and "Max." present the time-series averages of the monthly cross-sectional mean, standard deviation, and the minimum, first percentile, fifth percentile, 25th percentile, median, 75th percentile, 95th percentile, 99th percentile, and maximum, respectively, values of $MLER$. The column labeled "n" shows the time-series average of the number of observations in each month. The set of stocks included in the month t sample is all common shares of US-based firms that are listed on the NYSE, AMEX, or NASDAQ as of the end of month $t-1$. We also require that a return is available for each stock in each of months $t-12$ through $t-1$, and that the stock's market capitalization as of the end of month $t-1$ is available. Values of $MLER$ are calculated as of the end of month $t-1$.

Period	Mean	S.D.	Min.	P_1	P_5	P_{25}	Median	P_{75}	P_{95}	P_{99}	Max.	n
196307-202212	−0.02	0.11	−0.61	−0.33	−0.22	−0.07	−0.00	0.05	0.12	0.19	0.38	4166
196307-197412	−0.01	0.09	−0.53	−0.28	−0.17	−0.06	−0.00	0.04	0.12	0.20	0.41	2245
197501-198412	−0.01	0.11	−0.75	−0.35	−0.21	−0.06	0.01	0.06	0.13	0.20	0.40	4332
198501-199412	−0.02	0.12	−0.66	−0.38	−0.24	−0.07	−0.00	0.05	0.15	0.26	0.57	5337
199501-200412	−0.04	0.12	−0.61	−0.36	−0.26	−0.11	−0.02	0.04	0.12	0.20	0.38	5792
200501-201412	−0.01	0.10	−0.53	−0.33	−0.21	−0.06	0.00	0.05	0.11	0.15	0.26	3956
201501-202212	−0.02	0.10	−0.57	−0.32	−0.22	−0.07	0.00	0.05	0.10	0.14	0.21	3486

our focal ML-based forecasts, $MLER$. Notably, small deviations from this model, such as using the MAE loss function, EW weights, or either r_{Std} or r_{Pctl} as the dependent variable, also perform well. This suggests that variation in performance across models is not spurious and that our optimization procedure is likely to have merit. Ex-post tests of the effectiveness of our optimization procedure, discussed in Section 6, confirm this hypothesis.

3.2. ML-based forecasts

We generate our focal ML-based forecasts, $MLER$, by applying the ML model to expanding-window past subsets of our sample, and using the functions generated by these expanding window fits to produce forecasts for subsequent periods.[14] In general, we define $MLER_{i,t}^{t_1,t_2}$ to be the forecast of the normalized excess return of stock i in month t that results from fitting our ML model on data from the period between months t_1 and t_2, inclusive. Specifically, for any t_1, $t_2 > t_1$, i, and t, we calculate $MLER_{i,t}^{t_1,t_2}$ as follows. First, we apply the ML process 30 times to observations for all stocks from sample months between t_1 and t_2, inclusive. The result is 30 different forecasting functions. We then apply each of these 30 functions to the cumulative returns of stock i in months $t-12$ through $t-1$. The result is 30 different forecasts of the normalized excess return of stock i in month t. We then take $MLER_{i,t}^{t_1,t_2}$ to be the average of these 30 forecasts. For example, to calculate $MLER_{X,200809}^{192701,196306}$, we first run the ML process 30 times using all observations in our sample from months t between 192701 and 196306, inclusive. We then apply each of the resulting forecasting functions to the monthly cumulative returns of stock X that would be observed in stock X's price plot covering the period from 200709 through 200808. The result of applying these functions to these cumulative returns is 30 forecasts of the normalized excess return of stock X in 200809. We then take $MLER_{X,200809}^{192701,196306}$ to be the average of these 30 forecast values.

The expanding windows to which we apply the ML process cover sample months t between 192701 and each of 196306, 197412, 198412, 199412, 200412, and 201412. As a result, we calculate $MLER_{i,t}^{192701,196306}$, $MLER_{i,t}^{192701,197412}$, $MLER_{i,t}^{192701,198412}$, $MLER_{i,t}^{192701,199412}$, $MLER_{i,t}^{192701,200412}$, and $MLER_{i,t}^{192701,201412}$ for each stock i and month t observation in our sample. The forecast normalized excess return of stock i in month t that we use in our focal empirical tests, $MLER_{i,t}$, is the forecast based on the most recent past execution of the fitting process. We therefore have:

$$MLER_{i,t} = \begin{cases} MLER_{i,t}^{192701,196306}, & \text{if } 196307 \leq t \leq 197412; \\ MLER_{i,t}^{192701,197412}, & \text{if } 197501 \leq t \leq 198412; \\ MLER_{i,t}^{192701,198412}, & \text{if } 198501 \leq t \leq 199412; \\ MLER_{i,t}^{192701,199412}, & \text{if } 199501 \leq t \leq 200412; \\ MLER_{i,t}^{192701,200412}, & \text{if } 200501 \leq t \leq 201412; \\ MLER_{i,t}^{192701,201412}, & \text{if } 201501 \leq t \leq 202212. \end{cases}$$

3.3. Summary statistics

Table 2 presents the time-series averages of monthly cross-sectional summary statistics for $MLER$ for the entire 196307-202212 test period and for subperiods of the test period. In interpreting the forecasts, recall that the forecasts are for the normalized future excess return, not the excess return itself. In the average month, $MLER$ has a mean of −0.02, a median very close to zero, and a cross-sectional standard deviation of 0.11. Extreme negative values of $MLER$ occur more frequently than extreme positive values, since the minimum, first percentile, and fifth percentile values are further below the mean than the 95th percentile, 99th percentile, and maximum values, respectively, are above it. The prevalence of large negative values of $MLER$ compared to large positive values should not impact most of our asset pricing tests, since we focus on portfolio analyses, which rely on the ordering (but not the magnitude or distribution) of $MLER$ across stocks. In the average test period month, our sample contains 4,166 stocks. The subperiod results indicate that the cross-sectional distribution of $MLER$ is highly consistent through time, since the salient characteristics of the cross-sectional distribution of $MLER$ do not change much across the different subperiods. The number of stocks in the sample ranges from an average of 2,245 stocks per month for the 196307-197412 subperiod to 5,792 stocks per month for the 199501-200412 subperiod.

4. Predictive power of ML-based forecasts

We turn now to our main objective, which is to investigate the relation between the ML-based return forecasts and the cross-section of future stock returns. The EMH predicts that $MLER$ should have no ability to predict cross-sectional variation in future stock returns.

[14] Similar to Gu et al. (2020), we use only past data to fit our model to ensure that our forecasts are based on patterns that could have been learned by the type of chartist described in Lo et al. (2000) and used for return prediction. Some papers (e.g. Kozak et al. (2020)) that describe the factor structure of returns with less of a focus on return prediction use primarily cross-validation to assess out-of-sample performance.

Table 3
Portfolio Analysis. This table presents the results of a portfolio analysis examining the ability of $MLER$ to predict the cross-section of future stock returns. At the end of each month $t-1$, all stocks in the month t sample are sorted into 10 portfolios based on an ascending ordering of $MLER$. The breakpoints used to determine which stocks are in which portfolio are the deciles of $MLER$ calculated using only stocks listed on the NYSE. The month t excess return of each portfolio is then taken to be the market capitalization-weighted average month t excess return of all stocks in the portfolio, with market capitalization calculated as of the end of month $t-1$. We also calculate the excess return of a zero-cost portfolio that is long portfolio 10 and short portfolio 1. The columns labeled "$MLER$ 1" through "$MLER$ 10" present results for decile portfolios 1 through 10. The column labeled "$MLER$ 10 − 1" presents results for the zero-cost long-short portfolio that is long the decile 10 portfolio and short the decile 1 portfolio. The rows labeled "$\bar{r}$" and "SD" present the time-series averages and standard deviations, respectively, of the monthly portfolio excess returns for each of the portfolios, reported in percent per month. The row labeled "Sharpe" presents the annualized Sharpe ratio of each portfolio. The values in parentheses are t-statistics, calculated following Newey and West (1987) using 12 lags, testing the null hypothesis that the average monthly excess return is equal to zero. The analysis covers return months t from July 1963 through December 2022, inclusive.

Value	$MLER$ 1	$MLER$ 2	$MLER$ 3	$MLER$ 4	$MLER$ 5	$MLER$ 6	$MLER$ 7	$MLER$ 8	$MLER$ 9	$MLER$ 10	$MLER$ 10 − 1
$\bar{r}$	−0.14	0.32	0.39	0.55	0.54	0.65	0.71	0.66	0.77	0.93	1.08
	(−0.55)	(1.39)	(1.89)	(2.77)	(2.91)	(3.94)	(4.10)	(3.99)	(4.36)	(5.18)	(5.51)
SD	6.46	5.66	5.26	5.11	4.71	4.63	4.50	4.38	4.47	5.03	4.79
Sharpe	−0.08	0.20	0.26	0.37	0.40	0.49	0.55	0.52	0.59	0.64	0.78

4.1. $MLER$-sorted portfolio returns

We test the EMH by examining the performance of portfolios formed by sorting stocks based on $MLER$. Each month t, we sort all stocks into decile portfolios based on an ascending ordering of $MLER$, which is calculated from data available at the end of month $t-1$. The breakpoints determining which stocks are in which portfolios are the deciles of $MLER$ calculated from the subset of stocks that are listed on the NYSE at the end of month $t-1$. The month t excess return of each portfolio is calculated as the weighted average excess return of all stocks in the portfolio, with weights proportional to market capitalization at the end of month $t-1$ (value-weighted hereafter).[15] We also calculate the excess return of a zero-cost portfolio that is long the decile 10 portfolio and short the decile one portfolio (10−1 portfolio hereafter). Our portfolio construction methodology follows Hou et al. (2020) and is intended to limit the impact of small-capitalization stocks on our analysis.

Table 3 presents the time-series averages of the monthly portfolio excess returns during the 196307-202212 test period. The average monthly excess returns increase nearly monotonically from −0.14% for decile portfolio one to 0.93% for portfolio 10. The average excess return of the $MLER$ 10 − 1 portfolio of 1.08% per month is highly statistically significant, with a Newey and West (1987) t-statistic of 5.51. This portfolio's annualized Sharpe ratio is 0.78, which is higher than that of the market factor (0.42), the factors in the Fama and French (2015) factor model (between 0.26 and 0.51), the momentum factor (0.53), and the reversal factor (0.51) during the same 196307-202212 period.[16] The patterns in average excess returns contradict the prediction of the EMH that technical analysis should be fruitless.

4.2. Risk-adjusted returns of $MLER$-sorted portfolios

A refined version of the EMH allows for profitable technical strategies if the associated average returns are compensation for risk. Asset pricing theory predicts that expected stock returns are a function of covariances between individual stock returns and innovations in priced risk factors, or betas. Stocks with similar betas, therefore, are likely to have both similar expected returns and similar past return patterns due to their covariances with factor innovations. However, by definition, factor innovations are unpredictable and serially uncorrelated, making it unlikely that covariances with factor innovations produce return patterns that repeat through time. For the ML-based forecasts to predict the cross-section of future stock returns, there must be repeated patterns in past returns that are informative about future expected returns. There is good economic reason, therefore, to think that the differences in average excess returns between the $MLER$-sorted portfolios are not a manifestation of exposure to systematic risk factors. Nonetheless, we investigate whether exposure to systematic risk factors can explain the performance of the $MLER$-sorted portfolios.

4.2.1. Full sample factor model regressions

We begin by using factor analysis to estimate the average risk-adjusted excess return (alpha) of the $MLER$ 10 − 1 portfolio.[17] The portfolio's alpha is calculated as the intercept coefficient from a time-series regression of excess portfolio returns on excess factor returns. A non-zero (zero) alpha indicates that exposures to the systematic risk factors captured by the factor model do not explain (explain) the portfolio's average excess return. Since the true factor model that prices securities is not known, we conduct the analyses using six previously-established empirical models: a one-factor market model (CAPM), the three-factor model of Fama and French (1993, FF), the four-factor model of Carhart (1997, FFC), the FFC model augmented with a short-term reversal factor (FFC+REV), the five-factor model of Fama and French (2015, FF5), and the four-factor model of Hou et al. (2015, Q).[18]

Table 4 Panel A shows that regardless of which model is used, the alpha of the $MLER$ 10 − 1 portfolio is positive, economically large, and statistically significant. Because the momentum and reversal factors are constructed by sorting stocks on variables calculated only from past returns, it is not surprising that the $MLER$ 10 − 1 portfolio's alpha is smaller when using the FFC and FFC+REV models than when using the CAPM, FF, and FF5 models. That the alpha with respect to the Q model is similar to that of the FFC model is consistent with Hou et al. (2015)'s finding that the Q model explains the momentum effect. However, even when using the FFC+REV model, which includes

[15] In Section VII and Table A3 (Section VIII and Table A4) of the Internet Appendix, we show that our results hold when using equal-weighted portfolios (breakpoints calculated from all stocks).

[16] Monthly excess factor returns are from Ken French's website. Factor Sharpe ratios are not tabulated.

[17] In Section IX and Table A5 of the Internet Appendix, we present the results of analyses examining risk-adjusted returns for each of the individual decile portfolios.

[18] Monthly excess returns for factors in the CAPM, FF, FFC, FFC+REV, and FF5 models are from Ken French's website. Monthly excess returns for the factors in the Q model are from Chen Xue's website: http://global-q.org/factors.html. Q model factor excess returns are available for the period from 196701-202112, thus all analyses using the Q factor model are subject to this data constraint.

Table 4

Risk-Adjusted Returns. This table presents the results of analyses examining whether the average returns of the portfolios formed by sorting on ML-based forecasts are compensation for risk. Panel A presents the results of factor model regressions of the excess returns of the $MLER$ 10 − 1 portfolio on the excess returns of the factors in different factor models. The $MLER$ 10 − 1 portfolio is the same as that whose average excess returns are shown in Table 3. The rows labeled "α", "$SD(\epsilon)$", and "Adj. R^2" present the intercept, residual standard deviation, and adjusted R^2, respectively, of the regression. The row labeled "Sharpe($\alpha + \epsilon$)" presents the annualized Sharpe ratio of the portfolio that is long the $MLER$ 10 − 1 portfolio and short the factor portfolios in amounts dictated by the slope coefficients estimated by the regression, calculated as the alpha divided by the residual standard deviation, times $\sqrt{12}$. Panel B presents results from 6-month non-overlapping subperiod factor model regressions estimated using daily excess return data. The row labeled "α" presents the average monthly alphas (intercept coefficients multiplied by 21) from these regressions. The row labeled "$SD(\epsilon)$" presents the standard deviation of the residuals from these regressions times $\sqrt{21}$. The standard deviation of the residuals from these regressions is calculated by using the residuals from the individual short subperiod regressions to generate a full time-series of daily residuals for the entire 196301-202212 period and calculating the standard deviation of this full time-series of residuals. The row labeled "Sharpe($\alpha + \epsilon$)" presents the annualized Sharpe ratio of the portfolio, calculated by dividing the average monthly alpha by the monthly standard deviation of the residuals and multiplying by $\sqrt{12}$. The row labeled "Adj. R^2" presents the average adjusted R^2 value from the short subperiod regressions. Except for the column labeled "Value", the column headers indicate the factor model used to generate the results in the given column. All excess returns, standard deviations, and alphas are reported in percent per month. The values in parentheses are t-statistics testing the null hypothesis that the average monthly alpha is equal to zero. The t-statistics in Panel A are calculated following Newey and West (1987) using 12 lags. The analyses cover return months t from July 1963 through December 2022, inclusive.

Panel A: $MLER$ 10 − 1 Portfolio Full Sample Factor Regressions

Value	CAPM	FF	FFC	FFC+REV	FF5	Q
α	1.20	1.24	0.79	0.45	1.02	0.71
	(6.43)	(7.28)	(5.23)	(3.58)	(6.67)	(3.78)
$SD(\epsilon)$	4.68	4.51	4.00	3.41	4.38	4.30
Sharpe($\alpha + \epsilon$)	0.89	0.95	0.68	0.46	0.81	0.57
Adj. R^2	4.18%	10.89%	29.90%	49.05%	15.90%	22.15%

Panel B: $MLER$ 10 − 1 Portfolio Average Short Subperiod Factor Regressions

Value	CAPM	FF	FFC	FFC+REV	FF5	Q
α	1.30	1.39	1.48	0.56	1.30	1.20
	(8.46)	(9.56)	(11.57)	(3.77)	(9.51)	(9.32)
$SD(\epsilon)$	4.27	3.75	3.45	3.24	3.57	3.60
Sharpe($\alpha + \epsilon$)	1.06	1.28	1.49	0.59	1.26	1.15
Adj. R^2	10.42%	25.22%	32.02%	37.36%	29.28%	28.74%

both the momentum and reversal factors, the $MLER$ 10 − 1 portfolio's alpha of 0.45% per month (t-statistic = 3.58) is economically large and highly significant. Furthermore, while the adjusted R^2 values from the FFC and FFC+REV model regressions of 29.90% and 49.05%, respectively, are higher than those of other models, even the model that includes both the momentum and reversal factors (FFC+REV) explains less than half of the $MLER$ 10 − 1 portfolio's variance. Thus, while there is overlap between the ML-based forecasts and each of momentum and reversal, the $MLER$ 10 − 1 portfolio's returns also contain a substantial component that is unrelated to these effects. We present further evidence supporting this conclusion in Section 5.3.

4.2.2. Short subperiod factor model regressions

A potential concern with the full-sample factor model regressions is that the betas of the $MLER$ 10 − 1 portfolio may be time-varying. Ehsani and Linnainmaa (2022) show that the momentum factor has time-varying factor betas. It is possible that the $MLER$ 10 − 1 portfolio has similar characteristics. If the factor exposures of the $MLER$ 10 − 1 portfolio are time-varying, then the alpha estimated from the full sample factor model regression may be a biased estimate of the portfolio's true alpha. To address this concern, we follow Lewellen and Nagel (2006) and run factor model regressions using daily data from non-overlapping short subperiods of our data, and examine the average alpha from these regressions. Specifically, for each 6-month subperiod during our 196307-202212 test period, we regress the daily excess returns of the $MLER$ 10−1 portfolio on the daily factor excess returns.[19] We multiply the intercept coefficient by 21 because there are approximately 21 days in the average month, and take this value to be the monthly alpha of the $MLER$ 10 − 1 portfolio during the given subperiod. The average alphas from these subperiod regressions, shown in Panel B of Table 4, are economically large and highly statistically significant for all factor models, ranging from 0.56% per month (t-statistic = 3.77) for the FFC+REV model to 1.48% per month (t-statistic = 11.57) for the FFC model. These results provide no indication that time-varying factor exposures explain the performance of the $MLER$ 10 − 1 portfolio.

