Deep Learning in Asset Pricing

Luyang Chen[*], Markus Pelger[†] and Jason Zhu[‡]

This draft: June 12, 2019
First draft: March 15, 2019

Abstract

We propose a novel approach to estimate asset pricing models for individual stock returns that takes advantage of the vast amount of conditioning information, while keeping a fully flexible form and accounting for time-variation. Our general non-linear asset pricing model is estimated with deep neural networks applied to all U.S. equity data combined with a substantial set of macroeconomic and firm-specific information. We estimate the stochastic discount factor that explains all asset returns from the conditional moment constraints implied by no-arbitrage. Our asset pricing model outperforms out-of-sample all other benchmark approaches in terms of Sharpe ratio, explained variation and pricing errors. We trace its superior performance to including the no-arbitrage constraint in the estimation and to accounting for macroeconomic conditions and non-linear interactions between firm-specific characteristics. Our generative adversarial network enforces no-arbitrage by identifying the portfolio strategies with the most pricing information. Our recurrent Long-Short-Term-Memory network finds a small set of hidden economic state processes. A feedforward network captures the non-linear effects of the conditioning variables. Our model allows us to identify the key factors that drive asset prices and generate profitable investment strategies.

Keywords: No-arbitrage, stock returns, conditional asset pricing model, non-linear factor model, machine learning, deep learning, neural networks, big data, hidden states, GMM
JEL classification: C14, C38, C55, G12

[*]Institute for Computational and Mathematical Engineering, Stanford University, Email: lych@stanford.edu.
[†]Department of Management Science & Engineering, Stanford University, Email: mpelger@stanford.edu.
[‡]Department of Management Science & Engineering, Stanford University, Email: jzhu121@stanford.edu.
We thank Svetlana Bryzgalova, John Cochrane, Robert Hodrick, Kay Giesecke, Serhiy Kozak, Martin Lettau and Neil Shephard and seminar and conference participants at Stanford, the Western Mathematical Finance Conference and INFORMS for helpful comments. We thank the China Merchants Bank for generous research support.

I. Introduction

The most fundamental question in asset pricing is to understand why different assets have different average returns. No-arbitrage pricing theory provides a clear answer - expected returns differ because assets have different exposure to systematic risk. All pricing information is summarized in the stochastic discount factor (SDF) or pricing kernel. The empirical quest in asset pricing for the last 40 years was to estimate a stochastic discount factor that can explain expected returns of all assets. There are four major challenges that the literature so far has struggled to overcome in a single model: (1) The SDF could by construction depend on all available information, which means that the SDF is a function of a potentially very large set of variables. (2) The functional form of the SDF is unknown and likely complex. (3) The SDF can have a complex dynamic structure and the risk exposure for individual assets can vary over time depending on economic conditions and changes in asset-specific attributes. (4) The risk premium of individual stocks has a low signal-to-noise ratio, which complicates the estimation of an SDF that explains the expected returns of all stocks.

In this paper we estimate a general non-linear asset pricing model with deep neural networks for all U.S. equity data based on a substantial set of macroeconomic and firm-specific information. Our crucial innovation is the use of the no-arbitrage condition as part of the neural network algorithm. We estimate the stochastic discount factor that explains all stock returns from the conditional moment constraints implied by no-arbitrage. The use of machine learning techniques like deep neural networks is a natural idea to deal with the high dimensionality of the problem. One crucial insight of our work is that it is essential to incorporate economic conditions into the machine learning problem. Including the no-arbitrage constraint in the learning algorithm significantly improves the risk premium signal and makes it possible to explain individual stock returns. Empirically our general model outperforms out-of-sample the leading benchmark approaches and provides a clear insight into the structure of the pricing kernel and the sources of systematic risk.

Asset pricing and optimal investment are just two sides of the same coin and the results of this paper are also relevant for optimal portfolio investment. Solving for the SDF is actually equivalent to obtaining the conditional mean-variance efficient portfolio. Furthermore, exposure to the SDF should predict future expected returns, which can be directly incorporated in trading strategies.[1] Finally, mispriced assets correspond to arbitrage opportunities that can be exploited.

Our estimation approach combines no-arbitrage pricing and three neural network structures in a novel way. It considers four key elements concurrently: First, we can explain the general functional form of the SDF as a function of the information set using a feedforward neural network. Second, we capture the time-variation of the SDF on macroeconomic conditions with a recurrent Long-Short-Term-Memory (LSTM) network that identifies a small set of macroeconomic state processes. Third, a generative adversarial network identifies the states and portfolios with the most unexplained pricing information which allows us to price all assets. Fourth, the no-arbitrage constraint helps

[1] Stocks with a higher conditional correlation with the SDF should have higher expected returns.

to separate the risk premium signal from the noise and serves as a regularization to identify the relevant pricing information.

The special case of our model that restricts the SDF to a linear functional form is a linear factor models with time-varying loadings. In this case our estimator selects a linear combination of long-short factors based on firm characteristics. The loadings to this SDF factor are a linear function of time-varying characteristics where our model selects the best instruments to capture the time variation. However, our model allows for a more general functional form that captures arbitrary non-linearities and interaction effects of the factor portfolio weights and loadings as a function of the time-varying characteristics and macroeconomic variables. We show empirically that this general functional form significantly improves the model's ability to explain the cross-section of expected returns out-of-sample.

Our paper makes several methodological contributions. First, estimating the SDF from the fundamental no-arbitrage moment equation is conceptionally a generalized method of moment (GMM) problem. The conditional asset pricing moments imply an infinite number of moment conditions. Our generative adversarial approach provides a method to find and select the most relevant moment conditions from an infinite set of candidate moments. Second, the SDF depends on the dynamic time series structure of a large number of potentially non-stationary time series. Our LSTM approach summarizes the dynamics of a large number of time series in a small number of economic states. It serves the purpose of finding hidden states in the time series, summarizing them in a small number of state processes and applying the most appropriate transformation to the non-stationary time series in a data-driven way. Third, the no-arbitrage condition identifies the components of the pricing kernel that carry a high risk premia but have only a weak variation signal. Intuitively, most machine learning methods in finance[2] fit a model that can explain as much variation as possible, which is essentially a second moment object. The no-arbitrage condition is based on explaining the risk premia, which is based on a first moment. We can decompose stock returns into a predictable risk premium part and an unpredictable martingale component. Most of the variation is driven by the unpredictable component that does not carry a risk premium. When considering average returns the unpredictable component is diversified away over time and the predictable risk premium signal is strengthened. However, the risk premia of individual stocks is time-varying and an unconditional mean of stock returns might not capture the predictable component. Therefore, we consider unconditional means of stock returns instrumented with all possible combinations of firm specific characteristics and macroeconomic information. This serves the purpose of pushing up the risk premium signal while taking into account the time-variation in the risk premium.

Our empirical analysis is based on a data set of all available U.S. stocks from CRSP with monthly returns from 1967 to 2016 combined with 46 time-varying firm-specific characteristics and 178 macroeconomic time series. It includes the most relevant pricing anomalies and forecasting variables for the equity risk premium. Our approach outperforms out-of-sample all other bench-

[2]These models include Gu et al. (2018), Messmer (2017), Feng et al. (2018b) or Kelly et al. (2018).

mark approaches, which include linear models and deep neural networks that forecast risk premia instead of solving a GMM type problem. We compare the models out-of-sample with respect to the Sharpe Ratio implied by the pricing kernel, the explained variation and explained average returns of individual stocks. Our model has an annual out-of-sample Sharpe Ratio of 2.6 compared to 1.7 for the linear special case of our model, 1.5 for the deep learning forecasting approach and 0.8 for the Fama-French five factor model. At the same time we can explain 8% of the variation of individual stock returns and explain 23% of the expected returns of individual stocks, which is substantially larger than the other benchmark models. On standard test assets based on single- and double-sorted anomaly portfolios, our asset pricing model reveals an unprecedented pricing performance. In fact, on all 46 anomaly sorted decile portfolios we achieve a cross-sectional R^2 higher than 90%.

Our empirical findings are eleven-fold: First, because of their ability to fit flexible functional forms with many covariates, deep neural network can provide better asset pricing models. However, off-the-shelf simple prediction approaches perform worse than even linear no-arbitrage models. It is the crucial innovation to incorporate the economic constraint in the learning algorithm that allows us to detect the underlying SDF structure.

Second, linear models, which are the workhorse models in asset pricing, perform surprisingly well. We find that when considering firm-specific characteristics in isolation, the SDF depends approximately linearly on most characteristics. This explains why specific linear risk factors work well on certain single-sorted portfolios.

Third, non-linearities matter for interactions between covariates. The strength of the flexible functional form of deep neural networks reveals itself when considering the interaction between several characteristics. Although in isolation firm characteristics have a close to linear effect on the SDF, the multi-dimensional functional form is complex. Linear models and also non-linear models that assume an additive structure in the characteristics (e.g. additive splines or kernels) rule out interaction effects and cannot capture this structure.

Fourth, macroeconomic states matter. Macroeconomic time series data have a low dimensional "factor" structure, which can be captured by four hidden state processes. The SDF structure depends on these economic states that are closely related to business cycles and times of economic crises. In order to find these states we need to take into account the whole time series dynamics of all macroeconomic variables. The conventional approach to deal with non-stationary macroeconomic time series is to use differenced data that capture changes in the time series. However, using only the changes as an input loses all dynamic information and renders the macroeconomic time series essentially useless. Even worse, including conventionally standardized macroeconomic variables as predictors leads to worse performance than leaving them out overall, because they have lost most of their informational content and make it harder to separate the signal from the noise.

Fifth, estimating a pricing model on individual stocks leads to a superior pricing model on portfolios. Our model can almost perfectly explain expected returns on standard test assets, e.g. size and book-to-market single- or double-sorted portfolios. In fact, our model has an excellent pricing performance on all 46 anomaly sorted decile portfolios with a cross-sectional R^2 higher than 90% on each of them, outperforming all other benchmark models.

Sixth, the SDF structure is surprisingly stable over time. We estimate the functional form of the SDF with the data from 1967 to 1986, which has an excellent out-of-sample performance for the test data from 1992 to 2016. The risk exposure to the SDF for individual stocks can vary over time because the firm-specific characteristics and macroeconomic variables are time-varying. However, the functional form of the SDF and the risk exposure with respect to these covariates does not change.

Seventh, the most relevant pricing information are price trends and liquidity. All benchmark models agree on the variable categories. However, the functional form of how the characteristics affect the pricing kernel varies among different models.

Eighth, our asset pricing model yields highly profitable investment strategies. The mean-variance efficient portfolio implied by the pricing kernel has an out-of-sample Sharpe Ratio of 2.6. At the same time, other risk measures like maximum loss or drawdown are smaller than for the other benchmark models. The results are qualitatively robust to considering only large capitalization stocks, which suggests that illiquidity should not be a concern. Turnover as a proxy for the trading costs is less or similar to the linear models and the forecasting approach.

Ninth, our estimation is only based on the fundamental no-arbitrage moments. However, our model can explain more variation out-of-sample than a comparable model with the objective to maximize explained variation. This illustrates that the no-arbitrage condition disciplines the model and yields better results among all dimensions.

Tenth, our model yields a one factor model with time-varying loadings for individual stocks. It is not necessary to use the diversion of multiple linear or non-linear risk factors. The exposure to this SDF factor has predictive power for future returns as we demonstrate with portfolios that are sorted according to the SDF exposure.

Eleventh, our general GMM formulation includes essentially all other asset pricing models as a special case. It allows us to understand the incremental effect of restrictive model assumptions. For example imposing a linear structure yields a conventional linear factor model. Conditioning on kernel functions based on size and book-to-market ratio corresponds to pricing the conventional Fama-French double-sorted portfolios. Our results suggest that increasing the space of test assets (or equivalently having no-arbitrage moments conditioned on more characteristics) is actually more relevant than the flexible functional form of the SDF. However, to fully capture the SDF the flexible functional form and the relevant test assets are necessary.

Our paper contributes to an emerging literature that uses machine learning methods for asset pricing. In their pioneering work Gu et al. (2018) conduct a comparison of machine learning methods for predicting the panel of individual US stock returns. Their estimates of the expected risk premia of stocks map into a cross-sectional asset pricing model. We use their best prediction model based on deep neural networks as a benchmark model in our analysis. We show that including the no-arbitrage constraint leads to better results in asset pricing and explained variation than a simple prediction approach. Furthermore, we clarify that it is essential to identify the dynamic pattern in macroeconomic time series before feeding them into a machine learning model and we are the first to do this in an asset pricing context. Messmer (2017) and Feng et al. (2018a) follow a similar approach as Gu et al. (2018) to predict stock returns with neural networks. Bianchi et al. (2019) provide a comparison of machine learning method for predicting bond returns in the spirit of Gu et al. (2018).[3] Feng et al. (2018b) impose a no-arbitrage constraint by using a set of pre-specified linear asset pricing factors and estimate the risk loadings with a deep neural network.[4] Rossi (2018) uses Boosted Regression Trees to form conditional mean-variance efficient portfolios based on the market portfolio and the risk-free asset. Our approach also yields the conditional mean-variance efficient portfolio, but based on all stocks. Gu et al. (2019) extend the linear conditional factor model of Kelly et al. (2018) to a non-linear factor model using an autoencoder neural network.[5]

The workhorse models in asset pricing in equity are based on linear factor models exemplified by Fama and French (1993) and Fama and French (2015). Recently, new methods have been developed to study the cross-section of returns in the linear framework but accounting for the large amount of conditioning information. Lettau and Pelger (2018) extend principal component analysis to account for no-arbitrage. They show that a no-arbitrage penalty term makes it possible to overcome the low signal-to-noise ratio problem in financial data and find the information that is relevant for the pricing kernel. Our paper is based on a similar intuition and we show that this result extends to a non-linear framework. Kozak et al. (2018) apply mean-variance optimization with an elastic net penalty to characteristic sorted factors.[6] Kelly et al. (2018) apply PCA to stock returns projected on characteristics to obtain a SDF that is linear in the characteristics. Pelger (2019) combines high-frequency data with PCA to capture non-parametrically the time-variation in factor risk. Pelger and Xiong (2018b) show that macroeconomic states are relevant to capture time-variation in PCA-based factors. Freyberger et al. (2017) use Lasso selection methods to approximate the SDF as a non-linear function of characteristics but rule out interaction effects.

Our approach uses the same fundamental insight as Bansal and Viswanathan (1993) who propose using the conditional GMM equations to estimate the SDF, but restrict themselves to a small number of conditioning variables. In order to deal with the infinite number of moment conditions we extend the classcial GMM setup of Hansen (1982) and Chamberlain (1987) by an adversarial

[3]Other related work includes Sirignano et al. (2016) who estimate mortgage prepayments, delinquencies, and foreclosures with deep neural networks, Moritz and Zimmerman (2016) who apply tree-based models to portfolio sorting and Heaton et al. (2017) who automate portfolio selection with a deep neural network. Horel and Giesecke (2019) propose a significance test in neural networks and apply it to house price valuation.

[4]Their analysis considers various sets of sorted portfolios but is not applied to individual stock returns.

[5]The intuition behind their and our approach can be best understood when considering the linear special cases. Our approach can be viewed as a non-linear generalization of Kozak et al. (2018) with the additional elements of finding the macroeconomic states and identifying the most robust conditioning instruments. Fundamentally, our object of interest is the pricing kernel. Kelly et al. (2018) obtain a multi-factor factor model that maximizes the explained variation. The linear special case applies PCA to a set of characteristic based factors to obtain a linear lower dimensional factor model, while their more general autoencoder obtains the loadings to characteristic based factors that can depend non-linearly on the characteristics.

[6]We show that the special case of a linear formulation of our model is essentially a version of their model and we include it as the linear benchmark case in our analysis.

network to select the optimal moment conditions. A similar idea has been proposed by Lewis and Syrgkanis (2018) for non-parametric instrumental variable regressions. Our problem is also similar in spirit to the Wasserstein GAN in Arjosvky et al. (2017) that provides a robust fit to moments. The Generative Adversarial Network approach was first proposed by Goodfellow et al. (2014) for image recognition. In order to find the hidden states in macroeconomic time series we propose the use of Recurrent Neural Networks with Long-Short-Term-Memory (LSTM). LSTMs are designed to find patterns in time series data and have been first proposed by Hochreiter and Schmidhuber (1997). They are among the most successful commercial AIs and are heavily used for sequences of data such as speech (e.g. Google with speech recognition for Android, Apple with Siri and the "QuickType" function on the iPhone or Amazon with Alexa).

The rest of the paper is organized as follows. Section II introduces the model framework and Section III elaborates on the estimation approach. Section IV provides some intuition for our estimator in a simulation setup. The empirical results are collected in Section V. Section VI concludes. The Internet Appendix collects additional empirical results.