4.2.3. Additional tests of risk-based explanation

We conduct two sets of additional analyses investigating a risk-based explanation for the differences in the average returns of the $MLER$ decile portfolios. First, we examine ex-ante hedged returns of

[19] To calculate daily excess returns of the $MLER$ 10 − 1 portfolio, we first calculate the daily excess returns of the individual $MLER$ decile portfolios. As in our focal analysis, the stocks held in each decile portfolio are updated monthly. The excess return of each decile portfolio on any day d is taken to be the value-weighted average excess stock return of all stocks in the portfolio on the given day, where the market capitalization values used for weighting are as of the close of the previous trading day. The $MLER$ 10 − 1 portfolio excess return is then taken to be the excess return of the $MLER$ 10 portfolio minus that of the $MLER$ 1 portfolio. The first 6-month subperiod covers 196307-196312 and the last 6-month subperiod covers 202207-202212.

Table 5

Portfolio Subperiod Analysis. This table describes the performance of the $MLER$ decile portfolios during different subperiods. The portfolios are the exact same portfolios whose performances are examined in Table 3. The column labeled "Subperiod" indicates the subperiod covered by each analysis. The columns labeled "$MLER$ 1" through "$MLER$ 10" and "$MLER$ 10 − 1" present results for the portfolio indicated in the column header. The rows labeled "$\overline{r}$" and "Sharpe" present the time-series averages of the monthly portfolio excess returns, reported in percent per month, and the annualized Sharpe ratios, respectively, for each of the portfolios in the given period. The values in parentheses are t-statistics, calculated following Newey and West (1987) using 12 lags, testing the null hypothesis that the average monthly excess return is equal to zero.

Subperiod	Value	$MLER$ 1	$MLER$ 2	$MLER$ 3	$MLER$ 4	$MLER$ 5	$MLER$ 6	$MLER$ 7	$MLER$ 8	$MLER$ 9	$MLER$ 10	$MLER$ 10 − 1
196307-197412	$\overline{r}$	−0.90	−0.47	−0.26	−0.14	−0.20	−0.03	0.15	0.07	−0.03	0.25	1.15
		(−1.39)	(−0.84)	(−0.49)	(−0.32)	(−0.50)	(−0.09)	(0.33)	(0.17)	(−0.07)	(0.62)	(3.10)
	Sharpe	−0.56	−0.32	−0.19	−0.10	−0.15	−0.03	0.12	0.06	−0.02	0.17	1.16
197501-198412	$\overline{r}$	0.16	0.27	0.42	0.65	0.65	0.66	0.85	0.81	1.03	1.47	1.32
		(0.31)	(0.58)	(0.89)	(1.38)	(1.46)	(1.60)	(1.90)	(2.00)	(2.28)	(3.13)	(4.41)
	Sharpe	0.09	0.17	0.30	0.47	0.50	0.50	0.63	0.62	0.73	0.93	1.15
198501-199412	$\overline{r}$	−0.17	0.49	0.44	0.63	0.85	0.78	0.79	0.84	0.98	1.00	1.17
		(−0.57)	(1.51)	(1.54)	(2.08)	(2.49)	(2.49)	(2.67)	(2.68)	(3.19)	(2.77)	(4.07)
	Sharpe	−0.13	0.37	0.34	0.51	0.65	0.59	0.59	0.63	0.71	0.67	1.16
1995012-20041	$\overline{r}$	−0.14	0.66	0.50	0.68	0.55	0.93	0.72	0.94	1.10	1.57	1.71
		(−0.18)	(1.15)	(0.96)	(1.17)	(1.14)	(2.28)	(1.52)	(2.60)	(2.51)	(4.06)	(3.21)
	Sharpe	−0.06	0.35	0.30	0.41	0.40	0.67	0.58	0.83	0.96	1.16	0.90
200501-201412	$\overline{r}$	0.48	0.49	0.91	0.62	0.59	0.97	0.88	0.64	0.56	0.40	−0.08
		(0.67)	(0.67)	(1.51)	(1.09)	(1.19)	(2.35)	(2.13)	(1.28)	(1.18)	(0.71)	(−0.15)
	Sharpe	0.24	0.27	0.57	0.42	0.45	0.76	0.72	0.52	0.46	0.27	−0.05
201501-202212	$\overline{r}$	−0.19	0.66	0.43	1.07	1.00	0.70	1.03	0.77	1.17	1.01	1.20
		(−0.23)	(0.99)	(0.71)	(1.94)	(1.83)	(1.40)	(2.16)	(2.04)	(2.99)	(3.16)	(2.13)
	Sharpe	−0.09	0.38	0.23	0.62	0.61	0.46	0.73	0.58	0.85	0.78	0.75

the $MLER$ portfolios. In these tests, the factor hedge ratios of the portfolios are estimated from regressions that use data from prior to portfolio formation. Because the hedge ratios are known at the time of portfolio formation and updated monthly, this approach ensures that the resulting alphas are those of a tradable portfolio while also accommodating time-varying betas. The results of these tests, described in Section X and Tables A6 and A7 of the Internet Appendix, indicate that the alpha of the $MLER$ 10 − 1 portfolio is positive, economically large, and highly statistically significant for all factor models.

Second, we examine the total risk of the $MLER$ decile portfolios. If the dispersion in average excess returns of the $MLER$ decile portfolios is due to risk, we would expect a positive, negative, negative, and negative relation between the average excess returns and the standard deviation, skewness, value at risk, and expected shortfall, respectively, of these portfolios. Analyses of the total risk of the $MLER$ decile portfolios, shown in Section XI and Table A8 of the Internet Appendix, find no evidence of these patterns.

In sum, our results suggest that the patterns in the average returns of the portfolios formed by sorting on the ML-based forecasts do not have a risk-based explanation.

4.3. Predictive power in subperiods

We next examine whether the predictive power of the ML-based forecasts is specific to certain subperiods. The cumulative performance of the $MLER$ 10 − 1 portfolio, plotted in Fig. 2, indicates that the predictive power of $MLER$ is stable through time. With the exception of large losses in 2009, the cumulative excess return of the $MLER$ 10 − 1 portfolio has a positive trajectory. In Section XII and Tables A9 and A10 of the Internet Appendix, we show that the large losses in 2009 are driven by an unusually large momentum tilt of the $MLER$ 10 − 1 portfolio, combined with large momentum losses, during 200904 and 200908.

Table 5 reports the average excess returns of the $MLER$-sorted portfolios during several subperiods. Consistent with Fig. 2, with the exception of the 200501-201412 subperiod, the $MLER$ 10 − 1 portfolio generates a large, positive, and statistically significant average excess return of more than 1% per month in each subperiod. During the most recent 201501-202212 subperiod, the average excess return of 1.20% per month is highly significant (t-statistic = 2.13). However, even excluding 200501-201412, the Sharpe ratio of the $MLER$ 10 − 1 portfolio is lower in later subperiods than in earlier subperiods. Thus, while there is evidence that the $MLER$ 10 − 1 portfolio continues to generate alpha during the most recent subperiod, there is also evidence that its performance is weakening through time.

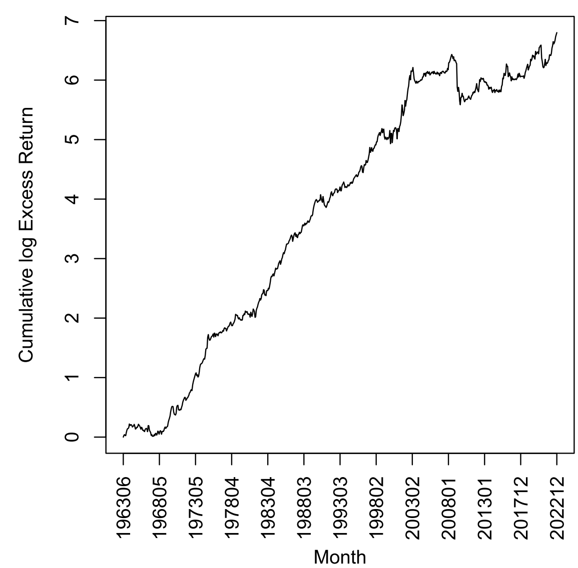

Fig. 2. Plot of Cumulative Performance of $MLER$ Long-Short Portfolio. This figure plots the cumulative log excess returns of the $MLER$ 10 − 1 portfolio. The cumulative log excess return of the $MLER$ 10 − 1 portfolio in any month t is taken to be the cumulative log return (excess return plus risk-free asset return) of the portfolio in that month minus the cumulative log risk-free asset return for the same month.

Table 6
Predictive Power Among Large Stocks. This table presents the results of portfolio analyses examining the ability of the ML-based forecasts to predict the cross-section of future stock returns among large stocks. The construction of the portfolios is exactly the same as was used to construct the portfolios whose performances are examined in Table 3, except that we vary the set of stocks used in the calculation of the portfolio breakpoints and the set of stocks that are sorted into the portfolios. The column labeled "Breakpoints" indicates the set of stocks used to calculate the breakpoints. The column labeled "Holdings" indicates the set of stocks sorted into the portfolios. The $Size > P_{20}^{\text{NYSE}}$ (NYSE/$Size > P_{20}^{\text{NYSE}}$) and $Size > P_{50}^{\text{NYSE}}$ (NYSE/$Size > P_{50}^{\text{NYSE}}$) subsets include all (NYSE-listed) stocks with market capitalizations greater than the 20th and 50th percentile values, respectively, among NYSE-listed stocks as of the end of month $t-1$. The Top 500 subset includes the top 500 stocks by market capitalization at the end of month $t-1$. The columns labeled "$MLER$ 1" through "$MLER$ 10" and "$MLER$ 10 − 1" present results for the portfolio indicated in the column header. The table shows the average excess return of each portfolio. All excess returns are reported in percent per month. The values in parentheses are t-statistics, calculated following Newey and West (1987) using 12 lags, testing the null hypothesis that the average monthly excess return is equal to zero. The analyses cover return months t from July 1963 through December 2022, inclusive.

Breakpoints	Holdings	$MLER$ 1	$MLER$ 2	$MLER$ 3	$MLER$ 4	$MLER$ 5	$MLER$ 6	$MLER$ 7	$MLER$ 8	$MLER$ 9	$MLER$ 10	$MLER$ 10 − 1
$Size > P_{20}^{\text{NYSE}}$	$Size > P_{20}^{\text{NYSE}}$	−0.13	0.34	0.39	0.57	0.53	0.60	0.75	0.62	0.75	0.92	1.06
		(−0.52)	(1.43)	(1.92)	(2.90)	(2.86)	(3.53)	(4.37)	(3.78)	(4.26)	(5.14)	(5.68)
NYSE/$Size > P_{20}^{\text{NYSE}}$	$Size > P_{20}^{\text{NYSE}}$	0.00	0.35	0.44	0.61	0.51	0.63	0.74	0.64	0.74	0.91	0.91
		(0.01)	(1.53)	(2.31)	(3.14)	(2.85)	(3.75)	(4.50)	(3.77)	(4.31)	(5.07)	(5.10)
$Size > P_{50}^{\text{NYSE}}$	$Size > P_{50}^{\text{NYSE}}$	0.04	0.33	0.48	0.52	0.52	0.61	0.64	0.62	0.77	0.85	0.81
		(0.17)	(1.52)	(2.34)	(2.78)	(2.92)	(3.68)	(3.72)	(3.70)	(4.52)	(4.60)	(4.63)
NYSE/$Size > P_{50}^{\text{NYSE}}$	$Size > P_{50}^{\text{NYSE}}$	0.10	0.37	0.46	0.61	0.55	0.62	0.66	0.66	0.75	0.87	0.77
		(0.42)	(1.80)	(2.27)	(3.38)	(3.21)	(3.63)	(3.97)	(3.91)	(4.37)	(4.82)	(4.66)
Top 500	Top 500	0.10	0.43	0.37	0.58	0.48	0.64	0.58	0.66	0.73	0.82	0.72
		(0.41)	(1.93)	(1.88)	(3.10)	(2.95)	(3.82)	(3.26)	(3.91)	(4.14)	(4.56)	(4.37)

4.4. *Predictive power for large stocks*

Recent work shows that many return anomalies are concentrated among small stocks (Fama and French (2008, 2018), Hou et al. (2020)) and that trading costs make such anomalies unprofitable to trade (Novy-Marx and Velikov (2016), Detzel et al. (2023)). The portfolios examined to this point are value-weighted and constructed using breakpoints calculated from only NYSE-listed stocks. Hou et al. (2020) show that this approach minimizes the influence of small stocks on the results of the analyses. Nonetheless, to ensure that our results persist when small stocks are excluded, we repeat the portfolio analysis using subsets of our sample that exclude small stocks from both the set of stocks held in the portfolios and the set of stocks used to calculate the breakpoints. We define the $Size > P_{20}^{\text{NYSE}}$ (NYSE/$Size > P_{20}^{\text{NYSE}}$) and $Size > P_{50}^{\text{NYSE}}$ (NYSE/$Size > P_{50}^{\text{NYSE}}$) subsets to be the sets of (NYSE-listed) stocks with market capitalizations greater than the 20th and 50th percentile values, respectively, among NYSE-listed stocks, and the Top 500 subset to include only the top 500 stocks by market capitalization. Each of the market capitalization-based criteria are evaluated monthly using all stocks in the sample for the given month. We continue to use value-weighted portfolios for all of these tests.

The results of these analyses are shown in Table 6. When we use the $Size > P_{20}^{\text{NYSE}}$ subset both to calculate the breakpoints and as the set of stocks held in the portfolios, the $MLER$ 10 − 1 portfolio generates an average excess return of 1.06% per month (t-statistic = 5.68). This result is extremely similar to that of our main test whose results are shown in Table 3. Indeed, the correlation between the excess returns of these two portfolios is 0.96 (untabulated). These results strongly suggest that the influence of small stocks on our focal tests is minimal. As we impose more stringent restrictions on the size of the stocks used in the analyses, the average excess returns of the $MLER$ 10 − 1 portfolio and associated t-statistics become slightly smaller. However, even when we use only the largest 500 stocks both to calculate the breakpoints and as the set of stocks held in the portfolios, the average excess return of the $MLER$ 10 − 1 portfolio of 0.72% per month (t-statistic = 4.37) remains economically large and highly statistically significant. Furthermore, regardless of the set of stocks used to calculate the breakpoints or held in the portfolios, the decile one (10) portfolio generates the lowest (highest) average excess return of all portfolios. These tests show that the predictive power of the ML-based forecasts is strong among large stocks.

5. Characterization of ML-based forecasts

Having demonstrated that the ML-based forecasts have strong ability to predict the cross-section of future stock returns, we proceed now to further characterize this predictive power.

5.1. *Stability of forecasting relation*

Our first tests characterizing the ML-based forecasts examine the stability of the relation between past price patterns and future stock returns (forecasting relation hereafter). Specifically, we investigate whether price patterns associated with high (low) future stock returns in the early part of our test period continue to be associated with high (low) stock returns for the duration of our test period. If the forecasting relation is stable, these relations could be learned over a prolonged period of time, thereby further suggesting the viability of charting.

To assess the stability of the forecasting relation, we define $MLER_{i,t}^{192701,196306}$, $MLER_{i,t}^{196307,197412}$, $MLER_{i,t}^{197501,198412}$, $MLER_{i,t}^{198501,199412}$, $MLER_{i,t}^{199501,200412}$, and $MLER_{i,t}^{200501,201412}$ to be the forecasts for stock i in month t based on the forecasting function generated by running the ML process on data from the months between and inclusive of those in the superscripts. If the forecasting functions learned using data from two subperiods are similar and both predict the cross-section of returns, it would indicate stability in the forecasting relation.

We test whether the forecasting functions learned using data from two different subperiods are similar in three ways. First, we calculate the monthly cross-sectional Spearman rank correlations between the forecasts for all months in the test period that are not included in the data used to generate either forecasting function. For example, we calculate the correlations between $MLER^{196307,197412}$ and $MLER^{199501,200412}$ using forecasts from all months in 196307-202212 excluding 196307-197412 and 199501-200412. The time-series averages of these correlations, shown in Table 7 under "Forecast Correlations", are all positive and of substantial magnitude, but decrease as the time between the fitting periods increases.

Table 7

Stability of Forecasting Relation. This table presents the results of analyses examining the stability of the forecasting relation. We define $MLER^{t_1,t_2}$ as the ML-based forecast generated by applying the ML process to data from months t between t_1 and t_2, inclusive. The columns under the "Forecast Correlations" heading present the time-series averages of monthly cross-sectional Spearman rank correlations between pairs of forecasts. The columns under the "$10-1$ Common Holdings" heading present the time-series averages of the monthly percentages, in decimal format (i.e. $0.01 = 1\%$) of holdings in the $10-1$ portfolios that are common to both portfolios. The percentage of common holdings in any pair of $10-1$ portfolios in any month is the number of stocks held in the same direction (both long or both short) in both portfolios divided by the average of the number of stocks in the portfolios. The columns under the "$10-1$ Return Correlations" heading present the time-series Pearson product-moment correlations between the excess returns of the $10-1$ portfolios. With the exception of the sort variable, the $10-1$ portfolios are constructed in exactly the same manner as those whose returns are examined in Table 3. All analyses use data from all months in the 196307-202212 test period except those months whose data are used by the ML process to generate the forecasting function upon which either forecast is based.