II. Model

A. No-Arbitrage Asset Pricing

Our goal is to explain the differences in the cross-section of returns R for individual stocks. Let $R_{t+1,i}$ denote the return of asset i at time $t+1$. The fundamental no-arbitrage assumption is equivalent to the existence of a stochastic discount factor (SDF)[7] such that for any return in excess of the risk-free rate $R^e_{t+1,i} = R_{t+1,i} - R^f_{t+1}$, it holds

$$\begin{aligned} \mathbb{E}_t\left[M_{t+1}R^e_{t+1,i}\right] = 0 \qquad &\Leftrightarrow \\ \mathbb{E}_t[R^e_{t+1,i}] = \underbrace{\left(-\frac{\mathrm{Cov}_t(R^e_{t+1,i}, M_{t+1})}{\mathrm{Var}_t(M_{t+1})}\right)}_{\beta_{t,i}} &\cdot \underbrace{\frac{\mathrm{Var}_t(M_{t+1})}{\mathbb{E}_t[M_{t+1}]}}_{\lambda_t}, \end{aligned}$$

where $\beta_{t,i}$ is the exposure to systematic risk and λ_t is the price of risk. $E_t[.]$ denotes the expectation conditional on the information at time t. The stochastic discount factor is an affine transformation of the tangency portfolio [8]. Without loss of generality[9] we consider the SDF formulation

$$M_{t+1} = 1 - \sum_{i=1}^{N} \omega_{t,i} R^e_{t+1,i} = 1 - \omega_t^\top R^e_{t+1}.$$

[7]also labeled as pricing kernel or change of measure to the martingale measure.

[8] See Back (2010)

[9]As we work with excess returns we have an additional degree of freedom. Following Cochrane (2003) we use the following normalized relationship between the SDF and the mean-variance efficient portfolio. We consider the SDF based on the projection on the asset space.

The fundamental pricing equation $\mathbb{E}_t[R^e_{t+1}M_{t+1}] = 0$ implies the SDF weights

$$\omega_t = \mathbb{E}_t[R^e_{t+1}{R^e_{t+1}}^\top]^{-1}\mathbb{E}_t[R^e_{t+1}], \tag{1}$$

which are the portfolio weights of the conditional mean-variance efficient portfolio.[10] We define the tangency portfolio as the SDF factor $F_{t+1} = \omega_t^\top R^e_{t+1}$. The asset pricing equation can now be formulated as

$$\begin{aligned}\mathbb{E}_t[R^e_{t+1,i}] &= \frac{\mathrm{Cov}_t(R^e_{t+1,i}, F_{t+1})}{\mathrm{Var}_t(F_{t+1})} \cdot \mathbb{E}_t[F_{t+1}] \\ &= \beta_{t,i}\mathbb{E}_t[F_{t+1}].\end{aligned}$$

Hence, no-arbitrage implies a one-factor model

$$R^e_{t+1,i} = \beta_{t,i}F_{t+1} + \epsilon_{t+1,i}$$

with $\mathbb{E}_t[\epsilon_{t+1,i}] = 0$ and $\mathrm{Cov}_t(F_{t+1}, \epsilon_{t+1,i}) = 0$. Conversely, the factor model formulation implies the stochastic discount factor formulation above. Furthemore, if the idiosyncratic risk $\epsilon_{t+1,i}$ is diversifiable and the SDF factor is systematic,[11] then knowledge of the risk loadings is sufficient to construct the SDF factor:

$$\left(\beta_t^\top \beta_t\right)^{-1} \beta_t^\top R^e_{t+1} = F_{t+1} + \left(\beta_t^\top \beta_t\right)^{-1} \beta_t^\top \epsilon_{t+1} = F_{t+1} + o_p(1).$$

The fundamental problem is to find the SDF portfolio weights ω_t and risk loadings β_t. Both are time-varying and general functions of the information set at time t. The knowledge of ω_t and β_t solves three problems: (1) We can explain the cross-section of individual stock returns. (2) We can construct the mean-variance efficient tangency portfolio. (3) We can decompose stock returns into their predictable systematic component and their non-systematic unpredictable component.

B. Generative Adversarial Methods of Moments

Finding the SDF weights is equivalent to solving a method of moment problem. The conditional no-arbitrage moment condition implies infinitely many unconditional moment conditions

$$\mathbb{E}[M_{t+1}R^e_{t+1,i}g(I_t, I_{t,i})] = 0$$

for any function $g(.) : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^d$, where $I_t \times I_{t,i} \in \mathbb{R}^{p+q}$ denotes all the variables in the information set at time t and d is the number of moment conditions. We denote by I_t all q macroeconomic conditioning variables that are not asset specific, e.g. inflation rates or the market

[10] Any portfolio on the globally efficient frontier achieves the maximum Sharpe Ratio. These portfolio weights represent one possible efficient portfolio.

[11] Denote the conditional residual covariance matrix by $\Sigma_t^\epsilon = \mathrm{Var}_t(\epsilon_t)$. Then, sufficient conditions are $\|\Sigma_t^\epsilon\|_2 < \infty$ and $\frac{\beta^\top\beta}{N} > 0$ for $N \to \infty$ i.e. Σ_t^ϵ has bounded eigenvalues and β_t has sufficiently many non-zero elements.

return, while $I_{t,i}$ are p firm-specific characteristics, e.g. the size or book-to-market ratio of firm i at time t. The unconditional moment condition can be interpreted as the pricing error for a choice of portfolios and times determined by $g(.)$. The challenge lies in finding the relevant moment conditions to identify the SDF.

A well-known formulation includes 25 moments that corresponds to pricing the 25 size and value double sorted portfolios of Fama and French (1993). For this special case each g corresponds to an indicator function if the size and book-to-market values of a company are in a specific quantile. Another special case is to consider only unconditional moments, i.e. setting g to a constant. This corresponds to minimizing the unconditional pricing error of each stock.

The SDF portfolio weights $\omega_{t,i} = \omega(I_t, I_{t,i})$ and risk loadings $\beta_{t,i} = \beta(I_t, I_{t,i})$ are general functions of the information set, i.e.

$$\omega : \mathbb{R}^p \times \mathbb{R}^q \to R \qquad \beta : \mathbb{R}^p \times \mathbb{R}^q \to R.$$

For example, the SDF weights and loadings in the Fama-French 3 factor model are a special case, where both functions are approximated by a two-dimensional kernel function that depends on the size and book-to-market ratio of firms. The Fama-French 3 factor model only uses firm-specific information but no macroeconomic information, e.g. the loadings cannot vary based on the state of the business cycle.

We use an adversarial approach to select the moment conditions that lead to the largest mis-pricing:

$$\min_{\omega} \max_{g} \frac{1}{N} \sum_{j=1}^{N} \left\| \mathbb{E}\left[\left(1 - \sum_{i=1}^{N} \omega(I_t, I_{t,i}) R^e_{t+1,i} \right) R^e_{t+1,j} g(I_t, I_{t,j}) \right] \right\|^2, \tag{2}$$

where the function ω and g are normalized functions chosen from a specified functional class. This is a minimax loss minimization problem. These types of problems can be modeled as a zero-sum game, where one player, the asset pricing modeler, wants to choose an asset pricing model, while the adversary wants to choose conditions under which the asset pricing model performs badly. This can be interpreted as first finding portfolios or times that are the most mis-priced and then correcting the asset pricing model to also price these assets. The conventional GMM approach assumes a finite number of moments that identify a finite dimensional parameter. The moments are selected to achieve the most efficient estimator within this class. Our problem is different in two ways that rule out using the same approach. First, we have an infinite number of candidate moments without the knowledge of which moments identify the parameters. Second, our parameter set is also of infinite dimension, and we consequently do not have an asymptotic normal distribution with a feasible estimator of the covariance matrix. In contrast, our approach selects the moments based on robustness.[12]

[12]See Blanchet et al. (2016) for a discussion on robust estimation with an adversarial approach.

Note, that our moment conditions allow the SDF weights to be general functions of the information set, while equation 1 gives an explicit solution in terms of the conditional second and first moment of stock returns. Without strong parametric assumptions it is practically not possible to estimate the inverse of a large dimensional conditional covariance matrix for stocks reliably. Even in the unconditional setup the estimation of the inverse of a large dimensional covariance matrix is already challenging. In order to avoid this problem, we do not explicitly impose these restrictions on the SDF weights. Instead we use the insight that if our SDF explains the unconditional moments in equation 2 for any choice of g, then it must also satisfy equation 1. The generative adversarial formulation allows us to side-step solving explicitly an infeasible mean-variance optimization. In other words, if we allow the SDF weights to be general functions of the information set, but require this SDF to explain prices of any possible portfolio, then this SDF factor has to correspond to the conditional mean-variance efficient portfolio.

Once we have obtained the SDF factor weights, the loadings are proportional to the conditional moment $\mathbb{E}_t[F_{t+1}R^e_{t+1,i}]$. A key element of our approach is to avoid estimating directly conditional means of stock returns. Our empirical results show that we can better estimate the conditional co-movement of stock returns with the SDF factors, which is a second moment, than the conditional first moment. Note, that in the no-arbitrage one-factor model, the loadings are proportional to $\mathrm{Cov}_t(R^e_{t+1,i}, F_{t+1})$ and $\mathbb{E}_t[F_{t+1}R^e_{t+1,i}]$, where the last one has the advantage that we avoid estimating the first conditional moment.

C. *Alternative Models*

Instead of minimizing the violation of the no-arbitrage condition, one can directly estimate the conditional mean. Note that the conditional expected returns $\mu_{t,i}$ are proportional to the loadings in the one factor formulation:

$$\mu_{t,i} := \mathbb{E}_t[R^e_{t+1,i}] = \beta_{t,i}\mathbb{E}_t[F_{t+1}].$$

Hence, up to a time-varying proportionality constant the SDF factor weights and loadings are equal to $\mu_{t,i}$. This reduces the cross-sectional asset pricing problem to a simple forecasting problem. Hence, we can use the forecasting approach pursued in Gu et al. (2018) for asset pricing.

Another benchmark model that we consider assumes a linear structure in the factor portfolio weights $\omega_{t,i} = \theta^\top I_{t,i}$ and linear conditioning in the no-arbitrage moment condition:

$$\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}\left[\left(1-\frac{1}{N}\sum_{i=1}^{N}\theta^\top I_{t,i}R^e_{t+1,i}\right)R^e_{t+1,j}I_{t,j}\right]=0 \qquad \Leftrightarrow \qquad \mathbb{E}\left[\left(1-\theta^\top\tilde{F}_{t+1}\right)\tilde{F}^\top_{t+1}\right]=0,$$

where $\tilde{F}_{t+1} = \frac{1}{N}\sum_{i=1}^{N} I_{t,i}R^e_{t+1,i}$ are q characteristic managed factors. Such characteristic managed factors based on linearly projecting onto quantiles of characteristics are exactly the input to PCA in

Kelly et al. (2018) or the elastic net mean-variance optimization in Kozak et al. (2018).[13] The solution to minimizing the sum of squared errors in these moment conditions is a simple mean-variance optimization for the q characteristic managed factors i.e $\theta = \left(\mathbb{E}\left[\tilde{F}_{t+1}\tilde{F}_{t+1}^{\top}\right]\right)^{-1}\mathbb{E}\left[\tilde{F}_{t+1}\right]$ are the weights of the tangency portfolio based on these factors.[14] We choose this specific linear version of the model as it maps directly into the linear approaches that have already been successfully used in the literature. This linear framework essentially captures the class of linear factor models. For comparison we will also include the conventional Fama-French 3 and 5 factor models from Kenneth French's website.

III. Estimation

A. *Loss Function and Model Architecture*

The empirical loss function of our model minimizes the weighted sample moments which can be interpreted as weighted sample mean pricing errors:

$$L(\omega|\hat{g}, I_t, I_{t,i}) = \frac{1}{N}\sum_{i=1}^{N}\frac{T_i}{T}\left\|\frac{1}{T_i}\sum_{t\in T_i}M_{t+1}R^e_{t+1,i}\hat{g}(I_t, I_{t,i})\right\|^2. \tag{3}$$

for a given conditioning function $\hat{g}(.)$ and information set. We deal with an unbalanced panel in which the number of time series observations T_i varies for each asset. As the convergence rates of the moments under suitable conditions is $1/\sqrt{T_i}$, we weight each cross-sectional moment condition by $\sqrt{T_i}/\sqrt{T}$, which assigns a higher weight to moments that are estimated more precisely and down-weights the moments of assets that are observed only for a short time period.

For a given conditioning function $\hat{g}(.)$ and choice of information set the SDF factor portfolio weights are estimated by a feedforward network that minimizes the pricing error loss

$$\hat{\omega} = \min_{\omega} L(\omega|\hat{g}, I_t, I_{t,i}).$$

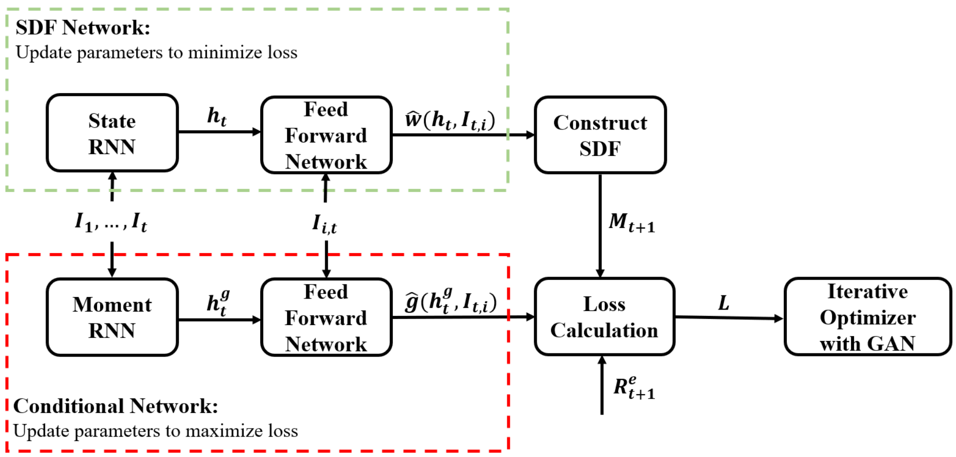
We refer to this network as the SDF network.

We construct the conditioning function $\hat{g}$ via a conditional network with a similar neural network architecture. The conditional network serves as an adversary and competes with the SDF network to identify the assets and portfolio strategies that are the hardest to explain. The macroeconomic information dynamics are summarized by macroeconomic state variables h_t which are obtained by a Recurrent Neural Network (RNN) with Long-Short-Term-Memory units. The model architecture is summarized in Figure 1 and each of the different components are described in detail in the next subsections.[15]

In contrast, forecasting returns similar to Gu et al. (2018) uses only a feedforward network and

[13] Kozak et al. (2018) consider also cross-products of the characteristics.
[14] As before we define as tangency portfolio one of the portfolios on the mean-variance efficient frontier.
[15] See Goodfellow et al. (2016) for a textbook treatment.

Figure 1. Model Architecture



Model architecture of GAN (Generative Adversarial Network) with RNN (Recurrent Neural Network) with LSTM cells.

is labeled as FFN. It estimates conditional means $\mu_{t,i} = \mu(I_t, I_{t,i})$ by minimizing the average sum of squared prediction errors:

$$\hat{\mu} = \min_{\mu} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{N_t} \sum_{i=1}^{N_t} \left(R_{t+1,i}^e - \mu(I_t, I_{t,i})\right)^2 .$$

We only include the best performing feedforward network from Gu et al. (2018)'s comparison study. Within their framework this model outperforms tree learning approaches and other linear and non-linear prediction models. In order to make the results more comparable with Gu et al. (2018) we follow the same procedure as outlined in their paper. Thus, the simple forecasting approach does not include an adversarial network or an RNN network to condense the macroeconomic dynamics.

B. *Feedforward Network (FFN)*

A feedforward network (FFN)[16] is a universal approximator that can learn any functional relationship between an input and output variable with sufficient data:

$$y = f(x).$$

We will consider four different FFN: For the covariates $x = [I_t, I_{t,i}]$ we estimate (1) the optimal weights in our GAN network ($y = \omega$), (2) the optimal instruments for the moment conditions in our GAN network ($y = g$), (3) the conditional mean return ($y = R_{t+1,i}^e$) and (4) the second moment ($y = R_{t+1,i}^e F_{t+1}$).

We start with a one-layer neural network. It combines the raw predictor variables (or features)

[16]FFN are among the simplest neural networks and treated in detail in standard machine learning textbooks, e.g. Goodfellow et al. (2016).

$x = x^{(0)} \in \mathbb{R}^{K^{(0)}}$ linearly and applies a non-linear transformation. This non-linear transformation is based on an element-wise operating activation function. We choose the popular function known as the rectified linear unit (ReLU)[17], which component-wise thresholds the inputs and is defined as

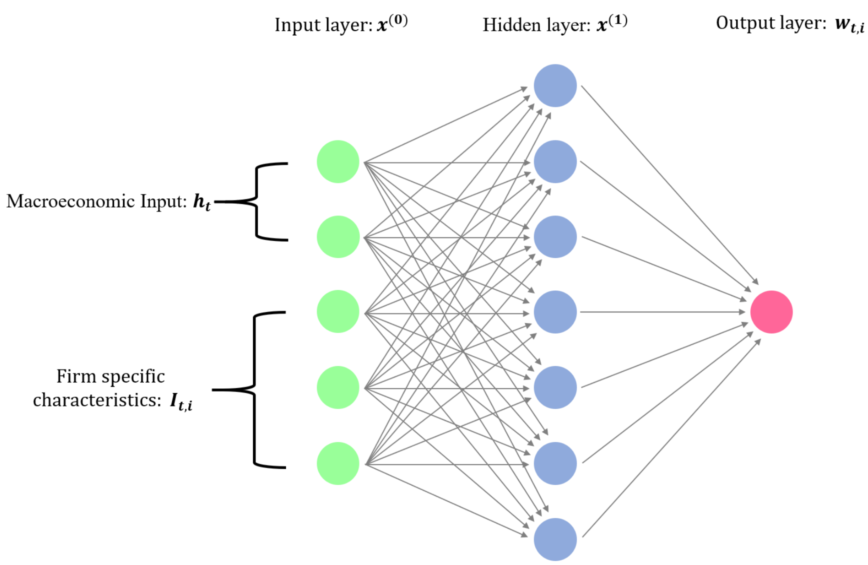
$$\text{ReLU}(x_k) = \max(x_k, 0).$$

The result is the hidden layer $x^{(1)} = (x_1^{(1)}, ..., x_{K^{(1)}}^{(1)})$ of dimension $K^{(1)}$ which depends on the parameters $W^{(0)} = (w_1^{(0)}, ..., w_{K^{(0)}}^{(0)})$ and the bias term $w_0^{(0)}$. The output layer is simply a linear transformation of the output from the hidden layer.

$$x^{(1)} = \text{ReLU}(W^{(0)\top}x^{(0)} + w_0^{(0)}) = \text{ReLU}\left(w_0^{(0)} + \sum_{k=1}^{K^{(0)}} w_k^{(0)} x_k^{(0)}\right)$$
$$y = W^{(1)\top}x^{(1)} + w_0^{(1)} \qquad \text{with } x^{(1)} \in \mathbb{R}^{K^{(1)}}, W^{(0)} \in \mathbb{R}^{K^{(1)} \times K^{(0)}}, W^{(1)} \in \mathbb{R}^{K^{(1)}}$$

Note, that without the non-linearity in the hidden layer, the one-layer network would reduce to a generalized linear model.

Figure 2. Feedforward Network with Single Hidden Layer



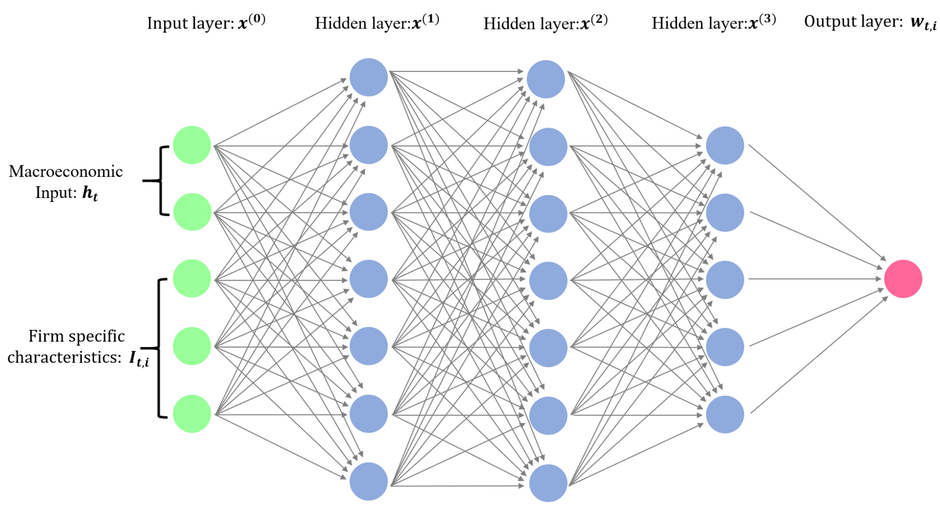
The deep neural network considers L layers as illustrated in Figure 3. Each hidden layers takes the output from the previous layer and transforms it into an output as

$$x^{(l)} = \text{ReLU}\left(W^{(l-1)\top}x^{(l-1)} + w_0^{(l-1)}\right) = \text{ReLU}\left(w_0^{(l-1)} + \sum_{k=1}^{K^{(l-1)}} w_k^{(l-1)} x_k^{(l-1)}\right)$$
$$y = W^{(L)\top}x^{(L)} + w_0^{(L)}$$

[17] Other activation functions include sigmoid, hyperbolic tangent function and leaky ReLU.

with hidden layer outputs $x^{(l)} = (x_1^{(l)}, ..., x_{K^{(l)}}^{(l)}) \in \mathbb{R}^{K^{(l)}}$ and parameters $W^{(l)} = (w_1^{(l)}, ..., w_{K^{(l)}}^{(l)}) \in \mathbb{R}^{K^{(l)} \times K^{(l-1)}}$ for $l = 0, ..., L-1$ and $W^{(L)} \in \mathbb{R}^{K^{(L)}}$.

Figure 3. Feedforward Network with 3 Hidden Layers



C. Recurrent Neural Network (RNN)

A Recurrent Neural Network (RNN) with Long-Short-Term-Memory (LSTM)[18] estimates the hidden macroeconomic state variables. Instead of directly passing macroeconomic variables I_t as features to the feedforward network, we apply a non-linear transformation to them with a specific Recurrent Neural Network. There are four reasons why this step is necessary. First, many macroeconomic variables themselves are not stationary. Although we perform transformations as suggested in McCracken and Ng (2016), some variables still seem to have non-stationary patterns. The RNN with LSTM will find the appropriate stationary transformation of the variables such that their dynamics explain asset prices. The LSTM can take first differences of the data if necessary but will also perform more complex transformations. Second, there is no reason to assume that the pricing kernel has a Markovian structure with respect to the macroeconomic information. For example business cycles can affect pricing but the GDP growth of the last period is insufficient to learn if the model is in a boom or a recession. Hence, lagged values of the macroeconomic state variables need to be included. A conventional Recurrent Neural Network can take into account the time series behavior but can encounter problems with exploding and vanishing gradients when considering longer time lags. This is why we use Long-Short-Term-Memory cells. Third, the LSTM is designed to find hidden state processes allowing for lags of unknown and potentially long duration in the time series, which makes it well-suited to detect business cycles. Fourth, the macroeconomic time series variables seem to be strongly dependent, i.e. there is redundant information. Although

[18]LSTM belong to the most successful machine learning methods for sequences of data. For example Apple uses LSTM for the "QuickType" function on the iPhone and for Siri, Amazon uses it for Amazon Alexa and Google for speech recognition.

the regularization in a neural network can in principle deal with redundant variables, the finite sample of stock returns is of a modest size compared to other machine learning applications. Hence, the large number of predictor variables proves to negatively impact the feedforward network performance. The RNN with LSTM summarizes the macroeconomic time series dynamics in a small number of hidden state processes that provide a more robust fit when used as an input for the feedforward network.

Recurrent Neural Networks are a family of neural networks for processing sequences of data. They transform a sequence of input variables to another output sequence, with the same set of parameters at each step. A vanilla RNN model takes the current input variable x_t and the previous hidden state h_{t-1} and performs a non-linear transformation to get the current state h_t.

$$h_t = \sigma(W_h h_{t-1} + W_x x_t + w_0),$$

where σ is the activation function. Intuitively, we can think of a vanilla RNN as non-linear generalization of an autoregressive process where the lagged variables are transformations of the lagged observed variables. This type of structure is powerful if only the immediate past is relevant, but it is not suitable if the time series dynamics are driven by events that are further back in the past. We use the more complex LSTM model to capture long-term dependencies. We can think of an LSTM as a flexible hidden state space model for a large dimensional system. The dynamics of the macroeconomic time series are driven by a small number of hidden states that aggregate cross-sectional and time series patterns. A popular approach to aggregate a cross-section of macroeconomic time series is principal component analysis.[19] This aggregates the time series to a small number of latent factors that explain the correlation in the innovations in the time series, but PCA cannot identify the current state of the economic system. On the other hand, state space models, with the simple linear Gaussian state space model estimated by a Kalman filter as one of the most popular ones, are usually set up for a small number of time series under restrictive distributional assumptions. Our LSTM approach can deal with both the large dimensionality of the system and a very general functional form of the states while allowing for long-term dependencies.

The LSTM is composed of a cell (the memory part of the LSTM unit) and three "regulators", called gates, of the flow of information inside the LSTM unit: an input gate, a forget gate, and an output gate. Intuitively, the cell is responsible for keeping track of the dependencies between the elements in the input sequence. The input gate controls the extent to which a new value flows into the cell, the forget gate controls the extent to which a value remains in the cell and the output gate controls the extent to which the value in the cell is used to compute the output activation of the LSTM unit.

We take $x_t = I_t$ as the input sequence of macroeconomic information, and the output is the state processes h_t. At each step, a new memory cell $\tilde{c}_t$ is created with current input x_t and previous

[19]See e.g. Ludvigson and Ng (2007).

hidden state h_{t-1}

$$\tilde{c}_t = \tanh(W_h^{(c)} h_{t-1} + W_x^{(c)} x_t + w_0^{(c)}).$$

The input and forget gates control the memory cell, while the output gate controls the amount of information stored in the hidden state:

$$\begin{aligned}
\text{input}_t &= \sigma(W_h^{(i)} h_{t-1} + W_x^{(i)} x_t + w_0^{(i)}) \\
\text{forget}_t &= \sigma(W_h^{(f)} h_{t-1} + W_x^{(f)} x_t + w_0^{(f)}) \\
\text{out}_t &= \sigma(W_h^{(o)} h_{t-1} + W_x^{(o)} x_t + w_0^{(o)}).
\end{aligned}$$

The sigmoid function σ is an element-wise non-linear transformation. Denoting the element-wise product by $\circ$, the final memory cell and hidden state are given by

$$\begin{aligned}
c_t &= \text{forget}_t \circ c_{t-1} + \text{input}_t \circ \tilde{c}_t \\
h_t &= \text{out}_t \circ \tanh(c_t).
\end{aligned}$$

We use the state processes h_t instead of the macroeconomic variables I_t as an input to our SDF network. Note, that for any $\mathcal{F}_t$-measurable sequence I_t, the output sequence h_t is again $\mathcal{F}_t$-measurable, so that the transformation creates no look-ahead bias. Furthermore, h_t contains all the macroeconomic information in the past, while I_t only uses current information. The flow chart A.2 summarizes the structure of the LSTM unit.

D. Generative Adversarial Network (GAN)

The conditioning function g is the output of a second feedforward network. Enlightened by Generative Adversarial Networks (GAN), we chose the moment conditions that lead to the largest pricing discrepancy by having two networks compete against each other. One network creates the SDF M_{t+1}, and the other network creates the conditioning function.

We take three steps to train the model. We first obtain an initial guess of the SDF by updating the SDF network to minimize the unconditional loss. Then for a given SDF network we maximize the loss by varying the parameters in the conditional network. Finally, we fix the parameters in the conditional network and train the SDF network to minimize the conditional loss.[20] The logic behind this idea is that by minimizing the largest conditional loss among all possible conditioning functions, the loss for any function is small. Note that both, the SDF network and the conditional network each use a FFN network combined with an LSTM that estimates the macroeconomic hidden state variables, i.e. instead of directly using I_t as an input each network summarizes the whole

[20] A conventional GAN network iterates this procedure until convergence. We find that our algorithm converges already after the above three steps, i.e. the model does not improve further by repeating the adversarial game.

macroeconomic time series information in the state process h_t (respectively h_t^g for the conditional network):[21]

$$\{\hat{\omega}, \hat{h}_t, \hat{g}, \hat{h}_t^g\} = \min_{\omega, h_t} \max_{g, h_t^g} L(\omega | \hat{g}, h_t^g, h_t, I_{t,i}).$$

E. Hyperparameters and Ensemble Learning

Due to the high dimensionality and non-linearity of the problem, training a deep neural network is a complex task. Here, we discuss the implementation in more detail.

For training deep neural networks the vanilla stochastic gradient descend method has proven to be not an efficient method. A better approach is to use optimization methods that introduce an adaptive learning rate.[22] We use Adam which is an algorithm for gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments to continuously adjust the learning rate. It is morel likely to escape saddle points and hence is more accurate, while also providing faster convergence.[23] The optimization depends on the normalization of the input variables. For all SDF models the characteristics and macroeconomic states are centered around their mean. [24]

Regularization is crucial and prevents the model from over-fitting on the training sample. Although l_1/l_2 regularization might also be used in training other neural networks, Dropout is preferable and generally results in better performances.[25] The term "Dropout" refers to dropping out units in a neural network as illustrated in Figure A.1. By dropping out a unit, we mean temporarily removing it from the network, along with all its incoming and outgoing connections with a certain probability. Dropout can be shown to be a form of ridge regularization and is only applied during the training[26]. When doing out-of-sample testing, we keep all the units and their connections.

We split the data into a training, validation and testing sample. The validation set is used to tune the hyperparameters, which are included in Table I. We choose the best configuration among all possible combinations of hyperparameters by maximizing the Sharpe Ratio of the SDF factor on the validation data set.[27] The hyperparameters of the model with the highest validation Sharpe Ratio are selected for the test data set.

We use ensemble averaging to create a group of models that provide a significantly more robust estimation. A distinguishing feature of neural networks is that the estimation results can depend on the starting value used in the optimization. The standard practice is to train the models separately with different initial values chosen randomly from a certain distribution. Although each model

[21] We allow for potentially different macroeconomic states for the SDF and the conditional network as the unconditional moment conditions that identify the SDF can depend on different states as the SDF weights.

[22] See e.g. Ruder (2016) and Kingma and Ba (2014).

[23] Other adaptive gradient descent methods include Adagrad or Adadelta.

[24] In a previous version of the paper we applied a different normalization. The results are qualitatively the same, but the benchmark performance for all models has improved under this new normalization.

[25] See e.g. Srivastava et al. (2014).

[26] See Wager et al. (2013)

[27] We have used different criteria functions, including the error in minimizing the moment conditions, to select the hyperparameters. The results are virtually identical and available upon request.

Table I Selection of Hyperparameters for GAN

Notation	Hyperparameters	Candidates	Optimal
HL	Number of layers in SDF Network	2, 3 or 4	2
HU	Number of hidden units in SDF Network	64	64
SMV	Number of hidden states in SDF Network	4 or 8	4
CSMV	Number of hidden states in Conditional Network	16 or 32	32
CHL	Number of layers in Conditional Network	0 or 1	0
CHU	Number of hidden units in Conditional Network	4, 8, 16 or 32	8
LR	Initial learning rate	0.001, 0.0005, 0.0002 or 0.0001	0.001
DR	Dropout	0.95	0.95

might have high variance, the variance can be reduced at no cost to the bias by averaging the outputs from these models. Let's denote $\hat{w}^{(j)}$ to be the optimal portfolio weights given by the j^{th} model. The ensemble model is a weighted average of the outputs from models with the same architecture but different starting values for the optimization and gives more robust estimates:[28] $\hat{\omega} = \frac{1}{9}\sum_{j=1}^{9}\hat{\omega}^{(j)}$. We also apply the ensemble method to the simple forecasting approach.

In summary, the hyperparameter selection works as follows: (1) First, for each possible combination of hyperparameters (384 models) we fit the GAN model. (2) Second, we select the four best combinations of hyperparameters on the validation data set. (3) Third, for each of the four combinations we fit 9 models with the same hyperparameters but different initialization. (4) Finally, we select the ensemble model with the best performance on the validation data set. Table I reports the tuning parameters of the best performing model. The feedforward network estimating the SDF weights has 2 hidden layers (HL) each of which has 64 nodes (HU). There are four hidden states (SMV) that summarize the macroeconomic dynamics in the LSTM network. The conditional adversarial network generates 8 moments (CHU) in a 0-layer (CHL) network. The macroeconomic dynamics for the conditional moments are summarized in 32 hidden states (CSMV). This conditional network essentially applies a non-linear transformation to the characteristics and the hidden macroeconomic states and then combines them linearly. The resulting moments can, for example, capture the pricing errors of long-short portfolios based on characteristic information or portfolios that only pay off under certain macroeconomic conditions.

The FFN for the forecasting approach uses the optimal hyperparameters selected by Gu et al. (2018) which is a 3-layer neural network with $[32, 16, 8]$ hidden units, dropout of 0.95 and a learning rate of 0.001. This has the additional advantage of making our results directly comparable to their results.

F. Model Comparison

We evaluate the performance of our model by calculating the Sharpe Ratio of the SDF factor, the amount of explained variation and the pricing errors of the model. We compare our GAN

[28] Averaging over 9 models has proven to provide very stable results. The results for a larger number of model averages are available upon request.

model, with a simple forecasting feedforward network model labeled as FFN, the linear special case of GAN labeled as LS and a regularized linear model labeled as EN.