	Forecast Correlations					$10-1$ Common Holdings					$10-1$ Return Correlations				
	$MLER^{196307,197412}$	$MLER^{197501,198412}$	$MLER^{198501,199412}$	$MLER^{199501,200412}$	$MLER^{200501,201412}$	$MLER^{196307,197412}$	$MLER^{197501,198412}$	$MLER^{198501,199412}$	$MLER^{199501,200412}$	$MLER^{200501,201412}$	$MLER^{196307,197412}$	$MLER^{197501,198412}$	$MLER^{198501,199412}$	$MLER^{199501,200412}$	$MLER^{200501,201412}$
$MLER^{192701,196306}$	0.62	0.56	0.50	0.49	0.31	0.50	0.48	0.41	0.41	0.30	0.60	0.54	0.37	0.41	0.30
$MLER^{196307,197412}$		0.74	0.73	0.71	0.57		0.60	0.58	0.57	0.45		0.84	0.79	0.80	0.66
$MLER^{197501,198412}$			0.80	0.78	0.68			0.64	0.62	0.52			0.84	0.88	0.81
$MLER^{198501,199412}$				0.84	0.77				0.65	0.58				0.91	0.85
$MLER^{199501,200412}$					0.75					0.56					0.83

Second, using each forecast as the sort variable, we construct monthly decile portfolios using the exact same process as was used to construct the portfolios examined in section 4. For each pair of forecasts, we then calculate the percentage of stocks that are common to both $10-1$ portfolios. For any pair of portfolios in any given month, this percentage is calculated as the number of stocks that are held in the same direction in both portfolios, i.e. long in both portfolios or short in both portfolios, divided by the average number of stocks in the two portfolios. In Table 7 under the "$10-1$ Common Holdings" heading, we present the time-series averages of these percentages for each pair of portfolios, once again using only months during the test period that are not used in either fit. Consistent with the patterns in the forecast correlations, there is substantial commonality in the holdings of all pairs of $10-1$ portfolios, but this commonality decreases as the time between the fitting periods used by the ML process to generate the forecasting functions increases.

Third, we calculate the Pearson product-moment correlations between the monthly excess returns of each pair of $10-1$ portfolios using only excess returns from months in the test period not used in the fitting process upon which either forecast is based. These correlations, shown under "$10-1$ Return Correlations" in Table 7, are once again large for all pairs of forecasts, but decrease as the time between the fitting periods increases.

The results of all three of these analyses suggest that while the forecasting function changes through time, this change is slow. Furthermore, the similarities between the forecasts generated by applying the ML process to data from 192701-196306 and 200501-201412 indicate that there is a substantial time-invariant component of the forecasting function.

To test whether the ML-based forecasts generated by fitting data from each given subperiod predict the cross-section of returns in different subperiods, in Table 8 we present the average excess returns of the $10-1$ portfolios formed by sorting on each of $MLER_{i,t}^{192701,196306}$, $MLER_{i,t}^{196307,197412}$, $MLER_{i,t}^{197501,198412}$, $MLER_{i,t}^{198501,199412}$, $MLER_{i,t}^{199501,200412}$, and $MLER_{i,t}^{200501,201412}$ in all subperiods excluding that used by the ML process to learn the forecasting function. While analyses that use a forecast based on a fit period subsequent to the period whose returns are examined are not reflective of obtainable trading profits, such analyses are valid statistical tests of the stability of the forecasting relation. The results show that the $10-1$ portfolios formed by sorting on each of the ML-based forecasts generate economically large and in most cases statistically significant average excess returns in all subperiods except for 200501-201412, during which, as shown in Section 4.3, the $MLER$ $10-1$ portfolio performs poorly. Notably, the $MLER^{192701,196306}$ $10-1$ portfolio earns an average excess return of 0.90% per month (t-statistic $= 2.38$) during the 201501-202212 subperiod. Thus, forecasts based on a fit of the data ending in 196306 contain strong predictive power more than 50 years later. Taken together, the

Table 8

Subperiod-Based Forecasts and Returns. This table presents the results of portfolio analyses examining the ability of ML-based forecasts generated by applying the ML process to data from one subperiod to predict the cross-section of stock returns during other subperiods. The column labeled "Return Period" indicates the period for which the excess portfolio returns are examined. The remaining column names indicate the variable used to construct the portfolios. $MLER^{t_1,t_2}$ is the ML-based forecast generated by applying the ML process to data from months t between t_1 and t_2, inclusive. With the exception of the sort variable, the portfolio construction methodology is identical to that used to construct the portfolios examined in Table 3. The table presents the average monthly excess return of the $10-1$ portfolio formed by sorting on the given variable during the given subperiod. All excess returns are reported in percent per month. The values in parentheses are t-statistics, calculated following Newey and West (1987) using 12 lags, testing the null hypothesis that the average monthly excess return is equal to zero.

Return Period	$MLER^{192701,196306}$	$MLER^{196307,197412}$	$MLER^{197501,198412}$	$MLER^{198501,199412}$	$MLER^{199501,200412}$	$MLER^{200501,201412}$
196307-197412	1.15		1.35	1.46	1.43	0.82
	(3.93)		(3.49)	(3.57)	(3.62)	(1.85)
197501-198412	1.23	1.43		1.19	1.32	0.58
	(3.01)	(4.09)		(3.18)	(3.47)	(1.41)
198501-199412	0.86	1.27	1.28		1.21	1.05
	(2.94)	(3.79)	(3.83)		(3.16)	(2.57)
199501-200412	1.57	1.44	1.08	1.31		0.52
	(3.12)	(2.42)	(1.44)	(1.95)		(0.62)
200501-201412	0.02	−0.18	0.02	0.20	−0.02	
	(0.06)	(−0.45)	(0.05)	(0.35)	(−0.04)	
201501-202212	0.90	0.88	1.03	1.38	1.27	0.51
	(2.38)	(1.73)	(1.73)	(2.00)	(1.82)	(0.70)

Table 9

Expanding-Window and Rolling-Window Forecast Portfolios. This table presents the results of portfolio analyses examining the ability of ML-based forecasts based on expanding window ($MLER$) and rolling-window ($MLER^{Roll}$) fit periods to predict the cross-section of future stock returns. With the exception of the sort variable for the portfolios formed by sorting on $MLER^{Roll}$, the portfolio construction methodology is identical to that used to construct the portfolios examined in Table 3. The column labeled "Sort Variable" indicates the variable used to sort stocks into portfolios. The columns labeled "1" through "10" present results for decile portfolios 1 through 10 formed by sorting on the given variable. The column labeled "10 − 1" presents results for the zero-cost long-short portfolio that is long the decile 10 portfolio and short the decile 1 portfolio. The rows with "$\bar{r}$" and "SD" in the column labeled "Value" present the time-series averages and standard deviations, respectively, of the monthly portfolio excess returns for each of the portfolios, reported in percent per month. The rows with "Sharpe" in the "Value" column present the annualized Sharpe ratio of the given portfolio. The values in parentheses are t-statistics, calculated following Newey and West (1987) using 12 lags, testing the null hypothesis that the average monthly excess return is equal to zero. The analyses cover return months t from July 1963 through December 2022, inclusive.

Sort Variable	Value	1	2	3	4	5	6	7	8	9	10	10 − 1
$MLER$	$\bar{r}$	−0.14	0.32	0.39	0.55	0.54	0.65	0.71	0.66	0.77	0.93	1.08
		(−0.55)	(1.39)	(1.89)	(2.77)	(2.91)	(3.94)	(4.10)	(3.99)	(4.36)	(5.18)	(5.51)
	SD	6.46	5.66	5.26	5.11	4.71	4.63	4.50	4.38	4.47	5.03	4.79
	Sharpe	−0.08	0.20	0.26	0.37	0.40	0.49	0.55	0.52	0.59	0.64	0.78
$MLER^{Roll}$	$\bar{r}$	−0.13	0.30	0.43	0.56	0.60	0.63	0.72	0.65	0.76	0.84	0.96
		(−0.43)	(1.21)	(1.83)	(2.68)	(3.32)	(3.46)	(4.43)	(4.01)	(4.60)	(4.96)	(4.24)
	SD	7.17	6.40	5.76	5.26	4.89	4.66	4.50	4.45	4.28	4.61	5.37
	Sharpe	−0.06	0.16	0.26	0.37	0.43	0.47	0.55	0.50	0.61	0.63	0.62

results in Tables 7 and 8 indicate a high degree of stability in the forecasting function, although it may also contain a slowly time-varying component.

The combination of time-invariant and slowly time-varying components of the forecasting relation suggests that there is a tradeoff to be made when deciding what data to use in the ML process. The time-invariant component will be better estimated when more data are used in the learning process. The time-varying component will be better estimated by using only more recent data. We investigate this tradeoff by comparing the performance of portfolios formed by sorting on our focal forecasts ($MLER$) which use expanding-window fit periods, and forecasts based on forecasting functions generated from rolling-window fit periods ($MLER^{Roll}$). $MLER^{Roll}$ for stock i in month t is defined as $MLER^{192701,196306}$ for months t in 196307-197412, $MLER^{196307,197412}$ for months t in 197501-198412, $MLER^{197501,198412}$ for months t in 198501-199412, $MLER^{198501,199412}$ for months t in 199501-200412, $MLER^{199501,200412}$ for months t in 200501-201412, and $MLER^{200501,201412}$ for months t in 201501-202212. If the tradeoff between the use of more data from a prolonged period and less data from a shorter but more recent time period favors the former, then the $MLER$ $10-1$ portfolio should outperform the $MLER^{Roll}$ $10-1$ portfolio, and vice versa.

Table 9 shows that the average monthly excess return and Sharpe ratio of the $MLER^{Roll}$ $10-1$ portfolio of 0.96% (t-statistic = 4.24) and 0.62, respectively, are lower, but only by a small amount, than the corresponding values of 1.08% (t-statistic = 5.51) and 0.78, respectively, for the $MLER$ $10-1$ portfolio. The results suggest that the tradeoff between using more data from a longer time period and less data from a shorter time period in the ML process favors, but only slightly, the use of more data from a longer time period.

5.2. Nonlinearity and interactions in the forecasting function

The results in Section 5.1 demonstrate that the forecasting relation is highly stable through time, but gives no indication of its functional form. The main benefit of using neural networks to generate forecasts is that they are capable of discerning highly-complex relations that may include nonlinear components and components based on interactions of the input variables. The challenge that comes with this benefit is that the complexities of the forecasting function are difficult, if not impossible, to succinctly describe. If the nonlinear and interaction components are important for prediction, it would indicate that traditional techniques such as linear regression, as well as ML methodologies that do not capture interaction effects, are suboptimal for return forecasting. If nonlinearities are important but interactions are unimportant, it would indicate that the components of the forecasting function relevant for prediction can be described as the summation of components related to each of the historical cumulative returns, which would substantially facilitate description of the forecasts. If neither nonlinearities nor interactions are important, then the components of the forecasting function that are relevant for cross-sectional prediction are simply the slope coefficients associated with each past return. Our next tests, therefore, examine the prevalence of nonlinear and interaction components in the forecasting function, and the degree to which such components are important for predicting the cross-section of future stock returns.

We begin by graphically examining the forecasting function for evidence of nonlinearities and interactions. Fig. 3 presents univariate partial-dependence plots of the relation between $MLER$ and CR_k, for $k \in \{1, 2, \ldots, 12\}$, based on a fit of the data from 192701-196306. The plotted values of $MLER$ are calculated by setting CR_j for each stock and each $j \in \{1, 2, \ldots, 12\}$, $j \neq k$, to its mean among all observations in the sample from 196307-202212. We focus on plots based on the fit using data ending in 196306 because these data are included in each of the expanding windows used to calculate $MLER$, and thus affect the forecasting function for all fit periods.[20] The plots indicate that the forecasting function is highly nonlinear, and in most cases nonmonotonic, in the input variables. The plots for $CR_2, \ldots, CR_{10}$ are all highly nonmonotonic, and several of the plots indicate multiple sign changes for the slope of the relation between the given input variable and the forecast. Furthermore, the plots show that these instances of nonmonotonicity are not isolated to extreme values of the focal input variable, where the forecasting function may be unreliable due to a small number of observations with similar values in the data used in the learning process.

To graphically illustrate interactions in the forecasting function, Fig. 4 presents examples of bivariate partial-dependence plots depicting the relation between $MLER$ and CR_k calculated using the 10th, 30th, 50th, 70th, and 90th percentile values of one other input variable, CR_j, with CR_l, $l \in 1, 2, \ldots, 12$, $l \neq j$ and $l \neq k$ set to its mean, using the forecasting function based on a fit of the data from 192701-196306. The percentiles and mean values are based on all observations from 196307-

[20] In Section XIII and Figure A4 of the Internet Appendix, we present similar plots for other fit periods.

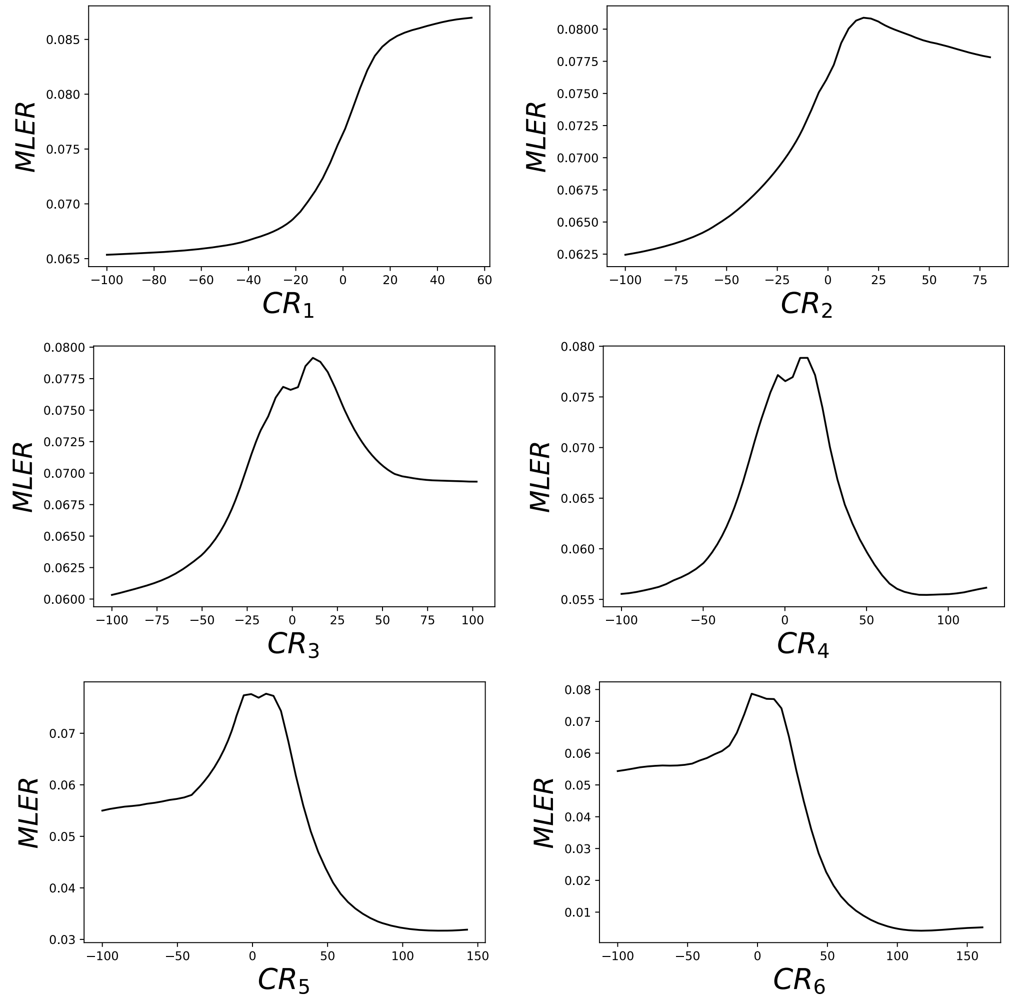

Fig. 3. Univariate Partial-Dependence Plots for ML-Based Forecasts. This figure presents partial dependence plots for the ML-based forecasting function. The function whose partial dependence plots are shown is based on a fit of the data from 192701 through 196306. Each plot depicts the relation between one of $CR_1, \ldots, CR_{12}$, indicated on the x-axis, and the value of $MLER$, indicated on the y-axis, with all other input variables set to their sample means based on all observations in the sample from 196307 through 202212. The range of values included on the x-axis in each plot is -100 to the 99th percentile value of the variable shown on the x-axis, calculated using observations in the sample from 196307 through 202212.

202212. If there are no interactions between CR_j and CR_k, the lines corresponding to the different percentiles of CR_j will be parallel.[21]

The left panel in Fig. 4 indicates strong interactions between CR_4 and CR_{12} in the forecasting function. For high values of CR_{12} (the 70th and 90th percentiles), $MLER$ is decreasing in CR_4 for values of CR_4 below and slightly above zero, and then increasing for higher values of CR_4. When CR_{12} is taken at its 30th or 50th percentile values, the slopes of $MLER$ take the opposite signs. The middle panel illustrates substantial interactions between CR_9 and CR_{10}. When CR_{10} is taken at its 30th (70th) percentile value, for values of CR_9 between approximately -50 and 0, $MLER$ is decreasing (increasing). Finally, the right panel shows an example where interactions are much less severe. The slopes of the relation between $MLER$ and CR_{10} are quite similar for all depicted percentiles of CR_4.

To formally assess the amount of cross-sectional variation in $MLER$ that is driven by the nonlinear and interaction components, we conduct a Fama and MacBeth (1973, FM hereafter) regression analysis of the relation between $MLER$ and $CR_1, \ldots, CR_{12}$. Specifically, each month t, we run a cross-sectional OLS regression of $MLER$ on $CR_1, \ldots, CR_{12}$. Panel A of Table 10 presents the time-series averages of the monthly cross-sectional coefficients and adjusted R^2 values. The average adjusted R^2 of these regressions is 37.55%, indicating that only slightly more than a third of the total variation in $MLER$ is explained by a linear combination of $CR_1, \ldots, CR_{12}$, and therefore that nonlinear and interaction components of the forecasting function account for most of the cross-sectional variation in the forecasts.

To test the importance of the nonlinearities and interactions for predicting future stock returns, we once again conduct FM regressions, this time using the future excess stock return as the dependent variable and $MLER$ and $CR_1, \ldots, CR_{12}$ as independent variables. The slope coefficient on $MLER$ measures the relation between future stock returns and the component of $MLER$ that is linearly orthogonal to the other independent variables, namely $CR_1, \ldots, CR_{12}$, and thus can be used to assess the importance of the nonlinear and interaction components of $MLER$ in prediction. We note, however, that our use of linear regression here assumes a linear relation between the component of $MLER$ that is orthogonal to $CR_1, \ldots, CR_{12}$ and future stock returns. Since our ML model uses the normalized excess stock return as the dependent variable, there is no strong reason to believe that the relation between this orthogonal component of $MLER$ and future (not normalized) excess stock returns is linear. While this may make economic interpretation of the magnitude of the slope coefficient difficult, a statistically significant coefficient on $MLER$ would be strong evidence that

[21] Plots for all pairs of input variables CR_j and CR_k are presented in Section XIV and Figure A5 of the Internet Appendix.