The one factor representation yields three performance metrics to compare the different model formulations. First, the SDF factor is by construction on the globally efficient frontier and should have the highest conditional Sharpe Ratio. We use the unconditional Sharpe Ratio of the SDF factor portfolio $SR = \frac{\mathbb{E}[F]}{Var[F]}$ as a measure to assess the pricing performance of models. The second metric measures how much variation the SDF factor explains. The explained variation is defined as $1 - \frac{\sum_{i=1}^N \mathbb{E}[\epsilon_i^2]}{\sum_{i=1}^N \mathbb{E}[R_i^e]}$ where ϵ_i is the residual of a cross-sectional regression on the loadings. As in Kelly et al. (2018) we do not demean returns due to their non-stationarity and noise in the mean estimation. Our explained variation measure can be interpreted as a time series R^2. The third performance measure is the average pricing error normalized by the average mean return to obtain a cross-sectional R^2 measure $1 - \frac{\frac{1}{N}\sum_{i=1}^N \mathbb{E}[\epsilon_i]^2}{\frac{1}{N}\sum_{i=1}^N \mathbb{E}[R_i]^2}$.

The output for our GAN model are the SDF factor weights $\hat{\omega}_{GAN}$. We obtain the risk exposure $\hat{\beta}_{GAN}$ by fitting a feedforward network to predict $R^e_{t+1}F_{t+1}$ and hence estimate $\mathbb{E}_t[R^e_{t+1}F_{t+1}]$. Note, that this loading estimate $\hat{\beta}_{GAN}$ is only proportional to the population value β but this is sufficient for projecting on the systematic and non-systematic component. The conventional forecasting approach yields the conditional mean $\hat{\mu}_{FFN}$, which is proportional to β and hence is used as $\hat{\beta}_{FNN}$ in the projection. At the same time $\hat{\mu}_{FFN}$ is proportional to the SDF factor portfolio weights and hence also serves as $\hat{\omega}_{FFN}$. Note that the linear model is a special case with an explicit solution

$$\begin{aligned}\hat{\theta}_{LS} &= \left(\frac{1}{T}\sum_{t=1}^T\left(\frac{1}{N}\sum_{i=1}^N R^e_{t+1,i}I_{t,i}\right)\left(\frac{1}{N}\sum_{i=1}^N R^e_{t+1,i}I_{t,i}\right)^\top\right)^{-1}\left(\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N R^e_{t+1,i}I_{t,i}\right)\\ &= \left(\frac{1}{T}\sum_{t=1}^T \tilde{F}_{t+1}\tilde{F}_{t+1}^\top\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^T \tilde{F}_{t+1}^\top\right)\end{aligned}$$

and SDF factor portfolio weights $\omega_{LS} = \hat{\theta}_{LS}^\top I_{t,i}$. The risk exposure $\hat{\beta}_{LS}$ is obtained by a linear regression of $R^e_{t+1}F_{t+1}$ on $I_{t,i}$. As the number of characteristics is very large in our setup, the linear model is likely to suffer from over-fitting. The non-linear models include a form of regularization to deal with the large number of characteristics. In order to make the model comparison valid, we add a regularization to the linear model as well. The regularized linear model EN adds an elastic net penalty to the regression to obtain $\hat{\theta}_{EN}$ and in the predictive regression for $\hat{\beta}_{EN}$:[29]

$$\hat{\theta}_{EN} = \arg\min_\theta \left(\frac{1}{T}\sum_{t=1}^T \tilde{F}_{t+1} - \frac{1}{T}\sum_{t=1}^T \tilde{F}_{t+1}\tilde{F}_{t+1}^\top\theta\right)^2 + \lambda_2\|\theta\|_2^2 + \lambda_1\|\theta\|_1.$$

The linear approach with elastic net is closely related to Kozak et al. (2018) who perform mean-variance optimization with an elastic net penalty on characteristic based factors.[30] In addition we

[29]We also use a lasso and ridge regularization, but the elastic net outperforms these approaches. The results are available upon request.

[30]There are five differences to their paper. First, they also include product terms of the characteristics. Second,

also report the maximum Sharpe Ratios for the tangency portfolios based on the Fama-French 3 and 5 factor models.[31]

For the four models GAN, FFN, EN and LS we obtain estimates of ω, which we use to construct the SDF factor, and estimates of β, which we need for the calculation of the residuals ϵ. We obtain the systematic and non-systematic return components by projecting returns on the estimated risk exposure $\hat{\beta}$:

$$\hat{\epsilon}_{t+1} = \left(I_N - \hat{\beta}_t (\hat{\beta}_t^\top \hat{\beta}_t)^{-1} \hat{\beta}_t^\top \right) R^e_{t+1}.$$

We calculate the following three performance metrics: (1) the unconditional Sharpe Ratio of the SDF factor

$$SR = \frac{\hat{\mathbb{E}}[F_t]}{\sqrt{\hat{Var}(F_t)}},$$

(2) the explained variation in individual stock returns

$$EV = 1 - \frac{\left(\frac{1}{T} \sum_{t=1}^{T} \frac{1}{N_t} \sum_{i=1}^{N_t} (\hat{\epsilon}_{t+1,i})^2 \right)}{\left(\frac{1}{T} \sum_{t=1}^{T} \frac{1}{N_t} \sum_{i=1}^{N_t} (R^e_{t+1,i})^2 \right)}$$

and (3) the cross-sectional mean[32] R^2

$$\text{XS-}R^2 \ = 1 - \frac{\frac{1}{N} \sum_{i=1}^{N} \frac{T_i}{T} \left(\frac{1}{T_i} \sum_{t \in T_i} \hat{\epsilon}_{t+1,i} \right)^2}{\frac{1}{N} \sum_{i=1}^{N} \frac{T_i}{T} \left(\frac{1}{T_i} \sum_{t \in T_i} \hat{R}_{t+1,i} \right)^2}.$$

These are generalization of the standard metrics used in linear asset pricing.

IV. Simulation Example

We illustrate with simulations that (1) the no-arbitrage condition in GAN is necessary to find the SDF in a low signal-to-noise setup, (2) the flexible form of GAN is necessary to correctly capture the interactions between characteristics, and (3) the RNN with LSTM is necessary to correctly incorporate macroeconomic dynamics in the pricing kernel. On purpose, we have designed the

their second moment matrix uses demeaned returns, i.e. the two approaches choose different mean-variance efficient portfolios on the globally efficient frontier. Third, they first apply PCA to the characteristics managed factors before solving the mean-variance optimization with elastic net penalty. Fourth, they constraint the tuning parameters based on economic priors. Fifth, we allow for different linear weights on the long and the short leg of the characteristic based factors. Blanchet et al. (2016) show that mean-variance optimization with regularization can be interpreted as an adversarial approach that perturbs the empirical distribution of returns.

[31] The tangency portfolio weights are obtained on the training data set and used on the validation and test data set.

[32] We weight the estimated means by their rate of convergence to account for the differences in precision.

simplest possible simulation setup to convey these points and to show that the forecasting approach or the simple linear model formulations cannot achieve these goals.[33]

Excess returns follow a no-arbitrage model with SDF factor F:

$$R^e_{t+1,i} = \beta_{t,i} F_{t+1} + \epsilon_{t+1,i}.$$

In our simple model the SDF factor follows $F_t \overset{i.i.d.}{\sim} \mathcal{N}(\mu_F, \sigma_F^2)$ and the idiosyncratic component $\epsilon_{t,i} \overset{i.i.d.}{\sim} N(0, \sigma_e^2)$. We consider two different formulations for the risk-loadings:

1. *Two characteristics:* The loadings are the multiplicative interaction of two characteristics

$$\beta_{t,i} = C^{(1)}_{t,i} \cdot C^{(2)}_{t,i} \qquad \text{with } C^{(1)}_{t,i}, C^{(2)}_{t,i} \overset{i.i.d.}{\sim} \mathcal{N}(0,1).$$

2. *One characteristic and one macroeconomic state process:* The loading depends on one characteristic and a state process h_t:

$$\beta_{t,i} = C_{t,i} \cdot b(h_t), \qquad h_t = sin(\pi * t/24) + \epsilon^h_t, \qquad b(h) = \begin{cases} 1 & \text{if } h > 0 \\ -1 & \text{otherwise.} \end{cases}$$

   We observe only the macroeconomic time series $Z_t = \mu_M t + h_t$, where we take $\mu_M = 0.05$. All innovations are independent and normally distributed: $C_{t,i} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ and $\epsilon^h_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 0.25)$.

The choice of the parameters is guided by our empirical results. The panel data set is $N = 500, T = 600$, where the first $T_{train} = 250$ are used for training, the next $T_{valid} = 100$ observations are the validation and the last $T_{test} = 250$ observations form the test data set. The SDF factor has $\sigma_F^2 = 0.1$ and $SR_F = 1$. The idiosyncratic noise variance is $\sigma_e^2 = 1$.

The first model setup with two characteristics has two distinguishing empirical features: (1) the loadings have a non-linear interaction effect for the two characteristics; (2) for many assets the signal-to-noise ratio is low. Because of the multiplicative form the loadings will take small values when two characteristics with values close to zero are multiplied. Figure 4 shows the form of the population loadings. The assets with loadings in the center are largely driven by idiosyncratic noise which makes it harder to extract their systematic component.

Table II reports the results for the first model. The GAN model outperforms the forecasting approach and the linear model in all categories. Note, that it is not necessary to include the elastic net approach as the number of covariates is only two and hence the regularization does not help. The Sharpe Ratio of the estimated GAN SDF factor reaches the same value as the population SDF factor used to generate the data. Based on the estimated loadings respectively the population loadings we project out the idiosyncratic component to obtain the explained variation and cross-sectional pricing errors. As expected the linear model is mis-specified for this setup and captures

[33] We have run substantially more simulations for a variety of different model formulations, where we reach the same conclusions. The other simulation results are available upon request.

neither the SDF factor nor the correct loading structure. Note, that the simple forecasting approach can generate a high Sharpe Ratio but fails in explaining the systematic component.

Table II Performance of Different SDF Models in Two Simulation Setups

	Sharpe Ratio			EV			Cross-sectional R^2		
Model	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
Two characteristics and no macroeconomic state variable									
Population	0.96	1.09	0.94	0.16	0.15	0.17	0.17	0.15	0.17
GAN	0.98	1.11	0.94	0.12	0.11	0.13	0.10	0.09	0.07
FFN	0.94	1.04	0.89	0.05	0.04	0.05	-0.30	-0.09	-0.33
LS	0.07	-0.10	0.01	0.00	0.00	0.00	0.00	0.01	0.01
One characteristic and one macroeconomic state variable									
Population	0.89	0.92	0.86	0.18	0.18	0.17	0.19	0.20	0.15
GAN	0.79	0.77	0.64	0.18	0.18	0.17	0.19	0.20	0.15
FFN	0.05	-0.05	0.06	0.02	0.01	0.02	0.01	0.01	0.02
LS	0.12	-0.05	0.10	0.16	0.16	0.15	0.15	0.18	0.14

Sharpe Ratio (SR) of the SDF factor, explained time series variation (EV) and cross-sectional mean R^2 for the GAN, FFN and LS model. EN is left out in this setup as there are only very few covariates. The data is generated with an SDF factor with Sharpe Ratio $SR = 1$ and $\sigma_F^2 = 0.1$ and the idiosyncratic noise has $\sigma_e^2 = 1$. $N = 500, T = 600$, $T_{train} = 250$, $T_{valid} = 100$ and $T_{test} = 250$.

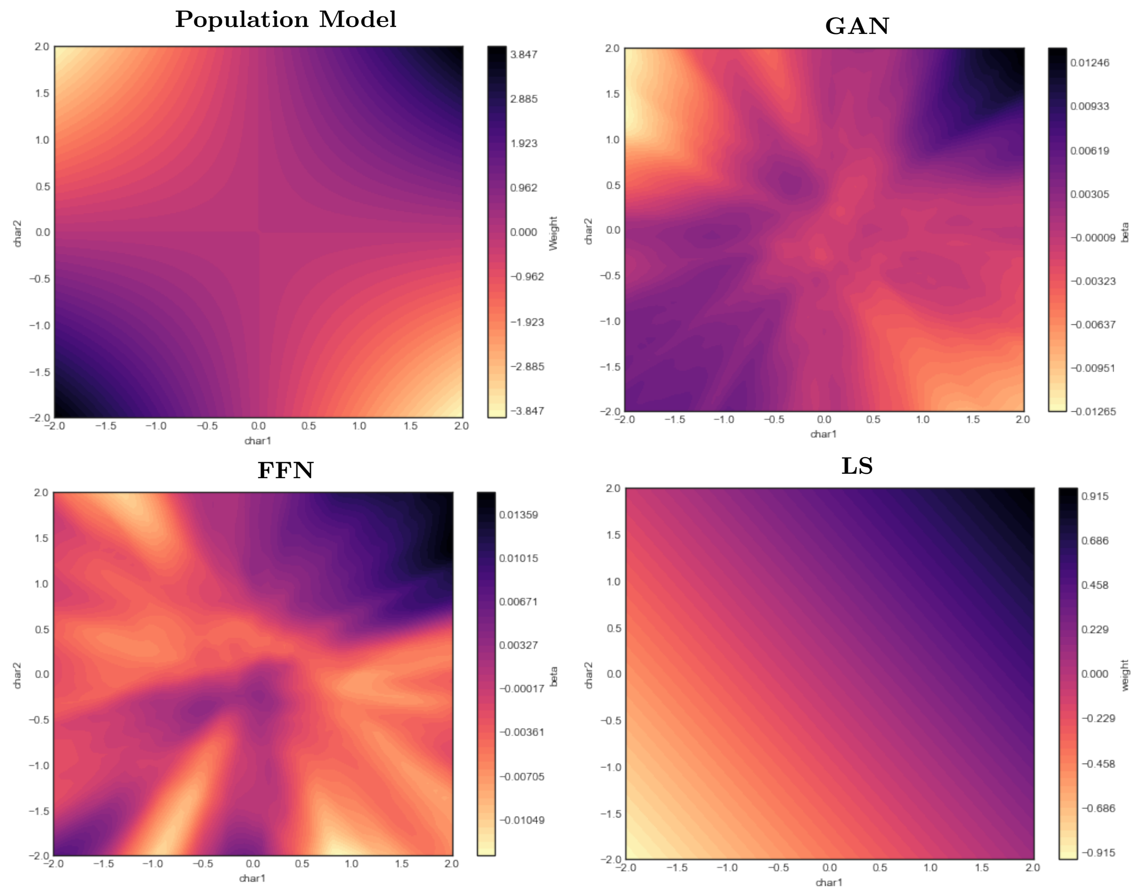
Figure 4 explains why we observe the above performance results. Note, that the SDF factor has large positive respectively negative weights on the extreme corner combinations of the characteristics. The middle combinations are close to zero. The GAN network captures this pattern and assigns positive weights on the combinations of high/high and low/low and negative weights for high/low and low/high. The FFN on the other hand generates a more diffuse picture. It assigns negative weights for low/low combinations. The FFN SDF factor still loads mainly on the extreme portfolios which results in the high Sharpe Ratio. However, the FFN fails to capture the loadings correctly which leads to high unexplained variation and pricing errors. The linear model can obviously not capture the non-linear interaction.

The second model setup with a macroeconomic state variable is designed to model the effect of a boom and recession cycle on the pricing model. In our model the SDF factor affects the assets differently during a boom and recession cycle. Note, that in general a macroeconomic variable can by construction only have a scaling effect on the loadings of the SDF factor but not change its cross-sectional distribution which can only depend on firm-specific information.

Figure 5 illustrates the path of the observed macroeconomic variable that has the distinguishing feature that we observe for most macroeconomic variables in our data set: (1) the macroeconomic process is non-stationary, i.e. it has a trend; (2) the process has a cyclical dynamic structure, i.e.

it is influenced by business cycles. For example GDP level has a similar qualitative behaviour. The conventional approach to deal with non-stationary data is to take first differences. Figure 5 shows that the differenced data does indeed look stationary but loses all information about the business cycle. The LSTM network in our GAN model can successfully extract the hidden state process. The models based on first differences can by construction not infer any dynamics in the macroeconomic variables.

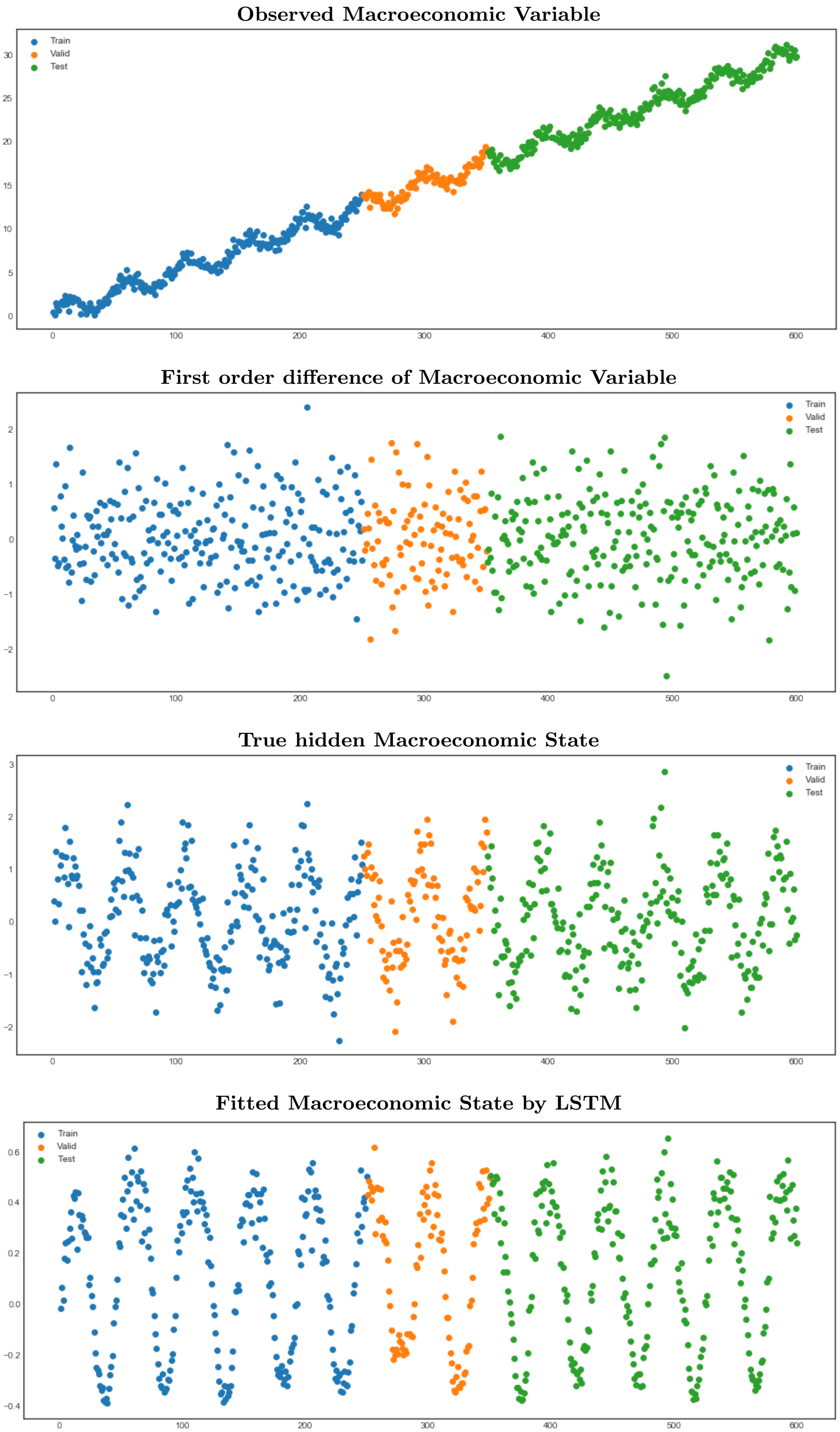
Figure 4. Loadings β for First Model with 2 Characteristics



Loadings β as the function of the two characteristics estimated by different methods.

Table II reports the results for the second model with macroeconomic state variable. As expected our GAN model strongly outperforms the forecasting and the linear model. Note, that the loading function here is linear and the macroeconomic state variable is only a time-varying proportionality constant for the loadings and SDF weights. As the projection on the systematic component is not affected by a proportionality constant, the linear model actually achieves the same explained variation and pricing errors as GAN. However, the Sharpe Ratio of the linear model collapses as for roughly half of the times it uses the wrong sign for the SDF weights.

Figure 5. Dynamics of Macroeconomic State Variable



The simulation section illustrates three findings: (1) All three evaluation metrics (SR, EV and XS-R^2) are necessary to assess the quality of the SDF factor. A model like FFN can achieve high Sharpe Ratios by loading on some extreme portfolios but it does not imply that it captures the loading structure correctly.[34] On the other hand the explained variation of a model can be high as for LS, but it does not capture the correct sign of the SDF weights and loadings that can depend on macroeconomic conditions. (2) It does not matter how flexible the model is (e.g. FFN), by conditioning only on the most recent macroeconomic observations, general macroeconomic dynamics are ruled out. (3) The no-arbitrage condition in the GAN model helps to deal with a low signal-to-noise ratio.

V. Empirical Results for U.S. Equities

A. *Data*

A.1. Returns and Firm Specific Characteristic Variables

We collect monthly equity return data for all securities on CRSP. The sample period spans January 1967 to December 2016, totaling 50 years. We divide the full data into 20 years of training sample (1967 - 1986), 5 years of validation sample (1987 - 1991), and 25 years of out-of-sample testing sample (1992 - 2016). We use the one-month Treasury bill rates from the Kenneth French Data Library as the risk-free rate to calculate excess returns.

In addition, we collect the 46 firm-specific characteristics listed either on Kenneth French Data Library or used by Freyberger et al. (2017).[35] All these variables are constructed either from accounting variables from the CRSP/Compustat database or from past returns from CRSP. We follow the standard conventions in the variable definition, construction and their updating. Yearly updated variables are updated at the end of each June following the Fama-French convention, while monthly changing variables are updated at the end of each month for the use in the next month. The full details on the construction of these variables are in Table A.VII. In Table A.VIII we sort the characteristics into the six categories *past returns, investment, profitability, intangibles, value and trading frictions*.

The number of all available stocks from CRSP is around 31,000. As in Kelly et al. (2018) or Freyberger et al. (2017), we are limited to the returns of stocks that have all firm characteristics information available in a certain month, which leaves us with around 10,000 stocks. This is the largest possible data set that can be used for this type of analysis.[36] Figure 11 plots the number of stocks available in each month.

[34] Pelger and Xiong (2018a) provide the theoretical arguments and show empirically in a linear setup why "proximate" factors that only capture the extreme factor weights correctly have similar time series properties as the population factors but might not have the correct loadings.

[35] We use the characteristics that Freyberger et al. (2017) used in the 2017 version of their paper.

[36] Using stocks with missing characteristic information requires data imputation based on model assumptions. Gu et al. (2019) replace a missing characteristic with the cross-sectional median of that characteristic during that month. However, this approach introduces an additional source of error and as it ignores the dependency structure in the characteristic space creates artificial time-series fluctuation in the characteristics, which we want to avoid.

For each characteristic variable in each month, we rank them cross-sectionally and convert them into quantiles. This is a standard transformation to deal with the different scales and has also been used in Kelly et al. (2018), Kozak et al. (2018) or Freyberger et al. (2017) among others. In the linear model the projection $\tilde{F}_{t+1} = \frac{1}{N}\sum_{i=1}^{N} I_{t,i} R^e_{t+1,i}$ results in long-short factors with an increasing positive weight for stocks that have a characteristic value above the median and a decreasing negative weight for below median values.[37] We increase the flexibility of the linear model by including the positive and negative leg separately for each characteristic, i.e. we take the rank-weighted average of the stocks with above median characteristic values and similarly for the below median values. This results in two "factors" for each characteristic. Note, that our model includes the conventional long-short factors as a special case where the long and short legs receive the same weight in the SDF. These factors are still zero cost portfolios as they are based on excess returns.[38]

A.2. Macroeconomic Variables

We collect 178 macroeconomic time series from three sources. We take 124 macroeconomic predictors from the FRED-MD database as detailed in McCracken and Ng (2016). Next, we add the cross-sectional median time series for each of the 46 firm characteristics. The quantile distribution combined with the median level for each characteristics is close to representing the same information as the raw characteristic information but in a normalized form. Third, we supplement the time series with the 8 macroeconomic predictors from Welch and Goyal (2007) which have been suggested as predictors for the equity premium.

We apply standard transformations to the time series data. We use the transformations suggested in McCracken and Ng (2016), and define transformations for the 46 median and the 8 time series from Welch and Goyal (2007) to obtain stationary time series. The transformations include: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$; and (7) $\Delta(x_t/x_{t-1} - 1.0)$. A detailed description of the macroeconomic variables as well as their corresponding transformations (tCode) are collected in Appendix G.A.

B. Cross Section of Individual Stock Returns

The GAN SDF factor has a higher out-of-sample Sharpe Ratio while explaining more variation and pricing than the other benchmark models. Table III reports the three main performance measures, Sharpe Ratio, explained variation and cross-sectional R^2, for the four model specifications. The annual out-of-sample Sharpe Ratio of GAN is around 2.6 and almost twice as high as with the simple forecasting approach FFN. The non-linear and interaction structure that GAN

[37] Kelly et al. (2018) and Kozak et al. (2018) also construct characteristic based factors in this way.

[38] In the first version of this paper we used the conventional long-short factors. However, our empirical results suggest that the long and short leg have different weights in the SDF and this additional flexibility improves the performance of the linear model. These findings are also in line with Lettau and Pelger (2018) who extract linear factors from the extreme deciles of single sorted portfolios and show that they are not spanned by long-short factors that put equal weight on the extreme deciles of each characteristic.

can capture results in a 50% increase compared to the regularized linear model. Hence, the more flexible form matters, but an appropriately designed linear model can already achieve an impressive performance. The in-sample results suffer from overfitting, but the annual in-sample Sharpe Ratio of GAN with 9.3 clearly stands out. The non-regularized linear model has the worst performance in terms of explained variation and pricing error. GAN explains 8% of the variation of individual stock returns which is twice as large as the other models. Similarly, the cross-sectional R^2 of 23% is substantially higher than for the other models. Interestingly, the regularized linear model based on the no-arbitrage objective function explains the time-series and cross-section of stock returns at least as good as the flexible neural network without the no-arbitrage condition. Each model here uses the optimal set of hyperparameters to maximize the validation Sharpe Ratio. In case of the LS, EN and FFN this implies to leave out the macroeconomic variables.[39]

Table III Performance of Different SDF Models

	SR			EV			Cross-Sectional R^2		
Model	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
LS	1.80	0.58	0.42	0.09	0.03	0.03	0.15	0.00	0.14
EN	1.37	1.15	0.50	0.12	0.05	0.04	0.17	0.02	0.19
FFN	0.45	0.42	0.44	0.11	0.04	0.04	0.14	-0.00	0.15
GAN	2.68	1.43	0.75	0.20	0.09	0.08	0.12	0.01	0.23

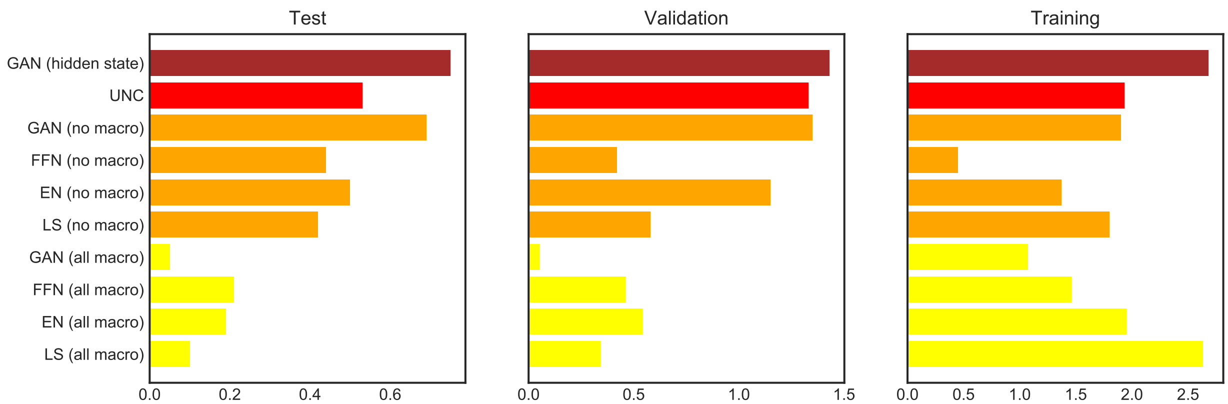
Monthly Sharpe Ratio (SR) of the SDF factor, explained time series variation (EV) and cross-sectional mean R^2 for the GAN, FFN, EN and LS model.

Figure 6 summarizes the effect of conditioning on the hidden macroeconomic state variables. First, we add the 178 macroeconomic variables as predictors to all networks without reducing them to the hidden state variables. The performance for the out-of-sample Sharpe Ratio of the LS, EN, FFN and GAN model completely collapses. First, conditioning only on the last normalized observation of the macroeconomic variables, which is usually an increment, does not allow to detect a dynamic structure, e.g. a business cycle. The decay in the Sharpe Ratio indicates that using only the past macroeconomic information results in a loss of valuable information. Even worse, including the large number of irrelevant variables actually lowers the performance compared to a model without macroeconomic information. Although the models use a form of regularization, a too large number of irrelevant variables makes it harder to select those that are actually relevant. The results for the in-sample training data illustrate the complete overfitting when the large number of macroeconomic variables is included. FFN, EN and LS without macroeconomic information perform better and that is why we choose them as the comparison benchmark models. GAN without the macroeconomic but only firm-specific variables has an out-of-sample Sharpe Ratio that

[39]The results are not affected by normalizing the SDF weights to have $\|\omega\|_1 = 1$. The explained variation and pricing results are based on a cross-sectional projection at each time step which is independent of any scaling. The internet appendix collects the additional results.

is around 10% lower than with the macroeconomic hidden states. This is another indication that it is relevant to include the dynamics of the time series. The UNC model uses only unconditional moments as the objective function, but includes the LSTM hidden states in the factor weights. The Sharpe Ratio is around a 20% lower than the GAN with hidden states. Hence, it is not only important to include the hidden states in the weights and loadings but also in the objective function to identify the times when they matter for pricing.

Figure 6. Performance of Models with Different Macroeconomic Variables



Sharpe Ratio of SDF factor for different inclusions of the macroeconomic information. The GAN (hidden states) is our reference model. UNC is a special version of our model that uses only unconditional moments (but includes LSTM macroeconomic states in the FFN network for the SDF weights). GAN (no macro), FFN (no macro), EN (no macro) and LS (no macro) use only firm specific information as conditioning variables but no macroeconomic variables. GAN (all macro), FFN (all macro), EN (all macro) and LS (all macro) include all 178 macro variables as predictors (respectively conditioning variables) without using LSTM to transform them into macroeconomic states.

Table IV SDF Factor Risk Measures

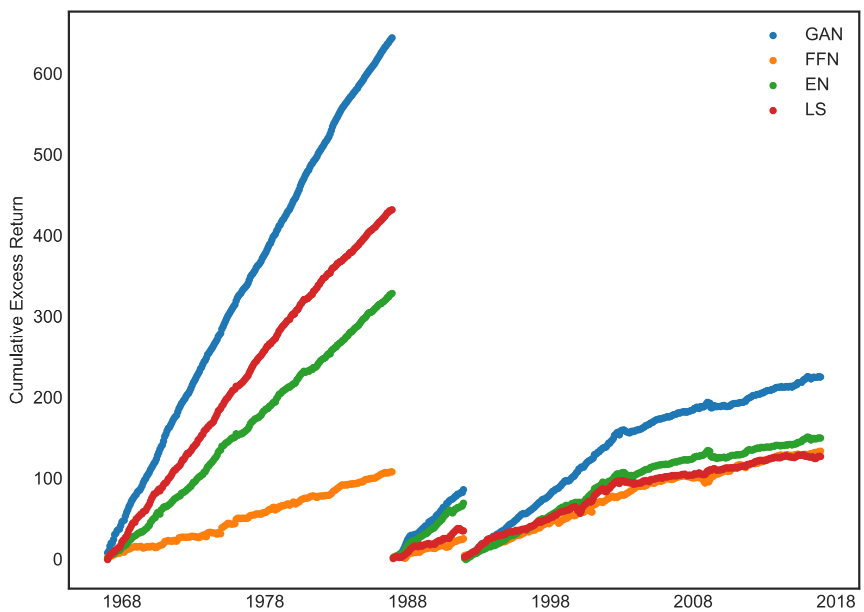
	SR			Max Loss			Max Drawdown		
Model	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
FF-3	0.27	-0.09	0.19	-2.45	-2.85	-4.31	7	10	10
FF-5	0.48	0.40	0.22	-2.62	-2.33	-4.90	4	3	7
LS	1.80	0.58	0.42	-1.96	-1.87	-4.99	1	3	4
EN	1.37	1.15	0.50	-2.22	-1.81	-6.18	1	3	5
FFN	0.45	0.42	0.44	-3.30	-4.61	-3.37	6	3	5
GAN	2.68	1.43	0.75	0.38	-0.28	-5.76	0	1	5

Sharpe Ratio, maximum 1-month loss and maximum drawdown of the SDF factor portfolios. We include the mean-variance efficient portfolio based on the 5 Fama-French factors.

The SDF factor is a tradeable portfolio with an attractive risk-return trade-off. Table IV reports the monthly Sharpe Ratios, maximum 1-month loss and maximum drawdown of the four benchmark models and also the Fama-French 3 and 5 factor models.[40] The number of consecutive losses as measured by drawdown and the maximum loss for the GAN model is comparable to the other models, while the Sharpe Ratio is by far the highest. Figure 7 plots the cumulative return for each model normalized by the standard deviation. As suggested by the risk-measures the GAN return exceeds the other models while it avoids fluctuations and large losses. Table A.II lists the turnover for the different approaches. The GAN factor has a comparative or even lower turnover than the other SDF factors. This suggests that all approaches are exposed to similar transaction costs and it is valid to directly compare their risk-adjusted return.