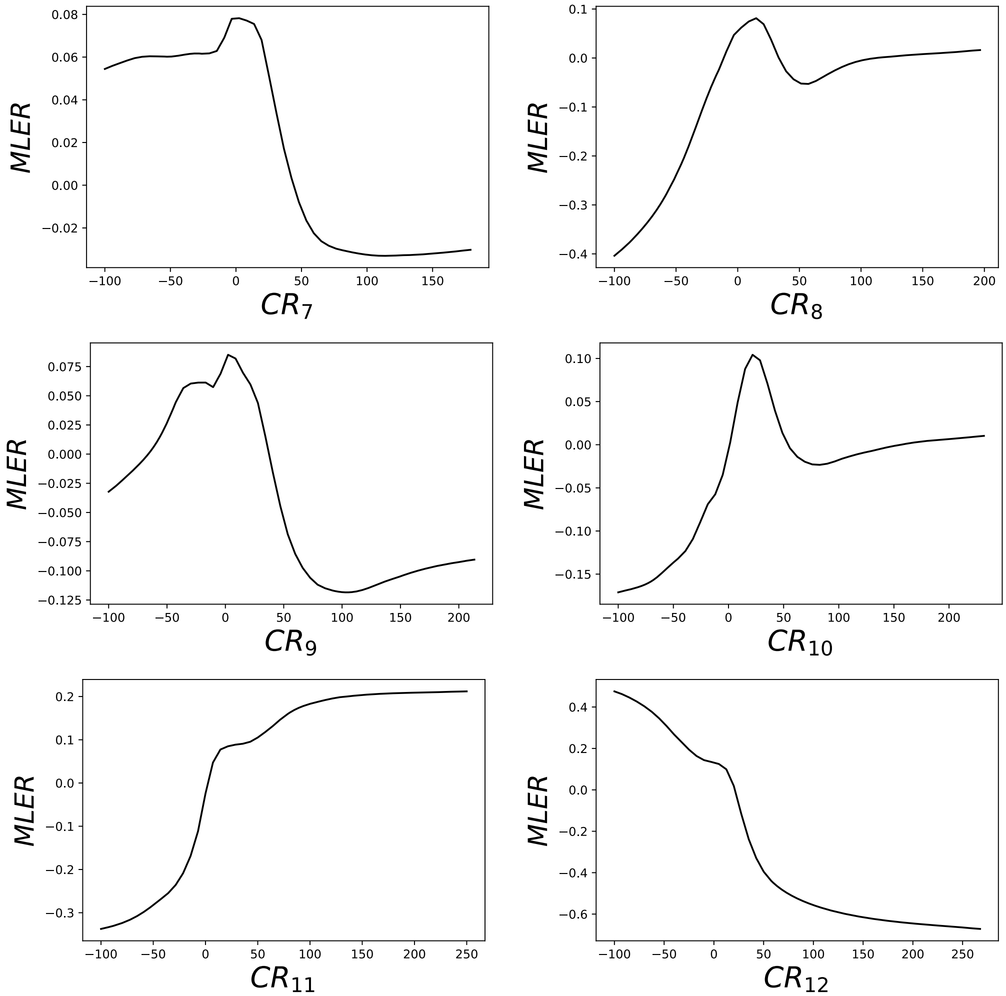

Fig. 3. (*continued*)

the nonlinear and/or interaction components of the ML-based forecasts are important for prediction. In addition to equal-weighted (i.e. OLS) regressions, for consistency with our value-weighted portfolio analyses, we also run the FM regression analysis using value-weighted (i.e. WLS) regressions with market capitalization as the weight variable. The results in Panel B of Table 10 show that the average coefficient on $MLER$ of 5.95 (t-statistic = 9.06) in the equal-weighted regressions and 2.81 (t-statistic = 5.47) in the value-weighted regressions are both highly statistically significant, indicating that nonlinearities and/or interactions in the forecasting function are important for future stock return prediction.

We next endeavor to discern between nonlinear and interaction components of the ML-based forecasts. To investigate the amount of cross-sectional variation in $MLER$ that is associated with the interaction component of the forecasts, we run a FM regression analysis using $MLER$ as the dependent variable and $CR_1, \ldots, CR_{12}$ as well as terms that capture the nonlinear components of the forecasts as independent variables. To capture the nonlinear components, we define $MLER_{CR_k}$ to be the value of the forecasting function evaluated using the true value of CR_k with values of CR_j, for $j \in 1, \ldots, 12$ and $j \neq k$, set to their mean values among all observations in the given month.[22] $MLER_{CR_k}$ captures the nonlinear component of the forecasting function with respect to CR_k. The component of $MLER$ that cannot be captured by a linear combination of the CR_k and $MLER_{CR_k}$, $k \in 1, 2, \ldots, 12$, therefore, must be driven by interactions in the forecasting function. Panel C of Table 10 shows that the average adjusted R^2 from these cross-sectional regressions is 55.36%, which is substantially higher than the 37.55% adjusted R^2 from the regressions that used only $CR_1, \ldots, CR_{12}$ as independent variables, but still far from a perfect 100%. This indicates that while a non-trivial portion of variation in $MLER$ is driven by the linear and nonlinear components of the forecasting function, there is also a substantial component, represented by the unexplained variation in the regression, that is attributable to interaction effects.

To investigate the importance of the interaction component of the ML-based forecasts for predicting the cross-section of future stock returns, we run equal-weighted and value-weighted FM regression analyses using the future excess stock return as the dependent variable and $CR_1, \ldots, CR_{12}$, $MLER_{CR_1}, \ldots, MLER_{CR_{12}}$, and $MLER$ as independent variables. In these regressions, the coefficient on $MLER$ measures the relation between the interaction component of the ML-based forecasts and future stock returns. Panel D of Table 10 shows that for both equal-weighted and value-weighted regressions, the average slope coefficient of $MLER$ is positive and highly significant, indicating that the interaction components of the forecasting function play an important role in prediction.

[22] If the forecasting function has no interaction components, then the choice of values for CR_j will have no impact on our results. Nonetheless, in Section XV and Table A11 of the Internet Appendix, we conduct similar analyses using zero, instead of the monthly mean values, for the CR_j when calculating $MLER_{CR_k}$, and find that this has no qualitative impact on our results.

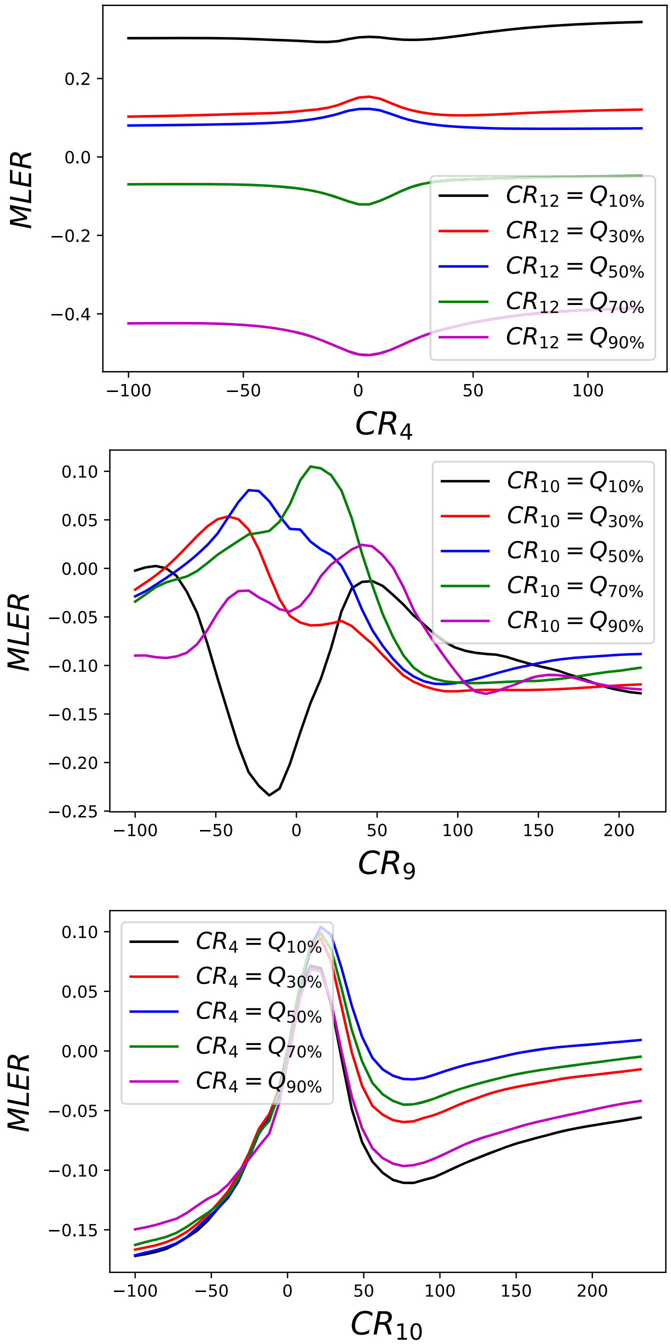

Fig. 4. Bivariate Partial-Dependence Plots for ML-Based Forecasts. This figure presents examples of bivariate partial dependence plots for the ML-based forecasting function. The plots depict the relation between CR_j (plotted on the x-axis) and $MLER$ (plotted on the y-axis) when CR_k is taken to be its 10th (black line), 30th (red line), 50th (blue line), 70th (green line), and 90th (purple line) percentile values, and CR_l, $l \in \{1, 2, \dots, 12\}$, $l \neq j$, and $l \neq k$, is taken to be its mean value calculated using all observations in the sample from 196307 through 202212. The plot on the top (middle) [bottom] takes $j = 4$ and $k = 12$ ($j = 9$ and $k = 10$) [$j = 10$ and $k = 4$]. The function whose partial dependence plots are shown is based on a fit of the data from 192701 through 196306. The range of values included on the x-axis is -100 to the 99th percentile value of CR_j, calculated using observations in the sample from 196307 through 202212. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

In sum, Figs. 3 and 4 and the results in Table 10 demonstrate that the ML-based forecasting function has substantial nonlinear and interaction components that play an important role in the ability of $MLER$ to predict the cross-section of future stock returns.

5.3. Momentum and reversal

Our next tests examine the extent to which the predictive power of the ML-based forecasts is related to the momentum and reversal effects. Our measure of momentum, Mom, is the cumulative stock return during the 11-month period covering months $t-12$ through $t-2$, inclusive (skipping month $t-1$). We measure reversal with Rev, defined as the stock return in month $t-1$. While the asset pricing literature has documented hundreds of variables that predict future stock returns, these two effects are of particular relevance here because both Mom and Rev can be discerned from historical price plots.[23] In fact, both Mom and Rev are very simple functions of the cumulative returns $CR_1, \dots, CR_{12}$ that are the inputs to the forecasting function. Specifically, $Mom = CR_{11}$ and $Rev = 100[(CR_{12}/100+1)/(CR_{11}/100+1)-1]$.[24] It is therefore highly plausible that the predictive power of $MLER$ is, at least in part, driven by the predictive power of Mom and Rev.

In addition to Mom and Rev, we calculate ML-based forecasts, which we denote $MLER^{Mom,Rev}$, that use only Mom and Rev as input variables. The ML model used to calculate $MLER^{Mom,Rev}$ is identical to that used to calculate $MLER$, except for the use of Mom and Rev, instead of $CR_1, \dots, CR_{12}$, as input variables. Our objective in calculating $MLER^{Mom,Rev}$ is to capture any nonlinearities and interaction effects in the predictive power of Mom and Rev. In Section XVI and Table A12 of the Internet Appendix, we construct decile portfolios by sorting on each of Mom, Rev, and $MLER^{Mom,Rev}$, and find that all three variables predict the cross-section of future stock returns in our test period sample.

5.3.1. Relations between ML-based forecasts, momentum, and reversal

We examine the strength of the relations between the ML-based forecasts, momentum, and reversal by running FM regression analyses with $MLER$ as the dependent variable and combinations of Mom, Rev, and $MLER^{Mom,Rev}$ as independent variables. Table 11 shows that when Mom (Rev) is the only independent variable, the average adjusted R^2 from the cross-sectional regressions is 10.14% (26.68%). When $MLER^{Mom,Rev}$ is the only independent variable, the average adjusted R^2 increases to 55.01%, indicating that while $MLER^{Mom,Rev}$ and $MLER$ have substantial commonality, a large component of $MLER$ is unrelated to $MLER^{Mom,Rev}$. When Mom, Rev, and $MLER^{Mom,Rev}$ are all included as independent variables, the average adjusted R^2 of 57.20% is only slightly higher than when $MLER^{Mom,Rev}$ is the only independent variable. This suggests that $MLER^{Mom,Rev}$ captures most of the commonality between $MLER$ and each of Mom and Rev. More importantly, the results indicate that, as expected, $MLER$ has components related to both momentum and reversal. However, $MLER$ also has a very substantial component that is unrelated to these effects.

5.3.2. Predictive power controlling for momentum and reversal

To investigate whether the predictive power of $MLER$ is, at least in part, distinct from momentum and reversal, we examine the performance of portfolios with similar levels of the control variable(s), Mom, Rev, and $MLER^{Mom,Rev}$, but different levels of $MLER$. We construct both bivariate portfolios, which control for Mom, Rev, and $MLER^{Mom,Rev}$ one at a time, and trivariate portfolios that simultaneously control for Mom and Rev. If the predictive power of $MLER$ is unexplained (explained) by the control variable(s), we expect to find a strong pattern (no pattern) in average excess returns across portfolios containing stocks with different levels of $MLER$ but similar levels of the control variable(s).

[23] See Hou et al. (2015), Harvey et al. (2016), McLean and Pontiff (2016), and Linnainmaa and Roberts (2018) for lists of variables that have been documented to predict the cross-section of future stock returns.

[24] The multiplication and divisions by 100 are because Mom, Rev, CR_{11}, and CR_{12} are recorded in percent.

Table 10
Nonlinearities and Interactions in the Forecasting Function. This table presents the results of Fama and MacBeth (1973) regression analyses examining nonlinearity and interactions of the ML-based forecasts $MLER$. Each month t we run a cross-sectional regression of the dependent variable on one or more independent variables. The results in Panel A are for regressions with $MLER$ as the dependent variable and $CR_1, \ldots, CR_{12}$ as independent variables. The results in Panel B are for regressions with the future excess stock return as the dependent variable and $MLER$ and $CR_1, \ldots, CR_{12}$ as independent variables. The results in Panel C are for regressions with $MLER$ as the dependent variable and $CR_1, \ldots, CR_{12}$, and $MLER_{CR_1}, \ldots, MLER_{CR_{12}}$ as independent variables. The results in Panel D are for regressions with the future excess stock return as the dependent variable and $MLER$, $CR_1, \ldots, CR_{12}$, and $MLER_{CR_1}, \ldots, MLER_{CR_{12}}$ as independent variables. $MLER_{CR_k}$ is the value of the ML-based forecast calculated using CR_k and the within-month means of CR_j for $j \in \{1, \ldots, 12\}$ and $j \neq k$, as the values of the input variables, and therefore captures the component of the ML-based forecast that is nonlinear in CR_k. In Panels B and D, the sections with "EW" and "VW" in the "Weight" column present results for equal-weighted and market capitalization-weighted regressions, respectively. The table presents the time-series averages of the monthly cross-sectional coefficients, along with t-statistics, calculated following Newey and West (1987) using 12 lags, testing the null hypothesis that the average coefficient is equal to zero (in parentheses). The column labeled "Adj. R^2" in Panels A and C presents the time-series average of adjusted R^2 values from the monthly cross-sectional regressions. In Panels A and C, we report the estimated slope coefficients times 100. The analyses cover sample months t from July 1963 through December 2022, inclusive.

Panel A: FM Regressions of $MLER$ on $CR_1, \ldots, CR_{12}$

CR_1	CR_2	CR_3	CR_4	CR_5	CR_6	
0.029	0.022	0.006	0.009	0.004	0.002	
(11.81)	(13.19)	(2.82)	(6.39)	(2.59)	(1.12)	
CR_7	CR_8	CR_9	CR_{10}	CR_{11}	CR_{12}	Adj. R^2
0.012	0.033	−0.065	0.131	0.251	−0.276	37.55%
(3.31)	(9.36)	(−11.18)	(16.94)	(17.00)	(−15.52)	

Panel B: FM Regressions of Future Excess Returns on $MLER$ and $CR_1, \ldots, CR_{12}$

Weight	$MLER$	CR_1	CR_2	CR_3	CR_4	CR_5	CR_6
EW	5.95	0.01	0.00	−0.00	0.00	−0.00	0.00
	(9.06)	(2.21)	(0.76)	(−1.89)	(1.69)	(−1.62)	(0.13)
		CR_7	CR_8	CR_9	CR_{10}	CR_{11}	CR_{12}
		−0.00	−0.00	0.00	−0.00	0.02	−0.01
		(−0.90)	(−0.75)	(0.50)	(−0.67)	(4.31)	(−2.52)
	$MLER$	CR_1	CR_2	CR_3	CR_4	CR_5	CR_6
VW	2.81	0.00	0.00	−0.01	0.02	−0.00	−0.01
	(5.47)	(0.96)	(1.08)	(−3.66)	(4.23)	(−0.13)	(−2.97)
		CR_7	CR_8	CR_9	CR_{10}	CR_{11}	CR_{12}
		0.01	0.00	−0.00	0.00	0.02	−0.01
		(2.43)	(0.04)	(−1.19)	(0.54)	(3.84)	(−2.43)

Panel C: FM Regressions of $MLER$ on $CR_1, \ldots, CR_{12}$, and $MLER_{CR_1}, \ldots, MLER_{CR_{12}}$

CR_1	CR_2	CR_3	CR_4	CR_5	CR_6	
−0.009	0.009	0.001	0.008	0.003	−0.000	
(−2.61)	(5.04)	(0.30)	(6.56)	(2.09)	(−0.15)	
CR_7	CR_8	CR_9	CR_{10}	CR_{11}	CR_{12}	
0.011	0.029	−0.080	0.131	0.197	−0.235	
(3.29)	(8.14)	(−13.47)	(16.48)	(16.04)	(−19.62)	
$MLER_{CR_1}$	$MLER_{CR_2}$	$MLER_{CR_3}$	$MLER_{CR_4}$	$MLER_{CR_5}$	$MLER_{CR_6}$	
282.503	89.942	51.901	31.352	23.228	7.921	
(11.97)	(8.20)	(7.20)	(6.76)	(8.76)	(4.05)	
$MLER_{CR_7}$	$MLER_{CR_8}$	$MLER_{CR_9}$	$MLER_{CR_{10}}$	$MLER_{CR_{11}}$	$MLER_{CR_{12}}$	Adj. R^2
18.418	20.680	−18.183	1.162	30.575	15.374	55.36%
(5.39)	(14.28)	(−11.48)	(0.86)	(22.06)	(8.78)	

(*continued on next page*)

To construct bivariate portfolios, each month t we sort stocks into five groups based on an ascending ordering of the control variable. We then sort all stocks within each control variable-based group into five portfolios based on an ascending ordering of $MLER$. We take the month t excess return for each of the resulting 25 portfolios to be the value-weighted average month t excess return of all stocks in the portfolio. To create a single portfolio for each $MLER$ quintile, we define the month t excess return of the bivariate $MLER$ quintile k portfolio to be the equal-weighted average of the month t excess returns of the five $MLER$ quintile k portfolios (one for each control variable quintile). Finally, we define the month t bivariate $MLER$ $5-1$ portfolio excess return to be the month t excess return of the bivariate $MLER$ quintile five portfolio minus that of the bivariate $MLER$ quintile one portfolio.

We construct the trivariate portfolios in a similar manner, except before sorting on $MLER$ we sort on both Mom and Rev. Specifically, each month t we sort the stocks in our sample into 25 groups based on the intersections of five Mom groups and five Rev groups con-

Table 10 (*continued*)

Panel D: FM Regressions of Future Excess Returns on $MLER$, $CR_1, \ldots, CR_{12}$ and $MLER_{CR_1}, \ldots, MLER_{CR_{12}}$

Weight	Slope Coefficients						
EW	$MLER$	CR_1	CR_2	CR_3	CR_4	CR_5	CR_6
	5.94	−0.00	0.00	−0.00	0.00	−0.00	0.00
	(10.29)	(−0.21)	(0.05)	(−2.09)	(1.12)	(−1.82)	(0.63)
		CR_7	CR_8	CR_9	CR_{10}	CR_{11}	CR_{12}
		−0.00	−0.00	0.00	−0.00	0.02	−0.01
		(−1.34)	(−0.50)	(0.37)	(−0.89)	(4.27)	(−2.63)
		$MLER_{CR_1}$	$MLER_{CR_2}$	$MLER_{CR_3}$	$MLER_{CR_4}$	$MLER_{CR_5}$	$MLER_{CR_6}$
		61.58	12.01	0.37	4.53	−8.79	4.35
		(1.68)	(1.00)	(0.04)	(0.96)	(−2.00)	(1.55)
		$MLER_{CR_7}$	$MLER_{CR_8}$	$MLER_{CR_9}$	$MLER_{CR_{10}}$	$MLER_{CR_{11}}$	$MLER_{CR_{12}}$
		−1.17	−0.86	−0.15	−0.16	0.52	−0.00
		(−0.56)	(−1.14)	(−0.22)	(−0.26)	(1.28)	(−0.00)
VW	$MLER$	CR_1	CR_2	CR_3	CR_4	CR_5	CR_6
	3.08	0.00	0.00	−0.01	0.01	−0.00	−0.01
	(6.47)	(0.69)	(0.66)	(−3.05)	(3.84)	(−0.37)	(−2.86)
		CR_7	CR_8	CR_9	CR_{10}	CR_{11}	CR_{12}
		0.01	−0.00	−0.00	0.00	0.02	−0.01
		(2.12)	(−0.24)	(−0.38)	(0.51)	(3.79)	(−1.73)
		$MLER_{CR_1}$	$MLER_{CR_2}$	$MLER_{CR_3}$	$MLER_{CR_4}$	$MLER_{CR_5}$	$MLER_{CR_6}$
		31.43	1.42	−4.79	−1.14	−2.45	0.82
		(0.69)	(0.07)	(−0.44)	(−0.17)	(−0.40)	(0.20)
		$MLER_{CR_7}$	$MLER_{CR_8}$	$MLER_{CR_9}$	$MLER_{CR_{10}}$	$MLER_{CR_{11}}$	$MLER_{CR_{12}}$
		−0.04	−0.07	0.75	0.16	−0.11	0.87
		(−0.02)	(−0.07)	(0.72)	(0.18)	(−0.20)	(1.61)

Table 11

Relations between ML-Based Forecasts, Momentum, and Reversal. This table presents the results of Fama and MacBeth (1973) regression analyses examining the ability of Mom, Rev, and $MLER^{Mom,Rev}$ to explain variation in $MLER$. The process for conducting the Fama and MacBeth (1973) regression analyses is exactly as described in Table 10. The dependent variable in the regressions is $MLER$ and the independent variables are combinations of Mom, Rev, and $MLER^{Mom,Rev}$. Results for different regression specifications are shown in different columns of the table. The table presents the time-series averages of monthly cross-sectional regression coefficients, along with t-statistics, calculated following Newey and West (1987) using 12 lags, testing the null hypothesis that the average coefficient is equal to zero (in parentheses). The rows labeled "Mom", "Rev", and "$MLER^{Mom,Rev}$" present results pertaining to the coefficient on the indicated variable. The row labeled "Adj. R^2" shows the average adjusted R^2 from the cross-sectional regressions. The analyses cover sample months t from July 1963 through December 2022, inclusive.

	(1)	(2)	(3)	(4)
Mom	0.001			0.000
	(31.50)			(10.93)
Rev		−0.004		−0.000
		(−72.27)		(−7.81)
$MLER^{Mom,Rev}$			0.878	0.815
			(207.32)	(102.38)
Adj. R^2	10.14%	26.68%	55.01%	57.20%

structed from independent sorts on Mom and Rev. Using stocks within each of the 25 groups, we then sort stocks into five $MLER$ portfolios. The result is 125 portfolios, for which we calculate the month t value-weighted excess returns. We then define the excess return of the trivariate quintile $MLER$ quintile k portfolio to be the equal-weighted average of the month t excess returns of the 25 portfolios (one for each intersection of the Mom and Rev groups) corresponding to the kth $MLER$ quintile. Finally, we define month t trivariate $MLER$ $5-1$ portfolio excess return to be the month t excess return of the trivariate $MLER$ quintile five portfolio minus that of the trivariate $MLER$ quintile one portfolio.

We use quintile portfolios, instead of decile portfolios, for these analyses to ensure that each portfolio is populated in each month. To enable comparison with a benchmark that does not control for either Mom or Rev, we also construct univariate $MLER$ quintile portfolios using exactly the same procedure as was used to construct the $MLER$ decile portfolios examined in Section 4, except with five, instead of 10, portfolios. The values of Mom, Rev, and $MLER$ used to construct the portfolios are measured as of the end of month $t-1$ and the breakpoints used in all sorts are calculated using only NYSE-listed stocks.

The average monthly univariate, bivariate, and trivariate $MLER$ quintile portfolio excess returns are shown in Table 12. The univariate $MLER$ $5-1$ portfolio generates an average excess return of 0.72% per month (t-statistic $= 4.64$) and the average excess returns of the individual quintile portfolios monotonically increase across the $MLER$ quintiles. The average excess return of the bivariate $MLER$ $5-1$ portfolio constructed to control for Mom is 0.72% per month (t-statistic $=$ 6.28), which is similar to that of the univariate $MLER$ $5-1$ portfolio. The reason for this similarity is that, as discussed in Section XII of the Internet Appendix, controlling for momentum causes the bivariate $MLER$ $5-1$ portfolio to avoid some large losses realized by the univariate $MLER$ $5-1$ portfolio that are associated with the momentum effect. The bivariate $MLER$ $5-1$ portfolio that controls for Rev has an average excess return of 0.61% per month (t-statistic $= 3.79$). When $MLER^{Mom,Rev}$ is used as the control variable, the $MLER$ $5-1$ portfolio generates 0.44% per month (t-statistic $= 3.99$). This is substantially less than the 0.72% generated by the univarate $MLER$ $5-1$ portfolio, but still economically large and highly statistically significant.

Finally, the trivariate $MLER$ $5-1$ portfolio has an average excess return of 0.40% per month (t-statistic $= 3.79$). The similarity of the average excess returns of the bivariate portfolio analysis that controls for

Table 12

Multivariate Portfolio Analysis - Control for Momentum and Reversal. This table presents the results of univariate, bivariate, and trivariate portfolio analyses examining the ability of the ML-based forecasts to predict the cross-section of future stock returns after controlling for momentum and reversal. The procedure used to generate the univariate portfolios is identical to that used to generate the portfolios in Table 3, except we create five quintile portfolios instead of 10 decile portfolios. The bivariate portfolios are constructed by sorting all stocks into quintiles based on a control variable, either Mom, Rev, or $MLER^{Mom,Rev}$, and then within each control variable quintile, into five $MLER$ portfolios. The trivariate portfolios are formed by independently sorting stocks into quintiles of Mom and Rev, and then sorting stocks in each of the 25 groups formed by the intersections of the Mom and Rev quintiles, into $MLER$ quintiles. Breakpoints for all sorts are calculated using only NYSE-listed stocks. Values of $MLER$, Mom, Rev, and $MLER^{Mom,Rev}$ are calculated as of the end of month $t-1$. The month t excess return of each of the resulting portfolios is taken to be the market capitalization-weighted average month t excess return of all stocks in the given portfolio, with market capitalization calculated as of the end of month $t-1$. For the bivariate (trivariate) portfolios, the month t excess return of the $MLER$ quintile k portfolio is taken to be the equal-weighted average, across all quintiles of the control variable (across all 25 Mom and Rev groups), of the $MLER$ quintile k portfolio. Finally, the $MLER$ $5-1$ portfolio excess return is taken to be the difference between the $MLER$ quintile five and quintile one portfolio excess returns. The table presents the time-series averages of the monthly portfolio excess returns for each of the $MLER$ quintile portfolios. The column labeled “Control Variable(s)” indicates the control variable(s) used to construct the portfolios. The columns labeled “$MLER$ 1” through “$MLER$ 5” and “$MLER$ $5-1$” present the average monthly excess portfolio returns along with t-statistics, calculated following Newey and West (1987) using 12 lags, testing the null hypothesis that the average monthly excess return of the given portfolio is equal to zero. All excess returns are reported in percent per month. The analyses cover return months t from July 1963 through December 2022, inclusive.

Control Variable(s)	$MLER$ 1	$MLER$ 2	$MLER$ 3	$MLER$ 4	$MLER$ 5	$MLER$ $5-1$
None	0.12	0.47	0.60	0.68	0.85	0.72
(Univariate)	(0.50)	(2.37)	(3.52)	(4.12)	(4.91)	(4.64)
Mom	0.12	0.40	0.62	0.75	0.84	0.72
(Bivariate)	(0.53)	(1.96)	(3.42)	(4.20)	(4.62)	(6.28)
Rev	0.22	0.53	0.63	0.64	0.83	0.61
(Bivariate)	(0.84)	(2.64)	(3.63)	(3.87)	(4.89)	(3.79)
$MLER^{Mom,Rev}$	0.32	0.50	0.57	0.64	0.76	0.44
(Bivariate)	(1.46)	(2.54)	(3.18)	(3.77)	(4.35)	(3.99)
Mom and Rev	0.39	0.52	0.62	0.74	0.79	0.40
(Trivariate)	(1.65)	(2.54)	(3.24)	(4.09)	(4.35)	(3.79)

$MLER^{Mom,Rev}$ and the trivariate portfolio analysis that controls for Mom and Rev is not surprising because the trivariate portfolio analysis captures both nonlinearities and interaction effects in the ability of Mom and Rev to predict the cross-section of future stock returns. As with the univariate portfolios, the bivariate and trivariate portfolios exhibit monotonically increasing average excess returns across the $MLER$ quintiles.

The results of the multivariate portfolio analyses show that while momentum and reversal contribute to $MLER$'s predictive power, $MLER$ also has substantial predictive power that is unrelated to these effects. In Section XVII and Table A13 of the Internet Appendix we show that FM regression analyses examining the ability of $MLER$ to predict the future stock returns after controlling for Mom, Rev, and $MLER^{Mom,Rev}$ lead to the same conclusion.

5.4. Image-based forecasts

In addition to momentum and reversal, there is some evidence of other technical signals that predict future stock returns. Most notably, Jiang et al. (2022, JKX hereafter) show that forecasts generated by applying a convolutional neural network to images depicting past price and volume data are informative about future returns. We therefore examine whether controlling for JKX's forecasts explains the predictive power of $MLER$. Specifically, we control for each of JKX's forecasts that are designed to predict one-month-ahead stock returns. These forecasts are based on charts of 60 days ($I60/R20$), 20 days ($I20/R20$), and 5 days ($I5/R20$) of past daily return and volume data.[25] The forecasts are available for sample months t from 200102-202001. All tests that use these variables cover this period.

5.4.1. Relations between ML-based forecasts and image-based forecasts

We begin by examining the strength of the relation between the JKX variables and $MLER$ by running FM regression analyses with $MLER$ as the dependent variable and combinations of $I60/R20$, $I20/R20$, and $I5/R20$ as independent variables. Table 13 shows that the average adjusted R^2s from these cross-sectional regressions are 7.23%, 2.43%, and 1.07% when only $I60/R20$, $I20/R20$, and $I5/R20$, respectively, are included as independent variables. When all three JKX variables are included in the regression specification, the average adjusted R^2 is 7.86%. In all specifications, the average slope coefficient on each of the JKX variables is significantly positive. The results therefore indicate that while there is some commonality in the forecasts, most of the variation in $MLER$ is unexplained by the JKX variables. The stronger relation between $MLER$ and $I60/R20$ than between $MLER$ and either $I20/R20$ or $I5/R20$ is likely due to the greater overlap in the past data used to generate these forecasts.

[25] We are very grateful to Bryan Kelly and Dacheng Xiu for sharing their stock, month-level forecasts.

Table 13
Relations between ML-Based Forecasts and Image-Based Forecasts. This table presents the results of Fama and MacBeth (1973) regression analyses examining the ability of $I60/R20$, $I20/R20$, and $I5/R20$ to explain variation in $MLER$. The process for conducting the Fama and MacBeth (1973) regression analyses is exactly as described in Table 10. The dependent variable in the regressions is $MLER$ and the independent variables are combinations of $I60/R20$, $I20/R20$, and $I5/R20$. Results for different regression specifications are shown in different columns of the table. The table presents the time-series averages of monthly cross-sectional regression coefficients, along with t-statistics, calculated following Newey and West (1987) using 12 lags, testing the null hypothesis that the average coefficient is equal to zero (in parentheses). The rows labeled "$I60/R20$", "$I20/R20$", and "$I5/R20$" present results pertaining to the coefficient on the indicated variable. The row labeled "Adj. R^2" shows the average adjusted R^2 from the cross-sectional regressions. The analyses cover sample months t from February 2001 through January 2020, inclusive.

	(1)	(2)	(3)	(4)
$I60/R20$	0.586			0.559
	(58.46)			(56.62)
$I20/R20$		0.300		0.024
		(30.03)		(2.46)
$I5/R20$			0.225	0.059
			(16.66)	(4.92)
Adj. R^2	7.23%	2.43%	1.07%	7.86%

5.4.2. Alphas relative to factor models with image-based forecast factors

To test whether the predictive power of $MLER$ is subsumed by that of the image-based forecasts of JKX, we first investigate whether the performance of the $MLER$ $10-1$ portfolio can be explained by factors based on $I60/R20$, $I20/R20$, and $I5/R20$. For each image based forecast $X \in \{I60/R20, I20/R20, I5/R20\}$, we construct a factor using a methodology similar to that of Fama and French (1993).[26] Specifically, each month t, we sort all stocks into two groups based on market capitalization, using the median market capitalization of NYSE-listed stocks at the end of month $t-1$ as the breakpoint. Independently, we sort all stocks into three groups based on X measured at the end of month $t-1$, using the 30th percentile and 70th percentile values of X calculated using only NYSE-listed stocks as the breakpoints. The intersections of the two market capitalization-based groups and three groups formed by sorting on X make six portfolios. We take the month t excess return of each of these portfolios to be the market capitalization-weighted average month t excess return of all stocks in the given portfolio. The excess return of the factor based on X, which we denote F_X, is defined as the average excess return of the two portfolios that contain stocks with values of X above the NYSE 70th percentile minus the average excess return of the two portfolios that contain stocks with values of X below the NYSE 30th percentile. We then augment each of the CAPM, FF, FFC, FFC+REV, FF5, and Q factor models with $F_{I60/R20}$, $F_{I20/R20}$, and $F_{I5/R20}$ and conduct factor model regressions of the $MLER$ $10-1$ portfolio excess returns using these augmented factor models.

Since the analyses using factors based on the image-based forecasts cover only the 200102-202001 subperiod for which the image-based forecasts are available, we begin by describing the performance of the $MLER$ $10-1$ portfolio during this subperiod. Table 14 shows that the average excess return of the $MLER$ $10-1$ portfolio during this period is 0.82% per month, which is economically large but statistically insignificant, with a t-statistic of 1.72. We attribute the statistical insignificance of the $MLER$ $10-1$ portfolio's average excess return during this period to the reduced test power associated with the shortened length of the sample period (19 years compared to 59.5 years for the full sample period).[27] Furthermore, the $MLER$ $10-1$ portfolio's average excess return of 0.82% per month during the 200102-202001 subperiod covered by these tests is lower than the 1.08% per month generated by this portfolio during the full sample period, suggesting that measures of the $MLER$ $10-1$ portfolio's performance based on this sample may be downward-biased estimates of the corresponding true population values.[28]

The top of Table 15 shows the monthly alphas of the $MLER$ $10-1$ portfolio during the 200102-202001 subperiod relative to previously-established factor models. Despite the concerns about bias and test power, the monthly alphas with respect to the CAPM, FF, FFC, and FFC+REV models of 1.15% (t-statistic = 2.50), 1.19% (t-statistic = 2.56), 0.75% (t-statistic = 2.38), and 0.64% (t-statistic = 2.48), respectively, are all economically large and highly statistically significant. The monthly alphas with respect to the FF5 and Q models of 0.65% (t-statistic = 1.88) and 0.45% (t-statistic = 1.16), respectively, are both insignificant. As shown in Section 4.2 and Table 4, the alphas with respect to all models are highly significant when estimated using the full sample period. We therefore attribute the insignificance of the FF5 and Q factor model alphas during the 200102-202001 subperiod to some combination of low test power and potential bias. The insignificance of these alphas makes it impractical to assess whether the predictive power of $MLER$ is subsumed by that of the image-based forecasts based on the statistical significance of the alphas from the augmented factor model regressions. Our interpretation of the results of the factor model regressions that use the augmented factor models will therefore focus on the reduction in alpha relative to the unaugmented model, the change in the adjusted R^2 relative to the unaugmented model, and the slope coefficients on $F_{I60/R20}$, $F_{I20/R20}$, and $F_{I5/R20}$. If the predictive power of $MLER$ is subsumed by that of the image-based forecasts, we expect the estimated alpha from the augmented factor model to be much lower than that of the unaugmented model and close to zero, the adjusted R^2 from the augmented model to be substantially higher than that of the unaugmented model, and the slope coefficients on $F_{I60/R20}$, $F_{I20/R20}$, and $F_{I5/R20}$ to be large and significant.