Gu et al. (2018) report high out-of-sample Sharpe Ratios for long-short portfolios based on the extreme quantiles of returns predicted by FFN. Table A.I compares the Sharpe Ratios for different extreme quantiles for equally and value-weighted long-short portfolios with FFN. We can replicate the high out-of-sample Sharpe Ratios when using extreme deciles of 10% or less and equally-weighted portfolios. However, for value-weighted portfolios the Sharpe Ratio drops by around 50%. This is a clear indication that the performance of these portfolios heavily depends on small stocks. For this reason we use the implied SDF factor as the benchmark portfolio for the FFN approach.

Figure 7. Cumulative Excess Returns for SDF Factor



Cumulative excess returns for the SDF factors for GAN, FFN, EN and LS. Each factor is normalized by its standard deviation for the time interval under consideration.

[40]Max Drawdown is defined as the maximum number of consecutive months with negative returns. The maximum 1-month loss is normalized by the standard deviation of the asset.

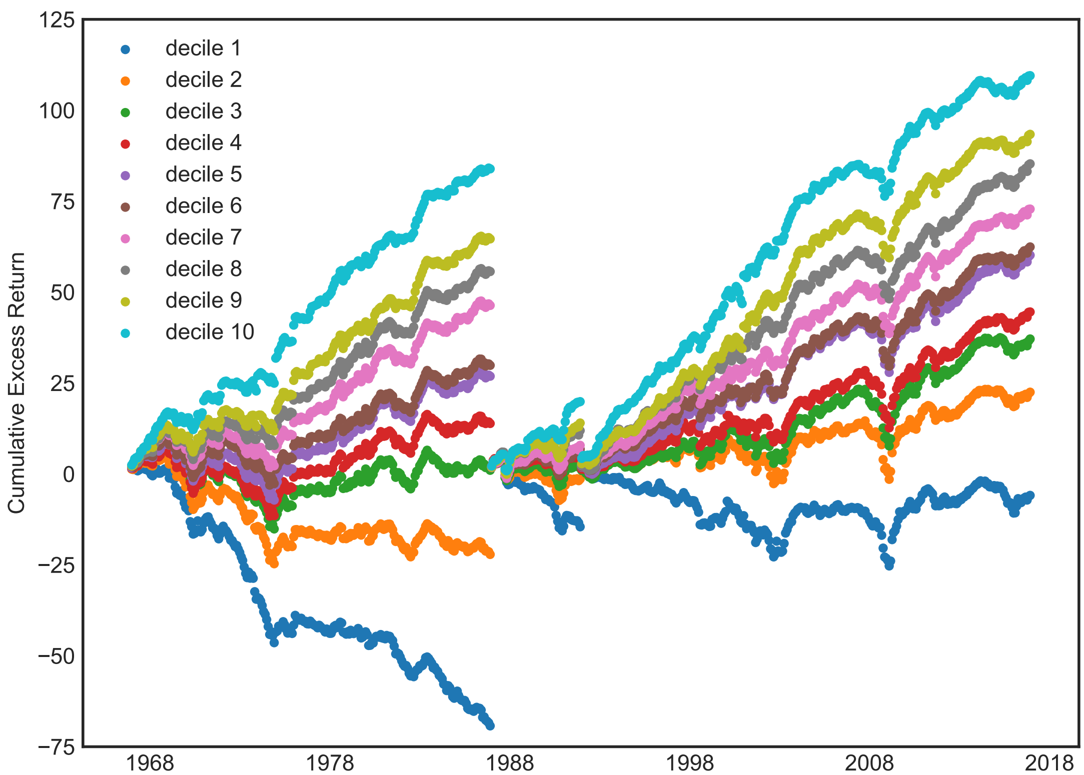
C. Predictive Performance

The no-arbitrage factor representation implies a connection between average returns of stocks and their risk exposure to the SDF factor measured by β. The fundamental equation

$$\mathbb{E}_t[R^e_{t+1,i}] = \beta_{t,i}\mathbb{E}_t[F_{t+1}]$$

implies that as long as the conditional risk premium $\mathbb{E}_t[F_{t+1}]$ is positive, assets with a higher risk exposure $\beta_{t,i}$ should have higher expected returns.[41] We test the predictive power of our model by sorting stocks into decile portfolios based on their risk loadings.

Figure 8. Cumulative Excess Return of Decile Sorted Portfolios by GAN



Cumulative excess return of decile sorted portfolios based on the risk loadings β. The first portfolio is based on the smallest decile of risk loadings, while the last decile portfolio is constructed with the largest loading decile. Within each decile the stocks are equally weighted.

In Figure 8 we plot the cumulative excess return of decile sorted portfolios based on risk loadings β's. Portfolios based on higher β's have higher subsequent returns. This clearly indicates that the risk loading predicts future stock returns. In particular the highest and lowest deciles clearly separate. The Internet Appendix collects the corresponding results for the other estimation approaches with qualitatively similar findings, i.e. the risk loadings predict future returns.

The systematic return difference is not explained by the market or Fama-French factors. Table

[41] We consider it a sensible and weak assumption that risk is compensated in the market and hence the conditional risk premium $\mathbb{E}_t[F_{t+1}]$ is positive.

V reports the time series pricing errors with corresponding t-statistics for the 10 decile-sorted portfolios for the three factor models. Obviously, the pricing errors are highly significant and expected returns of almost all decile portfolios are not explained by the Fama-French factors. The GRS test clearly rejects the null-hypothesis that either of the factor models prices this cross-section. These β-sorted portfolios equally weight the stocks within each decile. Figure A.6 and A.III in the Appendix show that the findings extend to value weighted β sorted portfolios.

Table V Time Series Pricing Errors for β-Sorted Portfolios

	Average Returns		Market-Rf				Fama-French 3				Fama-French 5			
	Whole	Test	Whole		Test		Whole		Test		Whole		Test	
Decile			α	t	α	t	α	t	α	t	α	t	α	t
1	-0.12	-0.02	-0.19	-8.92	-0.11	-3.43	-0.21	-12.77	-0.13	-5.01	-0.20	-11.99	-0.12	-4.35
2	-0.00	0.05	-0.07	-4.99	-0.04	-1.56	-0.09	-8.79	-0.05	-3.22	-0.09	-8.29	-0.05	-2.68
3	0.04	0.08	-0.02	-2.01	-0.00	-0.16	-0.04	-5.18	-0.02	-1.40	-0.04	-4.87	-0.01	-1.05
4	0.07	0.09	-0.00	-0.03	0.01	0.68	-0.02	-2.30	-0.00	-0.35	-0.02	-2.86	-0.01	-0.54
5	0.10	0.12	0.03	2.75	0.04	2.50	0.01	2.08	0.03	2.46	0.01	1.36	0.03	2.17
6	0.11	0.12	0.04	3.16	0.05	2.77	0.02	2.75	0.03	2.85	0.01	1.51	0.02	2.20
7	0.14	0.15	0.07	5.62	0.07	3.92	0.05	6.61	0.05	4.39	0.04	5.16	0.04	3.41
8	0.18	0.18	0.11	7.41	0.10	5.12	0.08	9.32	0.08	5.83	0.07	8.05	0.07	4.86
9	0.22	0.21	0.15	7.83	0.13	5.37	0.11	9.16	0.11	5.71	0.11	8.58	0.11	5.39
10	0.37	0.37	0.29	9.22	0.27	6.05	0.24	10.03	0.25	6.27	0.25	10.43	0.27	6.59
10-1	0.48	0.39	0.47	18.93	0.38	10.29	0.45	18.50	0.38	10.14	0.46	18.13	0.39	9.96
GRS Asset Pricing Test			GRS	p	GRS	p	GRS	p	GRS	p	GRS	p	GRS	p
			42.23	0.00	11.58	0.00	39.72	0.00	11.25	0.00	37.64	0.00	10.75	0.00

Average returns, time series pricing errors and corresponding t-statistics for β-sorted decile portfolios based on GAN. The pricing errors are based on the CAPM and Fama-French 3 and 5 factors models. Returns are annualized. The GRS-test is under the null hypothesis of correctly pricing all decile portfolios and includes the p-values. We consider the whole time period and the test period. Within each decile the stocks are equally weighted.

D. *Pricing of Characteristic Sorted Portfolios*

Our approach achieves unprecedented pricing performance on standard test portfolios. Asset pricing testing is usually conducted on characteristic sorted portfolios that isolate the pricing effect of a small number of characteristics. Our models provide risk loadings β_t's for each individual stock. We sort the stocks into value weighted decile and double-sorted 25 portfolios based on the characteristics.[42] The risk loadings β_t's for the portfolios are obtained by aggregating the corresponding stock specific loadings. We obtain the systematic and error components with a cross-sectional regression on β_t at each point in time. This is similar to a standard cross-sectional Fama-MacBeth regression in a linear model with the main difference that the β_t's are obtained from our SDF models on individual stocks. We normalize the variation and mean of the systematic component by the corresponding variation and mean of all portfolios. This is the same procedure

[42] Here we report only the results for value weighted portfolios. The results for equally weighted portfolios are similar.

as for EV and XS-R^2 but on the portfolio instead of the stock level. For each individual decile or double-sorted portfolio we also normalize its systematic variation by its overall variation. For the individual quantiles we also report the pricing error $\hat{\alpha}_i$ normalized by the root-mean-squared average returns of all corresponding quantile sorted portfolios, i.e. $\hat{\alpha}_i = \frac{\hat{\mathbb{E}}[\hat{\epsilon}_{t,i}]}{\sqrt{\frac{1}{N}\sum_{i=1}^N \hat{\mathbb{E}}[R_{t,i}]^2}}$.[43]

Table VI Explained Variation and Pricing Errors for Decile Sorted Portfolios

	EN	FFN	GAN	EN	FFN	GAN	EN	FFN	GAN	EN	FFN	GAN
	Short-Term Reversal						Momentum					
Decile	Explained Variation			Alpha			Explained Variation			Alpha		
1	0.84	0.74	0.77	-0.18	-0.21	-0.13	0.04	-0.06	0.33	0.37	0.39	0.11
2	0.86	0.81	0.82	0.00	-0.05	0.00	0.12	0.10	0.52	0.25	0.18	-0.01
3	0.80	0.82	0.84	0.13	0.04	0.06	0.19	0.25	0.66	0.14	0.05	-0.06
4	0.69	0.80	0.82	0.16	0.03	0.03	0.28	0.34	0.73	0.15	0.08	-0.02
5	0.58	0.68	0.71	0.13	-0.03	-0.04	0.37	0.46	0.80	0.19	0.09	0.02
6	0.43	0.66	0.75	0.22	0.05	0.01	0.45	0.58	0.78	0.02	-0.03	-0.09
7	0.23	0.64	0.77	0.20	0.03	-0.02	0.62	0.69	0.68	0.01	0.01	-0.05
8	-0.07	0.49	0.67	0.23	0.03	-0.05	0.58	0.71	0.64	-0.03	-0.04	-0.09
9	-0.25	0.29	0.58	0.30	0.09	-0.01	0.55	0.70	0.58	0.08	0.04	-0.03
10	-0.24	-0.04	0.35	0.47	0.38	0.18	0.51	0.53	0.53	0.24	0.29	0.19
	Explained Variation			Cross-Sectional R^2			Explained Variation			Cross-Sectional R^2		
All	0.43	0.58	0.70	0.45	0.79	0.94	0.26	0.27	0.54	0.66	0.71	0.93
	Book-To-Market						Size					
Decile	Explained Variation			Alpha			Explained Variation			Alpha		
1	0.38	0.66	0.70	0.03	-0.12	-0.08	0.80	0.75	0.79	0.09	-0.00	0.10
2	0.48	0.73	0.78	0.10	-0.05	-0.04	0.89	0.89	0.90	-0.11	-0.09	-0.06
3	0.71	0.84	0.86	0.07	-0.03	-0.01	0.91	0.80	0.91	-0.07	0.02	-0.02
4	0.76	0.88	0.89	0.00	-0.07	-0.07	0.90	0.77	0.91	-0.05	0.04	-0.01
5	0.82	0.87	0.88	0.05	0.02	0.01	0.90	0.78	0.91	0.01	0.10	0.04
6	0.77	0.82	0.88	0.06	0.04	0.02	0.88	0.80	0.91	0.03	0.09	0.02
7	0.81	0.81	0.87	0.03	0.08	0.03	0.84	0.81	0.89	0.04	0.05	-0.01
8	0.71	0.59	0.78	0.03	0.12	0.06	0.84	0.85	0.88	0.06	0.03	-0.02
9	0.80	0.72	0.80	-0.02	0.11	0.07	0.77	0.81	0.82	0.06	-0.01	-0.04
10	0.68	0.73	0.79	-0.05	-0.00	0.00	0.32	0.28	0.49	-0.04	-0.15	-0.10
	Explained Variation			Cross-Sectional R^2			Explained Variation			Cross-Sectional R^2		
All	0.70	0.75	0.82	0.97	0.94	0.98	0.83	0.78	0.86	0.96	0.95	0.97

Out-of-sample explained variation and pricing errors for decile-sorted portfolios based on Short-Term Reversal (ST_REV), Momentum (r12_2), Book to Market Ratio (BEME) and Size (LME).

[43]Note, that $XS\text{-}R^2 = 1 - \sum_{i=1}^N \hat{\alpha}_i^2$. The results for the unregularized linear model are the worst and available upon request.

Table VI starts with four sets of decile sorted portfolios. We choose short-term reversal and momentum as these are the two most important variables as discussed in the next sections and size and book-to-market sorted portfolios which are well-studied characteristics. GAN can substantially better capture the variation and mean return for short-term reversal and momentum sorted decile portfolios. EN and FFN have a very similar performance. The better GAN results are driven by explaining the extreme decile portfolios (the 10th decile for short-term reversal and the first decile for momentum). All approaches perform very similar for the middle portfolios. It turns out that book-to-market and size sorted portfolios are very "easy" to price. All models have time series R^2 above 70% and cross-sectional R^2 close to 1. Hence, all models seems to capture this pricing information almost perfectly, although the GAN results are still slightly better than for the other models.