Table 15 shows that the $MLER$ $10-1$ portfolio's alphas with respect to the augmented CAPM, FF, FFC, and FFC+REV factor models of 1.01% (t-statistic = 2.79), 1.10% (t-statistic = 3.09), 0.67% (t-statistic = 2.30), and 0.61% (t-statistic = 2.44), respectively, are once again all economically large and highly significant. Thus, for all unaugmented factor models that produced a significant $MLER$ $10-1$ portfolio alpha during 200102-202001 subperiod, the corresponding augmented model alpha is also significant. While the slope coefficient on $F_{I60/R20}$ when using the augmented CAPM and FF models is significantly positive, this slope coefficient is small and insignificant when using in the augmented FFC and FFC+REV models. Furthermore, the alphas and adjusted R^2 values from the augmented FFC and FFC+REV factor models are nearly unchanged from those of the corresponding unaugmented models. These results indicate that any commonality in the predictive power of $MLER$ and the image-based forecasts is because both $MLER$ and the image-based forecasts have components related to momentum and reversal. When using the FF5 and Q factor models, the results show that the augmented models produce alphas that are similar to, and adjusted R^2 values that are only slightly higher than, those of the corresponding unaugmented models. The slope coefficients on

[26] In Section XVIII and Table A14 of the Internet Appendix, we repeat our tests using factors defined as the excess return of the long-short portfolio constructed from decile portfolios formed by sorting on the JKX variables and find no qualitative change in the results of our analyses.

[27] Other papers with short sample periods (e.g. Martin (2017)) face similar issues of low statistical power.

[28] The $MLER$ $10-1$ produces an average excess return greater than 0.82% per month in 397 of 487 consecutive 19-year subperiods during the full 196307-202212 sample period.

Table 14
Portfolio Analysis - 200102-202001. This table presents the results of a portfolio analysis examining the ability of $MLER$ to predict the cross-section of future stock returns during the 200102-202001 period. The analysis is identical to that whose results are shown in Table 3 except that the results presented in this table are for return months t from February 2001 through January 2020, inclusive.

Value	$MLER$ 1	$MLER$ 2	$MLER$ 3	$MLER$ 4	$MLER$ 5	$MLER$ 6	$MLER$ 7	$MLER$ 8	$MLER$ 9	$MLER$ 10	$MLER$ 10 − 1
$\bar{r}$	−0.02	0.41	0.50	0.56	0.52	0.71	0.68	0.58	0.68	0.80	0.82
	(−0.03)	(0.86)	(1.22)	(1.46)	(1.56)	(2.58)	(2.37)	(1.90)	(2.36)	(2.32)	(1.72)
SD	7.52	6.21	5.57	5.30	4.44	4.22	4.00	3.98	3.88	4.50	5.75
Sharpe	−0.01	0.23	0.31	0.36	0.40	0.59	0.59	0.51	0.61	0.62	0.49

Table 15
Alphas Using Image-Based Forecast Factors. This table presents the results of factor model regressions of the excess returns of the $MLER$ 10 − 1 portfolio on the excess returns of factors in different factor models. The $MLER$ 10 − 1 portfolio is the same as that whose average excess returns are shown in Table 3. The excess return of the F_X factor, for $X \in \{I60/R20, I20/R20, I5/R20\}$, is that of the long-short portfolio formed by sorting on both market capitalization and X, constructed using a methodology, described in Section 5.4.2, similar to that of Fama and French (1993). The column labeled "Model" indicates the factor model, where "Base Model" refers to the model indicated in the header of columns other than the "Model" and "Value" columns and "Base Model+ $F_{I60/R20} + F_{I20/R20} + F_{I5/R20}$" refers to a model that augments the factors in the model shown in the column header with $F_{I60/R20}$, $F_{I20,R20}$, and $F_{I5/R20}$. The rows with "α" and "Adj. R^2" in the "Value" column present the intercept and adjusted R^2, respectively, of the regression. The rows with "β_{F_X}" in the "Value" column present the slope coefficient on F_X from the regression. Alphas are reported in percent per month. The values in parentheses are t-statistics, calculated following Newey and West (1987) using 12 lags, testing the null hypothesis that the average monthly alpha or slope coefficient is equal to zero. The analyses cover return months t from February 2001 through January 2020, inclusive.

Model	Value	CAPM	FF	FFC	FFC+REV	FF5	Q
Base Model	α	1.15	1.19	0.75	0.64	0.65	0.45
		(2.50)	(2.56)	(2.38)	(2.48)	(1.88)	(1.16)
	Adj. R^2	19.88%	22.42%	57.20%	65.38%	30.82%	48.96%
Base Model+ $F_{I60/R20} + F_{I20/R20} + F_{I5/R20}$	α	1.01	1.10	0.67	0.61	0.72	0.47
		(2.79)	(3.09)	(2.30)	(2.44)	(2.64)	(1.36)
	$\beta_{F_{I60/R20}}$	1.20	1.20	0.18	0.31	0.83	0.38
		(3.29)	(3.31)	(0.67)	(1.41)	(1.83)	(1.39)
	$\beta_{F_{I20/R20}}$	−0.33	−0.39	0.18	0.19	−0.31	−0.11
		(−1.10)	(−1.32)	(0.70)	(0.73)	(−1.08)	(−0.42)
	$\beta_{F_{I5/R20}}$	−0.23	−0.28	−0.12	−0.28	−0.22	−0.23
		(−0.78)	(−0.93)	(−0.51)	(−1.47)	(−0.68)	(−0.98)
	Adj. R^2	27.80%	29.82%	56.94%	65.61%	33.70%	50.70%

each of $F_{I60/R20}$, $F_{I20/R20}$, and $F_{I5/R20}$ are also insignificant in the augmented FF5 and Q factor models.

In sum, the results in Table 15 support the conclusion that the predictive power of $MLER$ remains strong after controlling for the image-based forecasts of JKX.

5.4.3. Additional tests that control for image-based forecasts

We conduct several additional tests to further investigate whether the predictive power of $MLER$ persists after controlling for the image-based forecasts of JKX. Our first such tests are bivariate portfolio analyses. The results of these tests, presented in Section XIX and Table A15 of the Internet Appendix, demonstrate that portfolios constructed to be neutral to one of the JKX variables while taking long (short) positions in stocks with high (low) values of $MLER$ generate economically large and highly statistically significant average excess returns. These tests provide further evidence that the predictive power of $MLER$ is distinct from that of $I60/R20$, $I20/R20$, and $I5/R20$.

We next investigate the ability of the JKX variables to explain the predictive power of $MLER$ using FM regression analyses with the one-month-ahead excess stock return as the dependent variable and combinations of $MLER$, $I60/R20$, $I20/R20$, $I5/R20$, Mom, Rev, and $MLER^{Mom,Rev}$ as independent variables. These analyses allow us to simultaneously control for all of the JKX variables, as well as momentum and reversal. The results of these tests, presented in Section XX and Table A16 of the Internet Appendix, show that regardless of which combination of the JKX variables, Mom, Rev, and $MLER^{Mom,Rev}$ is included as controls in the regression specification, the average coefficient on $MLER$ remains positive and highly statistically significant.

Our next tests investigate whether the predictive power of $MLER$ persists after accounting for any potential nonlinearity in the relation between the JKX variables and $MLER$. Including the JKX variables as controls in FM regression analyses may not effectively control for the relation between $MLER$ and these variables if the relation between the JKX variables and $MLER$ is not linear. To assess the degree of nonlinearity in the relations between these variables, we begin by generating plots of $MLER$ and the JKX variables. These plots, shown in Figure A6 of the Internet Appendix, provide no evidence of nonlinear relations between $MLER$ and the JKX variables. We then run FM regression analyses that include, in addition to the untransformed JKX variables, the natural log, square root, and square of $I60/R20$, $I20/R20$, and $I5/R20$ as independent variables in the regressions.[29] High correlations between the untransformed and transformed versions of the JKX variables cause multicollinearity challenges with these re-

[29] Our approach is similar to that of Kirby (2020), who includes the squared and cubed values, as well as pairwise interactions, of stock-level characteristics in cross-sectional regressions designed to estimate the returns associated with pure play factor portfolios when the relation between the characteristics and expected returns is nonlinear.

gressions that make it difficult to interpret the slope coefficients on the JKX variables. These issues do not affect the average slope coefficients on $MLER$, which remain positive and highly statistically significant in all such tests (untabulated). Finally, to address the concern of nonlinearity while overcoming the challenges associated with multicollinearity in the FM regression analyses, we use feed forward neural networks to create forecasts of $MLER$ based on the JKX variables, and include these forecasts, which we denote $E[MLER|\text{JKX}]$, as a control in FM regressions analyses. The calculation of $E[MLER|\text{JKX}]$ is described in detail in Section XX of the Internet Appendix. The results of tests using $E[MLER|\text{JKX}]$ are shown in Tables A17 and A18 of the Internet Appendix. In all specifications, the average slope coefficient on $MLER$ remains positive and highly statistically significant after controlling for $E[MLER|\text{JKX}]$. These tests, therefore, provide additional evidence that the predictive power of $MLER$ is distinct from that of the image-based forecasts of JKX.

Our final tests investigating whether the predictive power of $MLER$ persists after controlling for the JKX forecasts use the omitted factor methodology of Giglio and Xiu (2021, GX hereafter). As described in Section XXI and Table A19 of the Internet Appendix, we use the GX methodology to estimate the risk premium of the univariate $MLER$ $10-1$ portfolio based on a large set of base test assets that does not include portfolios formed by sorting on the JKX variables (JKX portfolios hereafter), and then again with an augmented set of test assets that includes JKX portfolios. We find that adding the JKX portfolios to the set of test assets has almost no effect on the estimate of the $MLER$ $10-1$ portfolio's risk premium, indicating that the JKX variables do not provide incremental ability to explain the predictive power of $MLER$ relative to the variables underlying the base set of test assets. Furthermore, regardless of whether the set of assets used in the GX methodology includes or excludes the JKX portfolios, the estimates of the $MLER$ $10-1$ portfolio's risk premium are much smaller than the average excess return of the $MLER$ $10-1$ portfolio and statistically insignificant, indicating that the predictive power of $MLER$ is not subsumed by a combination of the variables used in the construction of the test assets.

Summarily, the results of these portfolio analyses, FM regression analyses, and analyses using the GX methodology, indicate that the ability of $MLER$ to predict the cross-section of future stock returns remains strong after controlling for the image-based forecasts of JKX, momentum, and reversal.

5.5. Technical signals

We next examine whether the predictive power of our ML-based forecasts is subsumed by previously-studied technical trading signals. The technical signals we use come from Neely et al. (2014, NRTZ hereafter) and Freyberger et al. (2020, FNW hereafter). NRTZ study 14 technical signals, each of which is an indicator set to one (zero) for stocks with positive (negative) forecasts. These signals are well-suited for our purposes because, like $MLER$, they are based on monthly return data. Six of these signals are based on moving average prices, two are momentum-based signals, and the remaining six are derived from on-balance volume (OBV). The moving average signals ($I_{MA(s)\geq MA(l)}$) are set to one if the moving average price over the past s months ($MA(s)$) is greater than or equal to the moving average price over the past l months ($MA(l)$), and zero otherwise, where $s \in \{1,2,3\}$ and $l \in \{9,12\}$. The momentum signals ($I_{P_t \geq P_{t-m}}$) are set to one if the current price is greater than or equal to the price m months ago, and zero otherwise, for $m \in \{9,12\}$. The OBV signals ($I_{MAOBV(s)\geq MAOBV(l)}$) take the value one if the moving average OBV over the past s months ($MAOBV(s)$) is greater than or equal to the moving average OBV over the past l months ($MAOBV(l)$), for $s \in \{1,2,3\}$ and $l \in \{9,12\}$. The OBV of any stock i in any month t is the sum over all months k beginning when the stock is listed and ending in month t, of the stock's month k signed trading volume, where the signed trading volume is the stock's trading volume (the negative of the stock's trading volume) if the stock's price at the end of month k is greater than or equal to (less than) its price at the end of month $k-1$.[30]

FNW study the joint predictive power of 62 stock-level characteristic variables, many of which they characterize as "trading friction" variables. We use 14 trading frictions variables examined in FNW, most of which are calculated from past market data and thus may be discernable from plots depicting such data.[31] These variables include total volatility ($TotVol$) and idiosyncratic volatility ($IdioVol$) calculated following Ang et al. (2006), price relative to its 52-week high ($Price/High$, see George and Hwang (2004)), the maximum daily return in the past month (Max, see Bali et al. (2011)), the standard deviation of daily turnover (σ_{Turn}) and the standard deviation of daily volume (σ_{Volume}) of Chordia et al. (2001), beta (β_{FP}) calculated following Frazzini and Pedersen (2014), beta (β_{LN}) calculated following Lewellen and Nagel (2006), detrended market-adjusted turnover (DTO) and standardized unexpected volume (SUV) of Garfinkel (2009), turnover ($Turn$, see Datar et al. (1998)), total assets (AT), market capitalization ($MktCap$), and industry-demeaned market capitalization ($MktCap_{IndAdj}$). Detailed descriptions of the calculations of these variables are in Section XXII of the Internet Appendix. While AT, $MktCap$, and $MktCap_{IndAdj}$ are not calculated from histories of market data, we include them because they are plausibly related to market frictions that may manifest in patterns evident in historical price plots.

In Panel A of Table 16 we present the average excess returns of long-short portfolios formed by sorting on $MLER$ that are neutral to one of the NRTZ technical signals. For each NRTZ technical signal, we sort stocks with each value of the signal (zero or one) into decile $MLER$ portfolios using NYSE breakpoints. We calculate the excess returns of each of the resulting value-weighted portfolios, and take the $MLER$ $10-1$ portfolio excess return to be the equal-weighted average excess return of the two $MLER$ decile 10 portfolios minus that of the two $MLER$ decile one portfolios. The results show that the average excess returns of each of these $MLER$ $10-1$ portfolios is positive, large in magnitude, and highly statistically significant, with all associated t-statistics being 5.67 or higher.

Table 16 Panel B presents the average monthly excess returns of $MLER$ $5-1$ portfolios that are constructed to be neutral to a FNW technical variable. The methodology used to construct these portfolios is identical to that used to construct the $MLER$ $5-1$ portfolios that control for momentum and reversal, discussed in Section 5.3.2, except we use one of the FNW technical signals as the first sort variable. The average monthly excess returns of the $MLER$ $5-1$ portfolios that are neutral to the FNW technical signals are all positive, economically large, and highly statistically significant, ranging from 0.70% (t-statistic $=6.26$) when controlling for β_{FP} to 1.02% (t-statistic $=6.20$) when controlling for $MktCap$.

In Section XXIII and Table A20 of the Internet Appendix we use equal-weighted and value-weighted FM regression analyses to further investigate whether the predictive power of $MLER$ persists after controlling for the technical signals. Regardless of which set of technical variables we include as controls in the regression specification, the average coefficient on $MLER$ is positive and highly statistically significant. Finally, we use the excess returns of a long-short portfolio generated by FNW as a factor in factor model regressions.[32] FNW construct this portfolio by sorting stocks based on a composite forecast calculated by

[30] Our definitions follow those of NRTZ. Prices and volumes are adjusted for splits and stock dividends.

[31] FNW use 15 trading friction variables in their analyses. We use the 14 variables that are available for the entirety of our sample period. Due to data constraints discussed in Chung and Zhang (2014), the average daily bid-ask spread variable (denoted "Spread" in FNW, see their Table 1) cannot be calculated for a large portion of our sample period.

[32] We thank Michael Weber for sharing the excess returns of this portfolio.

Table 16
Multivariate Portfolio Analysis - Control for Technical Signals. This table presents the results of bivariate portfolio analyses examining the ability of the ML-based forecasts to predict the cross-section of future stock returns after controlling for other technical signals. The technical signals are those used by Neely et al. (2014, NRTZ) and Freyberger et al. (2020, FNW), described in Section 5.5. Panel A presents the average excess returns of $MLER$ $10-1$ portfolios designed to be neutral to different NRTZ technical signals (indicated in the column headers). At the end of each month $t-1$, stocks with each value of the given technical signal (either 0 or 1), are sorted into decile portfolios based on an ascending ordering of $MLER$ using breakpoints calculated from NYSE-listed stocks. The month t excess return of each portfolio is then taken to be the market capitalization-weighted average month t excess return of all stocks in the portfolio, with market capitalization calculated as of the end of month $t-1$. The $MLER$ $10-1$ portfolio excess return is taken to be the equal-weighted average excess return of the two decile 10 $MLER$ portfolios minus that of the two decile 1 portfolios. Panel B presents the average excess returns of $MLER$ $5-1$ portfolios designed to be neutral to different FNW technical signals (indicated in the column headers). The methodology used to construct these $MLER$ $5-1$ portfolios is identical to that used to construct the bivariate portfolios examined in Table 12 except that one of the FNW technical signals is used as the first sort variable. The values in parentheses are t-statistics, calculated following Newey and West (1987) using 12 lags, testing the null hypothesis that the average monthly excess return is equal to zero. The analyses cover return months t from July 1963 through December 2022, inclusive.

Panel A: $MLER$ $10-1$ Port. Avg. Excess Return - Control for NRTZ Variables

$I_{MA(1)\geq MA(9)}$	$I_{MA(1)\geq MA(12)}$	$I_{MA(2)\geq MA(9)}$	$I_{MA(2)\geq MA(12)}$	$I_{MA(3)\geq MA(9)}$	$I_{MA(3)\geq MA(12)}$	$I_{P_t\geq P_{t-9}}$	$I_{P_t\geq P_{t-12}}$	$I_{OBVMA(1)\geq OBVMA(9)}$	$I_{OBVMA(1)\geq OBVMA(12)}$	$I_{OBVMA(2)\geq OBVMA(9)}$	$I_{OBVMA(2)\geq OBVMA(12)}$	$I_{OBVMA(3)\geq OBVMA(9)}$	$I_{OBVMA(3)\geq OBVMA(12)}$
1.13	1.19	1.15	1.18	1.07	1.08	1.15	1.10	1.17	1.15	1.15	1.11	1.10	1.09
(5.67)	(7.01)	(6.70)	(7.63)	(6.50)	(6.68)	(6.73)	(6.77)	(6.07)	(6.06)	(6.10)	(6.03)	(5.77)	(6.07)

Panel B: $MLER$ $5-1$ Port. Avg. Excess Return - Control for FNW Variables

$TotVol$	$IdioVol$	$Price/High$	Max	σ_{Turn}	σ_{Volume}	β_{FP}	β_{LN}	DTO	SUV	$Turn$	AT	$MktCap$	$MktCap_{IndAdj}$
0.75	0.83	0.75	0.80	0.83	0.94	0.70	0.76	0.80	0.79	0.81	0.97	1.02	0.93
(6.85)	(6.88)	(4.66)	(6.87)	(6.17)	(5.76)	(6.26)	(6.59)	(6.02)	(5.11)	(6.16)	(6.55)	(6.20)	(5.60)

applying the adaptive group LASSO methodology to 62 stock-level characteristics, which include the technical variables mentioned above, past return-based variables including Mom and Rev, and many variables constructed from financial statement data. The results of these tests, discussed in detail and presented in Section XXIV and Table A21 of the Internet Appendix, show that the $MLER$ $10-1$ portfolio generates positive, economically large, and highly statistically significant alpha with respect to traditional factor models augmented with the excess returns of FNW's long-short portfolio.

We conclude from these tests that the predictive power of the ML-based forecasts is not subsumed by previously-studied technical signals.

5.6. Visual analysis of charts

Our final analyses characterizing the predictive power of $MLER$ aim to describe what charts that predict high or low future returns look like. We begin by plotting the average chart for the focal $MLER$ decile portfolios that we examined in Table 3. Specifically, for each month t and each decile portfolio, we calculate the value-weighted mean value of CR_k for $k \in \{1, 2, \ldots, 12\}$. In Fig. 5, we plot the time-series averages of these monthly mean values of CR_k for each portfolio.[33] As would be expected, the features that show up most prominently in the charts are the momentum and reversal effects. Specifically, the charts show that the average values of CR_{11}, which is identical to Mom, are increasing across the $MLER$ deciles. The large negative (positive) slope of the chart associated with $MLER$ decile 10 (1) stocks in month 12 reflects the large negative (positive) average returns, or values of Rev, of such stocks in the last month prior to portfolio formation.

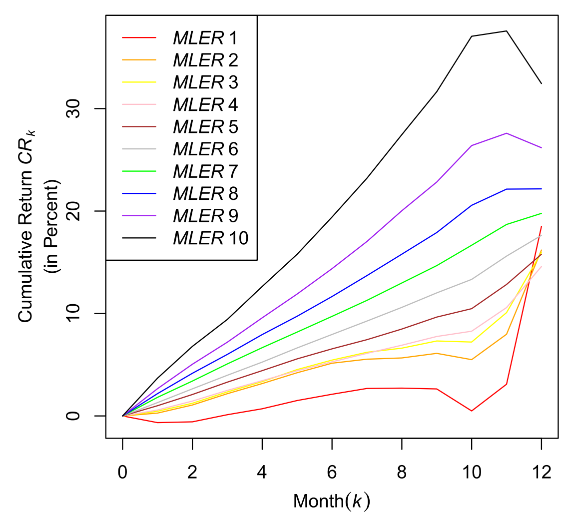

Fig. 5. Average Charts for $MLER$ Decile Portfolios. This figure depicts the average chart for stocks within each $MLER$ decile portfolio. The portfolios are the same portfolios as those whose returns are described in Table 3. For each portfolio in each month t, we calculate the value-weighted mean value of kth monthly cumulative return that would appear in charts of the stocks' cumulative monthly returns over the past year, CR_k. We then calculate the time-series averages of these monthly means for each decile portfolio and each value of $k \in \{1, 2, \ldots, 12\}$. The x-axis depicts the month k in the chart. The y-axis depicts the time-series average of the monthly mean cumulative returns. Values for different $MLER$ decile portfolios are shown in different colors, as indicated in the legend. For interpretation of the colors in the figure(s), the reader is referred to the web version of this article. The analysis covers sample months t from July 1963 through December 2022, inclusive.

[33] The charts generated using equal-weighted mean values of CR_k, presented in Section XXV and Figure A7 of the Internet Appendix, are qualitatively the same as those of the value-weighted charts.

To discern the features of charts that are important for return prediction but distinct from momentum or reversal, we examine a subsample of stocks in which the momentum and reversal effects do not exist. The exact process and empirical results that we use to identify such a subsample are described in detail in Section XXVI and Table A22 of the Internet Appendix. Here, we summarize. We begin by taking only stocks with moderate values of Mom and Rev. Specifically, each month t, we take only stocks with values of Mom and Rev that are between the 20th and 80th percentile values among NYSE-listed stocks of both of these variables in the given month. We find that while the momentum effect does not exist in this subsample, the reversal effect remains strong among these stocks. Thus, we incrementally shrink the range of values of Rev that we include in the subsample until we find a subsample in which the reversal effect does not exist. The subsample selected using this process includes the intersection of stocks with Mom between the 20th and 80th NYSE percentile values of Mom and stocks with Rev between the 40th and 60th NYSE percentile values of Rev. All percentiles in these analyses are calculated using only NYSE-listed stocks. In this subsample, the average excess return of the long-short quintile portfolio formed by sorting Mom is 0.04% per month (t-statistic = 0.31) and that of Rev is −0.06% per month (t-statistic = −0.52), both of which are economically small and statistically insignificant, indicating that the momentum and reversal effects are non-existent in this subsample of stocks. The long-short portfolio formed by sorting on $MLER$, however, produces an economically large and highly statistically significant average excess return of 0.37% per month (t-statistic = 2.23). Since the momentum and reversal effects are non-existent, the results demonstrate that the chart patterns picked up by the ML-based forecasts that are relevant for return prediction in this subsample are something other than momentum and reversal.

To visualize the features of the charts associated with the predictive power of the ML-based forecasts in this subsample, we create charts of the average values of CR_k among stocks with different values of $MLER$. Since this subsample does have substantial variation in Mom, to clearly expose features of the charts that are distinct from Mom, we separate this subsample into six different Mom-based groups. Specifically, we create separate charts for stocks with values of Mom between the 20th and 30th (decile 3), 30th and 40th (decile 4), 40th and 50th (decile 5), 50th and 60th (decile 6), 60th and 70th (decile 7), and 70th and 80th (decile 8) NYSE percentiles of Mom. We then create $MLER$-sorted quintile portfolios using stocks in each of these six groups and construct the chart of the average value-weighted means of CR_k for each of these portfolios.[34] These charts, shown in Fig. 6, exhibit a few striking patterns. First, in all six plots, the average stock in the $MLER$ quintile 5 (1) portfolio has a substantial increase (decrease) in value in month 10, followed by a substantial decrease (increase) in value in month 11. Furthermore, among stocks in Mom deciles 3-5, and to a lesser degree decile 6, we see that stocks with low values of $MLER$, and thus low forecast returns, actually tend to have the highest average cumulative returns in months 1-6 in the chart, and then tend to underperform in months 7-10.

The charts demonstrate that, even among stocks with similar values of Mom and Rev, stocks with high and low ML-based forecasts tend to have sufficiently different charts for the patterns to be recognized by a human chart reader. This, combined with the fact that these ML-based forecasts predict future stock returns, albeit not through the momentum and/or reversal channels, suggests that human chartists can produce forecasts that form the basis of successful investment strategies that are distinct from the momentum and reversal effects.

[34] In Section XXVII and Figure A8 of the Internet Appendix, we show that charts created from equal-weighted values of CR_k are qualitatively the same as those in Fig. 6.

6. Ex-post evaluation of optimization procedure

Our final tests evaluate the success of the optimization procedure that determined which ML model would be used to generate our focal forecasts. Specifically, we evaluate whether ML models that the optimization process, which used data from the 192701-196306 optimization period, identified as performing well, also produced superior out-of-sample forecasts during the 196307-202212 test period. To do so, we calculate forecasts using each of the 95 alternative ML models that were not selected by the optimization process. With the exception of the ML model, all other aspects of calculating these alternative ML-based forecasts are identical to those used to calculate $MLER$.

To evaluate the ability of the different forecasts to predict future stock returns, we repeat the decile portfolio analysis whose results are shown in Table 3 using each of these different forecasts. Table 17 presents the average monthly excess return of the 10 − 1 portfolio generated from each of these analyses. All but one of the 96 ML models examined (95 alternative models plus the selected model used to calculated $MLER$) produce 10 − 1 portfolios that have positive average excess returns, 90 of which are statistically significant at the 5% level. The fact that most models perform well during the test period suggests that our results are quite robust and not highly sensitive to the choice of ML model. Furthermore, while the model used to calculate $MLER$ performs well, there are 30 (32) alternative models whose 10 − 1 portfolios produce average monthly excess returns (t-statistics) that are greater than the 1.08% per month (t-statistic of 5.51) generated by the $MLER$ 10 − 1 portfolio. Thus, the model selected by the optimization process performs well relative to most other candidate models in out-of-sample tests, but as would be expected based on noise in the optimization process, there are models that performed better than that selected as our focal model.

In Fig. 7 we graphically depict the relation between performance of the forecasts as measured by the optimization process, and the out-of-sample test period performance of the associated 10 − 1 portfolio. Specifically, we plot on the x-axis the Spearman rank correlations between the forecasts and future stock returns that were used to evaluate model performance in the optimization period (reported in Table 1), and on the y-axis the annualized Sharpe ratios of the 10 − 1 portfolios associated with the given model during the test period. The plot illustrates a clear positive relation between the optimization period Spearman rank correlations and test period Sharpe ratios. The correlation between these values is 0.69.[35]

Taken together, the results in Table 17 and Fig. 7 demonstrate that the optimization process works as expected. The strong positive relation between performance during the optimization period and performance during the test period shows that ex-ante optimization is useful for determining which ML model to employ for out-of-sample use. However, the optimization process is noisy, and thus is unlikely to select the single model that will produce the best out-of-sample performance. Indeed, in our tests, approximately one third of the alternative models outperformed the model selected by the optimization process.

7. Conclusion

We test the weak form of the EMH by examining whether ML-based forecasts generated from past return data easily observable in historical price plots can predict the cross-section of future stock returns. We begin by using the 192701-196306 period to determine the optimal ML model to use to generate our ML-based forecasts. We find that using a convolutional neural network with long short-term memory as the ML architecture, the mean-squared-error as the loss function, weighting observations in the loss function in a manner that gives the same total

[35] The correlation between the optimization period Spearman rank correlations and the test period average excess returns is 0.68.

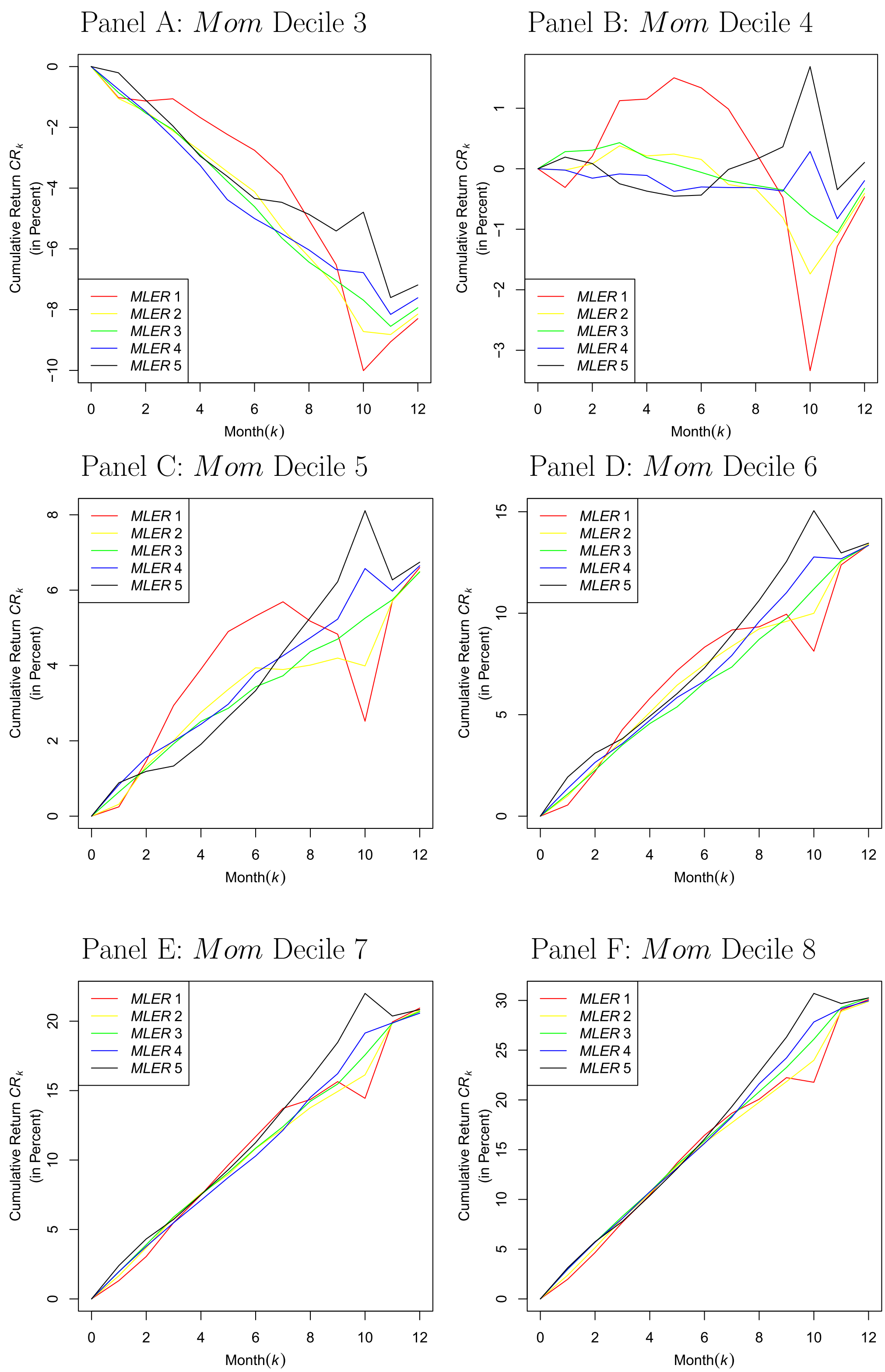

Fig. 6. Average Charts for $MLER$ Quintile Portfolios in Subsample without Momentum and Reversal Effects. This figure shows plots that depict the average chart for stocks within different subsets of our sample. The portfolios whose charts are depicted in Panels A, B, C, D, E, and F contain only stocks with values of Rev between the 40th and 60th percentile values of Rev and values of Mom between the 20th and 30th, 30th and 40th, 40th and 50th, 50th and 60th, 60th and 70th, and 70th and 80th percentile values of Mom, respectively. All percentiles are calculated on a monthly basis using only stocks listed on the NYSE. In each panel, all stocks in the given Rev- and Mom-based subset are sorted into 5 quintile portfolios based on $MLER$ using breakpoints calculated from only NYSE-listed stocks. For each portfolio in each month, we calculate the value-weighted mean value of the kth monthly cumulative return that would appear in charts of the stocks' cumulative monthly returns over the past year, CR_k, $k \in \{1, 2, \ldots, 12\}$. We then calculate the time-series averages of these monthly value-weighted means for each quintile portfolio and each value of k. The x-axis depicts the month k in the plot. The y-axis depicts the time-series average of the monthly mean cumulative return. Values for different $MLER$ quintile portfolios are shown in different colors, as indicated in the legends. For interpretation of the colors in the figure(s), the reader is referred to the web version of this article. The analysis covers sample months t from July 1963 through December 2022, inclusive.

Table 17

Alternative ML Models. This table presents the average monthly excess return for the long-short portfolio formed by sorting on ML-based forecasts generated using different ML models. We use each ML model described in Table 1 to generate forecasts of future stock returns using, with the exception of the ML model, the exact same procedure as was used to calculate $MLER$. We then construct decile portfolios formed by sorting on each of the different forecasts using the exact same decile portfolio construction methodology as was used in Table 3. The table presents the average monthly excess returns for the 10 − 1 portfolio formed using each of the forecasts. t-statistics, calculated following Newey and West (1987) using 12 lags, testing the null hypothesis that the average monthly excess return is equal to zero, are shown in parentheses. The column labeled "Dependent Variable" indicates the dependent variable. The column labeled "Weighting Methodology" indicates the weighting methodology. The remaining column headers indicate the ML architecture and loss function. The analyses cover return months t from July 1963 through December 2022, inclusive.

Dependent Variable	Weighting Methodology	FNN MAE	FNN MSE	CNN MAE	CNN MSE	LSTM MAE	LSTM MSE	CNNLSTM MAE	CNNLSTM MSE
r	EW	0.93	0.58	0.77	0.64	0.88	0.83	0.82	0.92
		(4.36)	(2.50)	(3.34)	(3.30)	(3.63)	(4.23)	(3.59)	(4.60)
	EWPM	0.90	0.61	0.71	0.44	0.97	0.80	0.88	1.07
		(4.44)	(2.40)	(3.33)	(1.98)	(4.50)	(3.83)	(4.04)	(5.62)
	EWPMVW	0.79	0.39	0.81	0.53	0.75	0.71	0.73	0.87
		(3.39)	(1.42)	(4.18)	(2.17)	(3.59)	(3.96)	(3.50)	(4.32)
r_{Std}	EW	0.62	0.03	1.15	1.28	1.03	1.29	1.03	1.41
		(2.42)	(0.12)	(5.60)	(6.34)	(4.67)	(7.73)	(4.95)	(7.54)
	EWPM	0.52	0.09	1.04	1.29	0.98	1.34	1.04	1.34
		(2.02)	(0.35)	(5.32)	(6.71)	(5.06)	(8.27)	(5.60)	(7.40)
	EWPMVW	−0.02	0.56	0.39	0.79	0.84	1.05	0.60	0.81
		(−0.08)	(2.37)	(2.54)	(4.59)	(5.92)	(7.46)	(4.10)	(5.67)
r_{Norm}	EW	1.03	0.66	1.15	1.19	1.16	1.19	1.17	1.21
		(4.32)	(2.59)	(5.30)	(5.52)	(5.11)	(5.49)	(5.20)	(5.87)
	EWPM	0.95	0.85	1.17	1.18	1.12	1.12	1.23	1.08
		(3.94)	(3.45)	(5.74)	(5.47)	(5.55)	(5.69)	(6.24)	(5.51)
	EWPMVW	0.92	0.68	0.76	0.79	1.06	1.13	0.97	0.92
		(4.15)	(3.01)	(4.84)	(4.59)	(6.36)	(7.23)	(5.59)	(6.94)
r_{Pctl}	EW	1.03	0.06	1.05	0.82	1.09	1.22	1.10	1.24
		(4.16)	(0.17)	(4.81)	(3.70)	(4.81)	(6.05)	(5.20)	(5.98)
	EWPM	1.12	0.26	0.94	0.91	1.16	1.18	1.19	1.22
		(4.66)	(0.93)	(4.96)	(4.68)	(5.81)	(6.37)	(6.51)	(6.67)
	EWPMVW	1.01	0.70	0.43	0.63	1.10	1.08	0.96	0.96
		(4.04)	(2.45)	(2.45)	(4.02)	(7.37)	(7.56)	(6.57)	(5.84)

weight to observations in each month and equal-weight to each stock within each month, and a normalized measure of the future stock return as the dependent variable, optimizes the ML model performance during the 192701-196306 optimization period. Our use of data from the early part of our sample period to determine the optimal ML model overcomes concerns about data mining and the out-of-sample validity of our results.