Table VII Explained Variation and Pricing Errors for Decile Sorted Portfolios

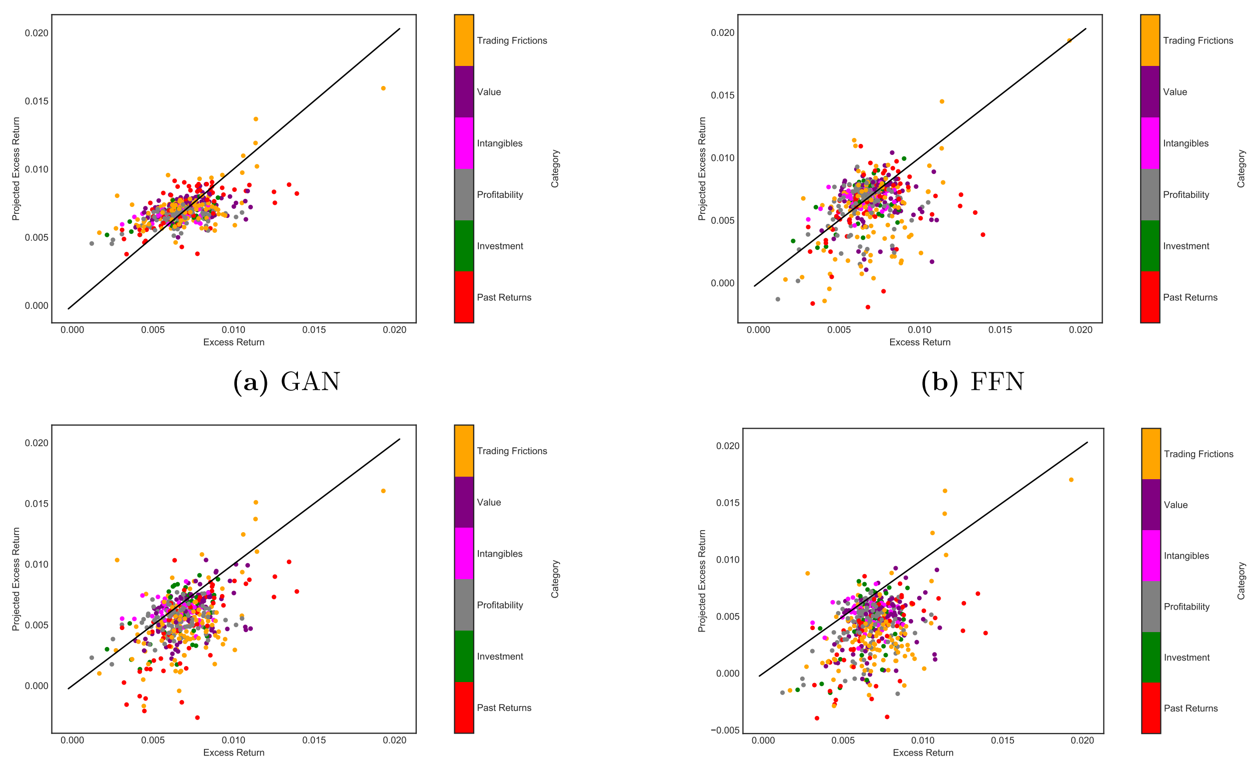
	Explained Variation			Cross-Sectional R^2				Explained Variation			Cross-Sectional R^2		
Charact.	EN	FFN	GAN	EN	FFN	GAN	Charact.	EN	FFN	GAN	EN	FFN	GAN
ST_REV	0.43	0.58	0.70	0.45	0.79	0.94	Q	0.68	0.70	0.78	0.97	0.92	0.96
SUV	0.42	0.75	0.83	0.64	0.97	0.99	Investment	0.54	0.65	0.75	0.91	0.94	0.98
r12_2	0.26	0.27	0.54	0.66	0.71	0.93	PM	0.52	0.42	0.68	0.90	0.86	0.93
NOA	0.58	0.69	0.78	0.94	0.96	0.95	DPI2A	0.57	0.70	0.78	0.90	0.95	0.97
SGA2S	0.52	0.63	0.73	0.93	0.95	0.96	ROE	0.59	0.56	0.76	0.91	0.86	0.97
LME	0.83	0.78	0.86	0.96	0.95	0.97	S2P	0.69	0.79	0.82	0.98	0.98	0.97
RNA	0.50	0.48	0.69	0.93	0.87	0.96	FC2Y	0.56	0.71	0.76	0.91	0.94	0.95
LTurnover	0.52	0.57	0.68	0.88	0.89	0.96	AC	0.63	0.79	0.82	0.96	0.98	0.98
Lev	0.52	0.63	0.73	0.90	0.92	0.95	CTO	0.59	0.73	0.79	0.92	0.96	0.97
Resid_Var	0.52	0.27	0.65	0.84	0.73	0.97	LT_Rev	0.60	0.59	0.72	0.93	0.85	0.94
ROA	0.51	0.44	0.70	0.92	0.93	0.98	OP	0.56	0.48	0.74	0.97	0.88	0.98
E2P	0.48	0.44	0.67	0.86	0.80	0.95	PROF	0.58	0.62	0.76	0.91	0.98	0.95
D2P	0.47	0.51	0.72	0.82	0.85	0.94	IdioVol	0.43	0.27	0.66	0.79	0.72	0.97
Spread	0.49	0.32	0.60	0.76	0.71	0.92	r12_7	0.37	0.42	0.66	0.84	0.86	0.93
CF2P	0.46	0.47	0.66	0.90	0.89	0.99	Beta	0.45	0.46	0.62	0.83	0.87	0.97
BEME	0.70	0.75	0.82	0.97	0.94	0.98	OA	0.65	0.78	0.83	0.88	0.92	0.93
Variance	0.48	0.27	0.61	0.74	0.72	0.90	ATO	0.58	0.70	0.77	0.96	0.98	0.99
D2A	0.57	0.71	0.78	0.96	0.96	0.97	MktBeta	0.44	0.44	0.64	0.81	0.85	0.97
PCM	0.66	0.79	0.82	0.97	0.98	0.99	OL	0.60	0.73	0.78	0.95	0.97	0.97
A2ME	0.72	0.79	0.83	0.97	0.96	0.98	C	0.51	0.65	0.73	0.90	0.93	0.95
AT	0.77	0.70	0.83	0.77	0.89	0.92	r36_13	0.54	0.53	0.69	0.92	0.82	0.93
Rel2High	0.46	0.33	0.60	0.90	0.83	0.97	NI	0.51	0.60	0.75	0.88	0.96	0.99
CF	0.61	0.64	0.78	0.89	0.85	0.96	r2_1	0.51	0.52	0.69	0.87	0.90	0.95

Out-of-sample explained variation and pricing errors for decile-sorted portfolios.

Tables A.IV and A.V in the appendix repeat the same analysis on short-term reversal and momentum double sorted and size and book-to-market double-sorted portfolios. The takeaways are similar to the decile sorted portfolios. GAN outperforms FFN and EN on the momentum related portfolios, while all three models are able to explain the size and value double sorted portfolios. Importantly, the linear EN becomes worse on the double-sorted reversal and momentum portfolios. This is due to the extreme corner portfolios, which are in parricular low momentum and high short-term reversal stocks. This implies that the linear model cannot capture the interaction between characteristics, while the GAN model successfully identifies the potentially non-linear interaction effects.[44]

Our findings generalize to other decile sorted portfolios. Table VII collects the explained variation and cross-sectional R^2 for all decile-sorted portfolios. It is striking that GAN is always better than the other two models in explaining variation. At the same time GAN achieves a cross-sectional R^2 higher than 90% for all characteristics. In the few cases where the other models have a slightly higher cross-sectional R^2, this number is very close to 1, i.e. all models can essentially perfectly explain the pricing information in the deciles. In summary GAN strongly dominates the other methods in explaining sorted portfolios. The results show (1) that the non-linearities and interactions matter as GAN is better than EN and (2) the no-arbitrage condition extracts additional information as GAN is better than FFN.

Figure 9. Predicted returns for value weighted characteristic sorted portfolios



(a) GAN **(b)** FFN

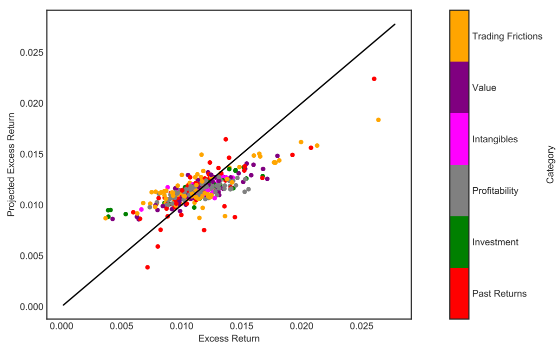
(c) EN **(d)** LS

Predicted and average excess returns for value weighted characteristic sorted decile portfolios.

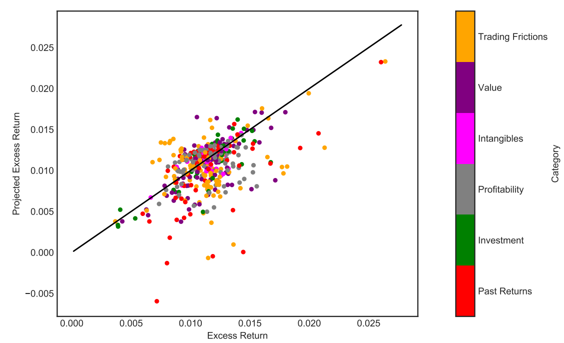
[44]The Internet Appendix collects the results for additional characteristic sorts with similar findings.

Figure 9 visualizes the ability of GAN to explain the cross-section of expected returns for all value weighted characteristic sorted deciles. We plot the average excess return and the model implied average excess return. In an ideal model the points would line up on the 45 degree line. The GAN SDF captures the correct monotonic behavior, but its prediction is biased to towards the mean. In contrast, the prediction of other three models show a larger discrepancy which holds for characteristics of all groups. Figure 10 shows the prediction results for equally weighted decile portfolios. All models seem to perform slightly better, but the general findings are the same.

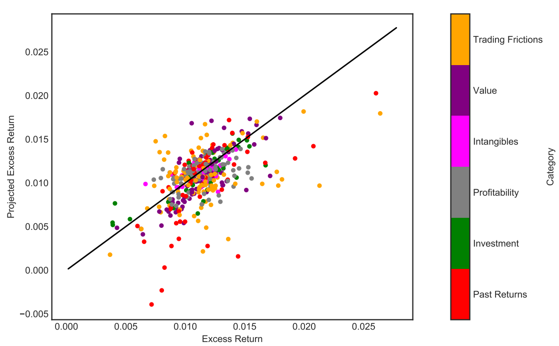
Figure 10. Predicted returns for equally weighted characteristic sorted portfolios



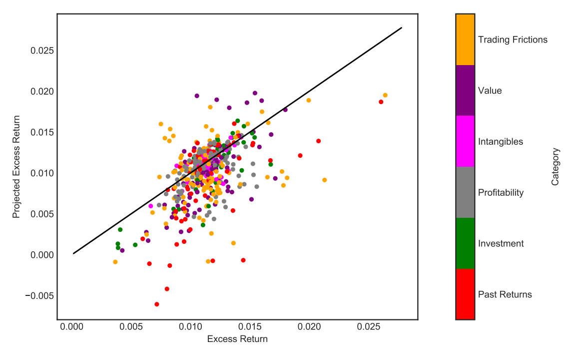
(a) GAN



(b) FFN



(c) EN



(d) LS

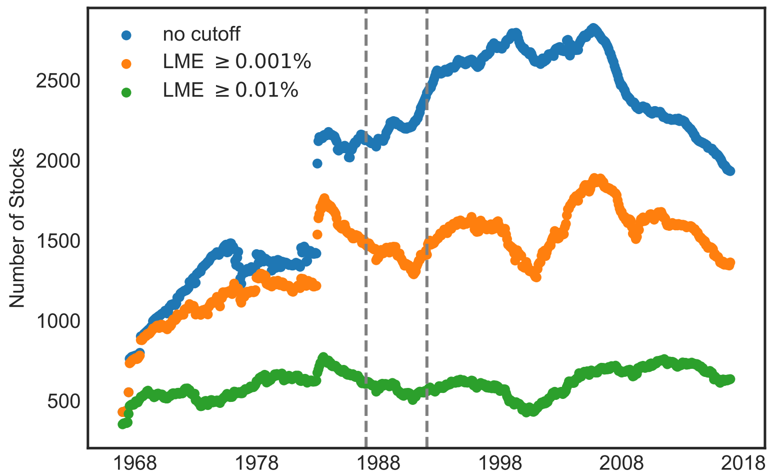
Predicted and average excess returns for equally weighted characteristic sorted decile portfolios.

E. *Robustness to Size*

The qualitative findings are robust to small cap stocks. It is well-known that penny stocks can achieve high Sharpe ratios and are hard to price by conventional asset pricing models. However, trading in these small cap stocks is limited due to low liquidity and high spreads. Hence, the high Sharpe ratios or large alphas of small cap stocks can potentially not be exploited. Here we compare the model performance restricted to medium and large cap stocks.

Our cross-section of stocks in the test data is composed of 2,000 to 3,000 individual stocks per month. Figure 11 shows that the restriction to the stocks with a market capitalization larger than 0.001% of the total market capitalization leaves us on average with the largest 1,500 stocks. Restricting the sample to stocks with market cap above 0.01% of the total market cap yields on average the largest 550 stocks, i.e. the sample is close to the S&P 500 index.

Figure 11. Number of Stocks per Month



Number of stocks per month in the total sample and for stocks with market capitalization larger than 0.01% or 0.001% of the total market capitalization.

Table VIII Different SDF Models Evaluated on Large Market Cap Stocks

	SR			EV			Cross-Sectional R^2		
Model	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
				Size $\geq$ 0.001% of total market cap					
LS	1.44	0.31	0.13	0.07	0.05	0.03	0.14	0.03	0.10
EN	0.93	0.56	0.15	0.11	0.09	0.06	0.17	0.05	0.14
FFN	0.42	0.20	0.30	0.11	0.10	0.05	0.19	0.08	0.18
GAN	2.32	1.09	0.41	0.23	0.22	0.14	0.20	0.13	0.26
				Size $\geq$ 0.01% of total market cap					
LS	0.32	-0.11	-0.06	0.05	0.07	0.04	0.13	0.05	0.09
EN	0.37	0.26	0.23	0.09	0.12	0.07	0.17	0.08	0.14
FFN	0.32	0.17	0.24	0.13	0.22	0.09	0.22	0.15	0.26
GAN	0.97	0.54	0.26	0.28	0.34	0.18	0.27	0.23	0.32

Monthly Sharpe Ratio (SR) of the SDF factor, explained time series variation (EV) and cross-sectional mean R^2 for the GAN, FFN, EN and LS model. The model is estimated on all stocks but evaluated on stocks with market capitalization larger than 0.01% or 0.001% of the total market capitalization.

Table VIII reports the model performance for these two subsets of the data. The SDF weights are obtained on all individual stocks, but the Sharpe-ratio and the explained time-series and cross-sectional variation is calculated on stocks with market cap larger than 0.001% respectively 0.01% of the total market capitalization. As expected the Sharpe ratios decline, but GAN still achieves an annual out-of-sample Sharpe ratio of 1.4 using only the 1,500 largest stocks. In contrast, the linear models collapse. Based on the 550 largest stocks the Sharpe ratio of GAN falls to 0.9, but is still larger than for the other models. Most importantly the explained variation of GAN is two to three times higher than for the linear or deep learning prediction model. Similarly, the gap in the cross-sectional R^2 is substantially wider on the larger stocks than on the whole sample. This suggests that FFN and the linear models are mainly fitting small stocks, while GAN also finds the systematic structure in the large cap stocks.

Table IX Different SDF Models Estimated on Large Market Cap Stocks

	SR			EV			Cross-Sectional R^2		
Model	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
LS	1.91	0.40	0.19	0.08	0.06	0.04	0.18	0.05	0.12
EN	1.34	0.92	0.42	0.13	0.13	0.07	0.23	0.09	0.19
FFN	0.37	0.19	0.28	0.13	0.13	0.07	0.21	0.10	0.21
GAN	3.57	1.18	0.42	0.24	0.23	0.14	0.23	0.13	0.26

Monthly Sharpe Ratio (SR) of the SDF factor, explained time series variation (EV) and cross-sectional mean R^2 for the GAN, FFN, EN and LS model. The models are estimated and evaluated on stocks with market capitalization larger than 0.001% of the total market capitalization.

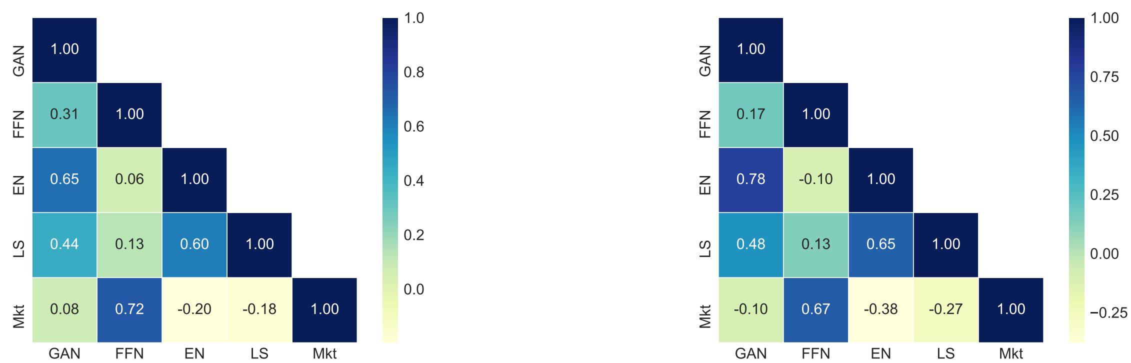
Table IX estimates and evaluates the different model on stocks with market capitalization larger than 0.001% of the total market capitalization.[45]The performance of GAN is essentially identical, suggesting that our approach finds the same SDF structure conditioned on large cap stocks if it is trained on all stocks or only the large stocks. In this sense our model is robust to the size of the companies. In contrast, the elastic net approach performs significantly better on large cap stocks when estimated on this sample. This is evidence that it overfits small stocks when applied to the whole sample in contrast to our approach. The prediction approach has a very similar performance on the large cap stocks when estimated on this subset or on the whole data set. This is indicative that it cannot capture the structure in large cap stocks. Even when optimally trained on the subset of large cap stocks the linear and prediction approach explain substantially less time-series and cross-sectional variation than GAN.

[45]We estimate the optimal tuning parameters for the model restricted to the large cap stocks. Using the same tuning parameters as for the total sample yields identical results.

F. *Variable Importance*

What is the structure of the SDF factor? As a first step in Figure 12 we calculate the correlation between different factors implied by different methods. It is apparent that the factors for each model are different. Our GAN factor has the highest correlation with the elastic net factor, i.e. with the factor based on the same model but restricted to a linear form. The GAN factor has only a correlation of 8% with the market factor. Surprisingly the FFN factor has a high correlation of over 70% with the market factor.

Figure 12. Correlation between SDF Factors for Different Models



(a) Whole Time Horizon

(b) Test Period

As a second step we compare the GAN factor with the Fama-French 5 factor model. None of the five factors has a high correlation with our factor with the profitability factor having the highest correlation with 17%. Next, we run a time series regression to explain the GAN factor portfolio with the Fama-French 5 factors. Only the market and profitability factors are significant. The strongly significant pricing error indicates that these factors fail to capture the pricing information in our SDF portfolio.