We use the selected ML model to generate out-of-sample forecasts of future stock returns during our 196307-202212 test period. Portfolio analyses demonstrate that the ML-based forecasts are a strong predictor of the cross-section of future stock returns. Factor analyses and other risk metrics provide no indication that the variation in average returns associated with the ML-based forecasts reflects compensation for risk. Further tests demonstrate that the predictive power of the ML-based forecasts is strong during most subperiods of our main test period, including the most recent 201501-202212 period. We also find that the forecasts are effective among the largest 500 stocks in our sample, indicating that the predictive power is not limited to small and illiquid stocks.

The forecasting function generated by the ML process is highly stable through time. Portfolios formed by sorting on forecasts based on fitting data from nonoverlapping subperiods have similar holdings and exhibit similar performance. Furthermore, we find only a small improvement in performance when using forecasts based on applying the learning process to data from an expanding window compared to a rolling window. The results of regression analyses show that the forecasting function has substantial nonlinear and interaction components, and that these components are important for prediction. While the ML-based forecasts contain nonnegligible components related to the momentum and reversal effects, the majority of the predictive power is unrelated to these phenomena. Similarly, ML-based forecasts based on images depicting historical price and volume data and previously-studied technical signals fail to explain the predictive power of our ML-based forecasts. Visual examination reveals differences in price charts associated with high and low future returns that are easily discernable by the human eye and distinct from momentum and reversal, suggesting the viability of charting by humans as a fruitful investment technique.

Finally, we conduct an ex-post analysis of the effectiveness of the optimization procedure we undertake to select the ML model used in our focal tests. Variation in the performance of different ML models during the 192701-196306 optimization period is strongly correlated with the out-of-sample performance of the associated forecasts during the 196307-202212 test period, indicating that the optimization procedure has substantial ability to discern the effectiveness of different ML models. However, approximately one third of the ML models examined have better out-of-sample performance than the selected model, suggesting that there is substantial noise in the optimization process. Long-short portfolios based on forecasts produced by all but one of the 96 ML models we examine generate positive out-of-sample average excess returns, almost all of which are statistically significant, indicating that our findings are robust to the use of different ML models.

Our results are strong evidence contrary to the main prediction of the EMH that profitable portfolios cannot be constructed from only information contained in historical returns. While one might reasonably argue that the momentum and reversal effects are already strong evidence contradicting the EMH, the complexity of the relations between

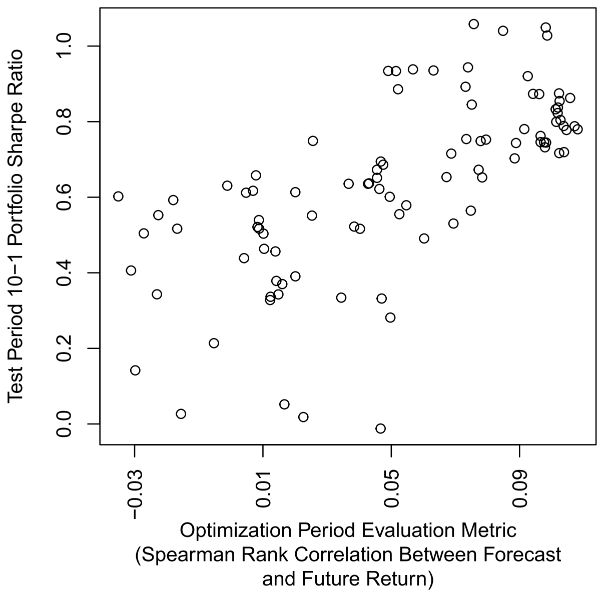

Fig. 7. Relation Between Optimization Period and Test Period Performance. This figure depicts the relation between the 192701-196306 optimization period performance and the 196307-202212 test period performance of ML-based forecasts calculated using different ML models. The x-axis depicts the Spearman rank correlations used to evaluate the ML model during the 192701-196306 optimization period, which are shown in Table 1. The y-axis depicts the annualized Sharpe ratios during the 196307-202212 test period of the $10-1$ portfolios formed by sorting stocks based on the forecasts generated by different ML models. Each point represents the results for a different ML model.

past price patterns and future returns indicate that violations of the EMH are much more intricate than previously understood. Our findings also suggest that technical analysis, or charting, has greater merit than acknowledged in academic work, and shed light on why this investment technique remains prevalent among investment practitioners.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data. However, the code to extract the required data from Wharton Research Data Services, along with all code required to reproduce the paper and an example of the sample we use with mostly randomized data, can be found at Mendeley Data under DOI https://dx.doi.org/10.17632/x63r376783.2.

Acknowledgements

Nikolai Roussanov was the editor for this article. We thank Nikolai Roussanov (the editor), an anonymous referee, Vikas Agarwal, Turan Bali, Bryan Kelly, Markus Pelger, Chip Ryan, Baozhong Yang, Guofu Zhou, and seminar participants at the 2020 Australasian Finance and Banking Conference, 2020 Conference on Asia-Pacific Financial Markets, 2020 International Risk Management Conference, 2020 Shanghai-Edinburgh Fintech Conference, 2020 World Finance & Banking Symposium, the 2021 Southwestern Finance Association Conference, 2021 University of Miami Research Conference on Machine Learning and Business, the University of Alabama, and Georgia State University for feedback that has substantially improved this paper. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562 (Towns et al. (2014)) and the Advanced Research Computing Technology and Innovation Core (ARCTIC) resources at Georgia State University's Research Solutions made available by the National Science Foundation Major Research Instrumentation (MRI) grant number CNS-1920024.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jfineco.2024.103791.

References

Allen, F., Karjalainen, R., 1999. Using genetic algorithms to find technical trading rules. J. Financ. Econ. 51, 245–271.
Ang, A., Hodrick, R.J., Xing, Y., Zhang, X., 2006. The cross-section of volatility and expected returns. J. Finance 61, 259–299.
Bajgrowicz, P., Scaillet, O., 2012. Technical trading revisited: false discoveries, persistence tests, and transaction costs. J. Financ. Econ. 106, 473–491.
Bali, T.G., Cakici, N., Whitelaw, R.F., 2011. Maxing out: stocks as lotteries and the cross-section of expected returns. J. Financ. Econ. 99, 427–446.
Bali, T.G., Engle, R.F., Murray, S., 2016. Empirical Asset Pricing: The Cross Section of Stock Returns. John Wiley & Sons.
Bali, T.G., Goyal, A., Huang, D., Jiang, F., Wen, Q., 2022. Predicting corporate bond returns: Merton meets machine learning. Available at SSRN: https://ssrn.com/abstract=3686164.
Bessembinder, H., Chan, K., 1998. Market efficiency and the returns to technical analysis. Financ. Manag., 5–17.
Bianchi, D., Büchner, M., Tamoni, A., 2021. Bond risk premiums with machine learning. Rev. Financ. Stud. 34, 1046–1089.
Brock, W., Lakonishok, J., LeBaron, B., 1992. Simple technical trading rules and the stochastic properties of stock returns. J. Finance 47, 1731–1764.
Bryzgalova, S., Pelger, M., Zhu, J., 2023. Forest through the trees: building cross-sections of stock returns. J. Finance. Forthcoming. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3493458.
Carhart, M.M., 1997. On persistence in mutual fund performance. J. Finance 52, 57–82.
Chang, P.K., Osler, C.L., 1999. Methodical madness: technical analysis and the irrationality of exchange-rate forecasts. Econ. J. 109, 636–661.
Chen, L., Pelger, M., Zhu, J., 2023. Deep learning in asset pricing. Manag. Sci. https://doi.org/10.1287/mnsc.2023.4695. Forthcoming.
Chinco, A., Clark-Joseph, A.D., Ye, M., 2019. Sparse signals in the cross-section of returns. J. Finance 74, 449–492.
Chordia, T., Subrahmanyam, A., Anshuman, V.R., 2001. Trading activity and expected stock returns. J. Financ. Econ. 59, 3–32.
Chung, K.H., Zhang, H., 2014. A simple approximation of intraday spreads using daily data. J. Financ. Mark. 17, 94–120.
Datar, V.T., Naik, N.Y., Radcliffe, R., 1998. Liquidity and stock returns: an alternative test. J. Financ. Mark. 1, 203–219.
Detzel, A.L., Novy-Marx, R., Velikov, M., 2023. Model comparison with transaction costs. J. Finance. https://doi.org/10.1111/jofi.13225. Forthcoming.
Ehsani, S., Linnainmaa, J.T., 2022. Factor momentum and the momentum factor. J. Finance 77, 1877–1919.
Fama, E.F., 1991. Efficient capital markets: II. J. Finance 46, 1575–1617.
Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. J. Finance 47, 427–465.
Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 33, 3–56.
Fama, E.F., French, K.R., 2008. Dissecting anomalies. J. Finance 63, 1653–1678.
Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. J. Financ. Econ. 116, 1–22.
Fama, E.F., French, K.R., 2018. Choosing factors. J. Financ. Econ. 128, 234–252.
Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: empirical tests. J. Polit. Econ. 81, 607–636.
Feng, G., He, J., Polson, N.G., 2018. Deep learning for predicting asset returns. arXiv preprint. https://arxiv.org/pdf/1804.09314.pdf.
Feng, G., He, J., Polson, N.G., Xu, J., 2023. Deep learning in characteristics-sorted factor models. J. Financ. Quant. Anal. Forthcoming. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3243683.
Frazzini, A., Pedersen, L.H., 2014. Betting against beta. J. Financ. Econ. 111, 1–25.
Freyberger, J., Neuhierl, A., Weber, M., 2020. Dissecting characteristics nonparametrically. Rev. Financ. Stud. 33, 2326–2377.
Garfinkel, J.A., 2009. Measuring investors' opinion divergence. J. Account. Res. 47, 1317–1348.
Gehrig, T., Menkhoff, L., 2006. Extended evidence on the use of technical analysis in foreign exchange. Int. J. Financ. Econ. 11, 327–338.
George, T.J., Hwang, C.Y., 2004. The 52-week high and momentum investing. J. Finance 59, 2145–2176.

Giglio, S., Xiu, D., 2021. Asset pricing with omitted factors. J. Polit. Econ. 129, 1947–1990.
Giglio, S., Liao, Y., Xiu, D., 2021. Thousands of alpha tests. Rev. Financ. Stud. 34, 3456–3496.
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.
Green, J., Hand, J.R., Zhang, X.F., 2017. The characteristics that provide independent information about average us monthly stock returns. Rev. Financ. Stud. 30, 4389–4436.
Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. Rev. Financ. Stud. 33, 2223–2273.
Gu, S., Kelly, B., Xiu, D., 2021. Autoencoder asset pricing models. J. Econom. 222, 429–450.
Guijarro-Ordonez, J., Pelger, M., Zanotti, G., 2023. Deep learning statistical arbitrage. Available at SSRN: https://ssrn.com/abstract=3862004.
Han, Y., Yang, K., Zhou, G., 2013. A new anomaly: the cross-sectional profitability of technical analysis. J. Financ. Quant. Anal. 48, 1433–1461.
Harvey, C.R., Liu, Y., Zhu, H., 2016. …and the cross-section of expected returns. Rev. Financ. Stud. 29, 5–68.
Hoseinzade, E., Haratizadeh, S., 2019. Cnnpred: CNN-based stock market prediction using a diverse set of variables. Expert Syst. Appl. 129, 273–285.
Hou, K., Xue, C., Zhang, L., 2015. Digesting anomalies: an investment approach. Rev. Financ. Stud. 28, 650–705.
Hou, K., Xue, C., Zhang, L., 2020. Replicating anomalies. Rev. Financ. Stud. 33, 2019–2133.
Jegadeesh, N., 1990. Evidence of predictable behavior of security returns. J. Finance 45, 881–898.
Jegadeesh, N., 2000. Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation: discussion. J. Finance 55, 1765–1770.
Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. J. Finance 48, 65–91.
Jiang, J., Kelly, B.T., Xiu, D., 2022. (Re-)imag(in)ing price trend. J. Finance. https://doi.org/10.1111/jofi.13268. Forthcoming.
Kelly, B.T., Pruitt, S., Su, Y., 2019. Characteristics are covariances: a unified model of risk and return. J. Financ. Econ. 134, 501–524.
Kirby, C., 2020. Firm characteristics, cross-sectional regression estimates, and asset pricing tests. Rev. Asset Pricing Stud. 10, 290–334.
Kozak, S., Nagel, S., Santosh, S., 2020. Shrinking the cross-section. J. Financ. Econ. 135, 271–292.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25, 1097–1105.
Krogh, A., Vedelsby, J., 1995. Validation, and active learning. Adv. Neural Inf. Process. Syst. 7 (7), 231–238.
LeBaron, B., 1999. Technical trading rule profitability and foreign exchange intervention. J. Int. Econ. 49, 125–143.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
Lehmann, B.N., 1990. Fads, martingales, and market efficiency. Q. J. Econ. 105, 1–28.
Lettau, M., Pelger, M., 2020. Factors that fit the time series and cross-section of stock returns. Rev. Financ. Stud. 33, 2274–2325.
Levich, R.M., Thomas III, L.R., 1993. The significance of technical trading-rule profits in the foreign exchange market: a bootstrap approach. J. Int. Money Financ. 12, 451–474.
Lewellen, J., Nagel, S., 2006. The conditional CAPM does not explain asset-pricing anomalies. J. Financ. Econ. 82, 289–314.
Linnainmaa, J.T., Roberts, M.R., 2018. The history of the cross-section of stock returns. Rev. Financ. Stud. 31, 2606–2649.
Lo, A.W., Hasanhodzic, J., 2010. The Heretics of Finance: Conversations with Leading Practitioners of Technical Analysis, vol. 16. John Wiley and Sons.
Lo, A.W., MacKinlay, A.C., 1990. Data-snooping biases in tests of financial asset pricing models. Rev. Financ. Stud. 3, 431–467.
Lo, A.W., Mamaysky, H., Wang, J., 2000. Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation. J. Finance 55, 1705–1765.
Martin, I., 2017. What is the expected return on the market? Q. J. Econ. 132, 367–433.
McLean, R.D., Pontiff, J., 2016. Does academic research destroy stock return predictability? J. Finance 71, 5–32.
Menkhoff, L., 2010. The use of technical analysis by fund managers: international evidence. J. Bank. Finance 34, 2573–2586.
Messmer, M., 2017. Deep learning and the cross-section of expected stock returns. Available at SSRN: https://ssrn.com/abstract=3081555.
Messmer, M., Audrino, F., 2020. The lasso and the factor zoo - expected returns in the cross-section. Available at SSRN: https://ssrn.com/abstract=2930436.
Moritz, B., Zimmermann, T., 2016. Tree-based conditional portfolio sorts: the relation between past and future stock returns. Available at SSRN: https://ssrn.com/abstract=2740751.
Moskowitz, T.J., Ooi, Y.H., Pedersen, L.H., 2012. Time series momentum. J. Financ. Econ. 104, 228–250.
Neely, C., Weller, P., Dittmar, R., 1997. Is technical analysis in the foreign exchange market profitable? A genetic programming approach. J. Financ. Quant. Anal. 32, 405–426.
Neely, C.J., 2002. The temporal pattern of trading rule returns and exchange rate intervention: intervention does not generate technical trading profits. J. Int. Econ. 58, 211–232.
Neely, C.J., Rapach, D.E., Tu, J., Zhou, G., 2014. Forecasting the equity risk premium: the role of technical indicators. Manag. Sci. 60, 1772–1791.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Novy-Marx, R., Velikov, M., 2016. A taxonomy of anomalies and their trading costs. Rev. Financ. Stud. 29, 104–147.
Osler, C.L., 2003. Currency orders and exchange rate dynamics: an explanation for the predictive success of technical analysis. J. Finance 58, 1791–1819.
Rapach, D.E., Strauss, J.K., Zhou, G., 2013. International stock return predictability: what is the role of the United States? J. Finance 68, 1633–1662.
Ready, M.J., 2002. Profits from technical trading rules. Financ. Manag. 31, 43–61.
Rossi, A.G., 2018. Predicting stock market returns with machine learning. Working Paper.
Schwert, G.W., 2003. Anomalies and market efficiency. In: Constantinides, G., Harris, M., Stulz, R.M. (Eds.), Handbook of the Economics of Finance. Elsevier Science, B.V., Netherlands, pp. 939–974. Chapter 15.
Shumway, T., 1997. The delisting bias in CRSP data. J. Finance 52, 327–340.
Sullivan, R., Timmermann, A., White, H., 1999. Data-snooping, technical trading rule performance, and the bootstrap. J. Finance 54, 1647–1691.
Sutskever, I., Vinyals, O., Le, Q.V., 2014. Sequence to sequence learning with neural networks. Adv. Neural Inf. Process. Syst. 27.
Sweeney, R.J., 1986. Beating the foreign exchange market. J. Finance 41, 163–182.
Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G.D., et al., 2014. XSEDE: accelerating scientific discovery. Comput. Sci. Eng. 16, 62–74.
Zhu, Y., Zhou, G., 2009. Technical analysis: an asset allocation perspective on the use of moving averages. J. Financ. Econ. 92, 519–544.