Table X Correlation of GAN-SDF Factor with Fama-French 5 Factors

	Mkt-RF	SMB	HML	RMW	CMA	intercept
Regression Coefficients	0.00	0.00	-0.04	0.08***	0.04	0.76***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.06)
Correlations	-0.10	-0.09	0.01	0.17	0.05	-

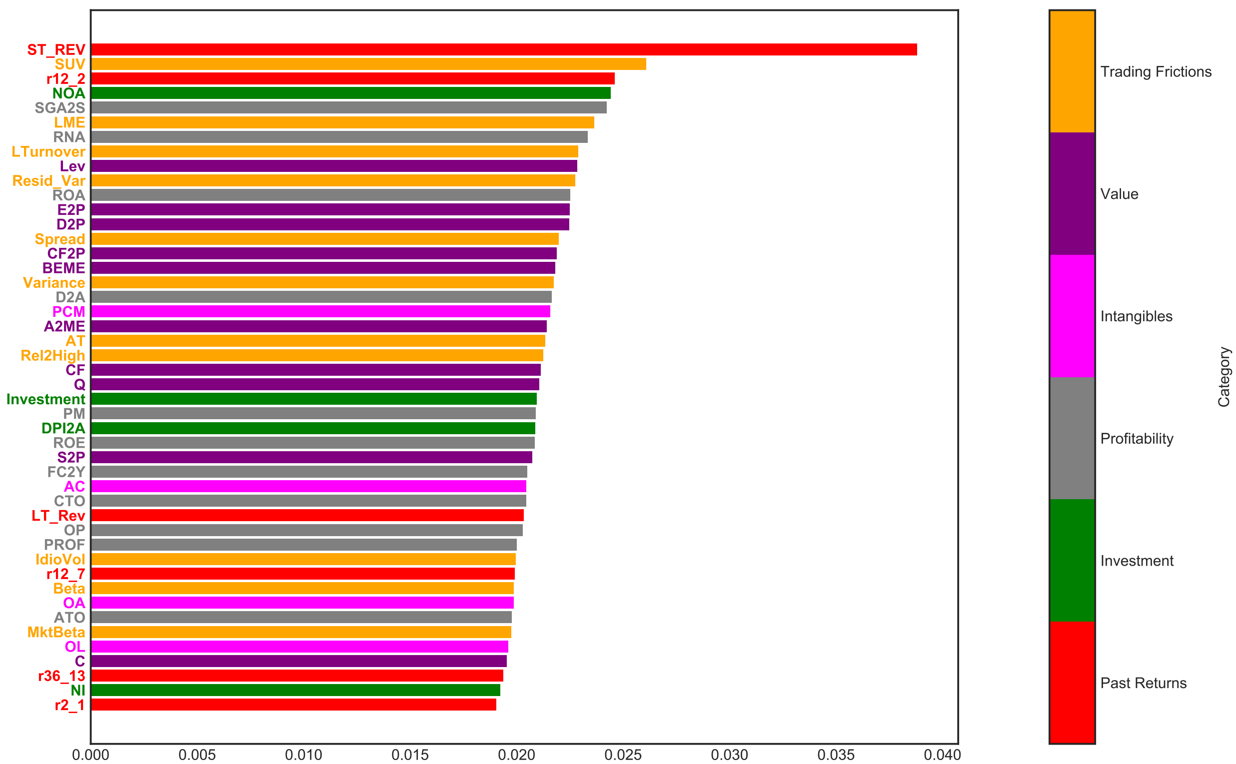
Out-of-sample correlation and regression of GAN SDF factor on the Fama-French 5 factors. The regression intercept is the monthly time series pricing error of the SDF portfolio. Standard errors are in parenthesis.

We rank the importance of firm-specific and macroeconomic variables for the pricing kernel based on the sensitivity of the SDF weight ω with respect to these variables. Our sensitivity analysis is similar to Sirignano et al. (2016) and Horel and Giesecke (2019) and based on the average absolute gradient. More specifically, we define the sensitivity of a particular variable as the average absolute derivative of the weight w with respect to this variable:

$$\text{Sensitivity}(x_j) = \frac{1}{C} \sum_{i=1}^{N} \sum_{t=1}^{T} \left| \frac{\partial w(I_t, I_{t,i})}{\partial x_j} \right|,$$

where C a normalization constant. This simplifies to the standard slope coefficient in the special case of a linear regression framework. A sensitivity of value z for a given variable means that the weight w will approximately change (in magnitude) by $z\Delta$ for a small change of Δ in this variable.

Figure 13. Characteristic Importance by GAN

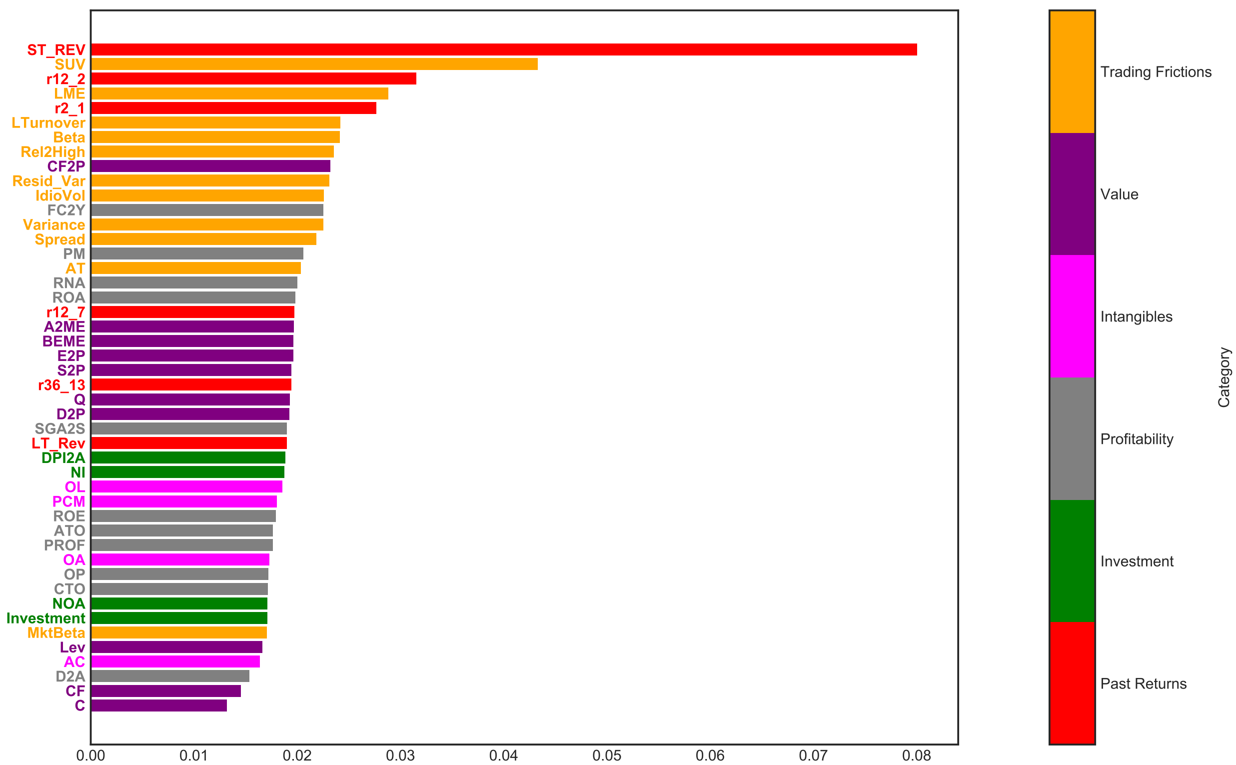


GAN variable importance ranking of the 46 firm-specific characteristics in terms of average absolute gradient (VI) in the test data. The values are normalized to sum up to one.

Figure 13 ranks the variable importance of the 46 firm-specific characteristics for GAN. The sum of all sensitivities is normalized to one. Figures 14, A.4 and A.5 collect the corresponding results for FFN, EN and LS.

Surprisingly, all three models GAN, FFN and EN select trading frictions and past returns as being the most relevant categories. The most important variables for GAN are Short-Term Reversal (`ST_REV`), Standard Unexplained Volume (`SUV`) and Momentum (`r12_2`). Importantly, for GAN all 6 categories are represented among the first 20 variables, which includes size, value, investment and profitability characteristics. The SDF composition is different for FNN, where the first 14 characteristics are almost only in the trading friction and past return category. More specifically, this SDF loads heavily on short-term reversal, illiquidity measured by unexplained volume and size, which confirms the suspicion that a simple forecasting approach focuses mainly on illiquid penny stocks. This in line with the findings of the previous section and Table A.2. The no-arbitrage condition is necessary to discipline the model to capture the pricing information in other characteristics. The linear model with regularization also selects variables from all six categories among the first 9 variables. Note, that the elastic net penalty will remove characteristics that are close substitutes, e.g. as the dividend-price ratio (D2P) and book-to-market ratio (BEME) capture similar information, the regularized model only selects one of them. The linear model without regularization cannot handle the large number of variables and not surprisingly results in a different ranking.

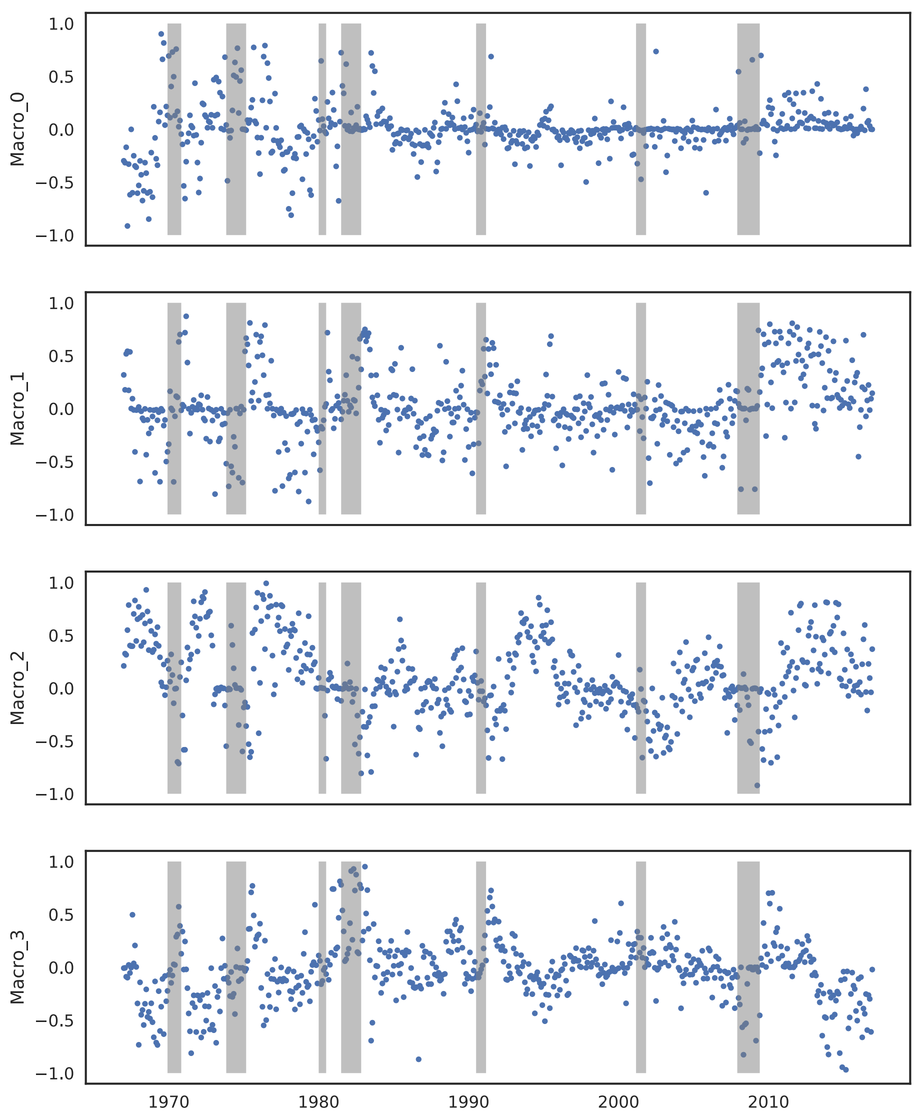
Figure 14. Characteristic Importance for FFN



FFN variable importance ranking of the 46 firm-specific characteristics in terms of average absolute gradient (VI) in the test data. The values are normalized to sum up to one.

Figure A.3 shows the importance of the macroeconomic variables for the GAN model. These variables are first summarized into the four hidden states processes before they enter the weights of the SDF factor. First, it is apparent that most macroeconomic variables have a very similar importance. This is in line with a model where there is a strong dependency between the macroeconomic time series which is driven by a low dimensional non-linear factor structure. A simple example would be the factor model in Ludvigson and Ng (2009) where the information in a macroeconomic data set very similar to ours is summarized by a small number of PCA factors. As the first PCA factor is likely to pick up a general economic market trend, it would affect all variables. If the SDF structure depends on this PCA factor, all macroeconomic variables will appear to be important (of potentially similar magnitude). It is important to keep in mind that a simple PCA analysis of the macroeconomic variables does not work in our asset pricing context. The reason is that the PCA factors would mainly be based on increments of the macroeconomic time series and hence would not capture the dynamic pattern.[46] The two most relevant variables that stand out in our importance ranking are the median bid-ask spread (`Spread`) and the federal fund rate (`FEDFUNDS`). This can be interpreted as capturing the overall economic activity level and overall market volatility.

Figure 15. Macroeconomic Hidden State Processes (LSTM Outputs)



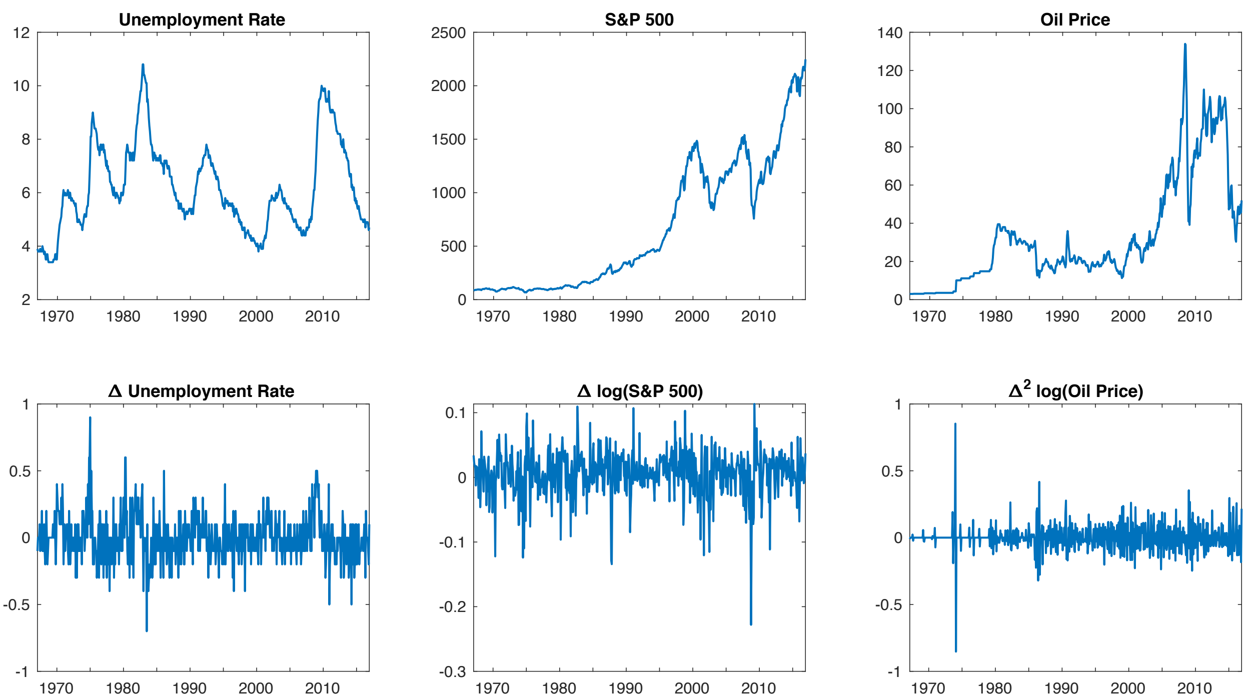
Macroeconomic hidden state processes from GAN. The gray areas mark NBER recession periods.

[46]The results for PCA based macroecononimc factors are available upon request. We also want to clarify that for other applications PCA based factors based on macroeconomic time-series might actually capture the relevant information.

We show that the hidden macroeconomic states are closely linked to business cycles and overall economic activity. Figure 15 plots the time series of the four hidden macroeconomic state variables. These variables are the outputs from the LSTM that encodes the history of macroeconomic information. The grey shaded areas indicate NBER recessions.[47] First, it is apparent that the state variables, in particular for the third and fourth state, peak during times of recessions. Second, the state processes seem to have a cyclical behavior which confirms our intuition that the relevant macroeconomic information is likely to be related to business cycles. The cycles and peaks of the different state variables do not coincide at all times indicating that they capture different macroeconomic risks.

Figure 16 are three examples of the complex dynamics and non-stationarities in the macroeconomic time series. We plot the time series of the U.S. unemployment rate, the S&P 500 price and the oil price together with the standard transformations proposed by McCracken and Ng (2016) to remove the obvious non-stationarities. Using only the last observation of the differenced data obviously results in a loss of information and cannot identify the cyclical dynamic patterns.

Figure 16. Examples of Macroeconomic Variables



Examples of macroeconomic time series with standard transformations proposed by McCracken and Ng (2016)

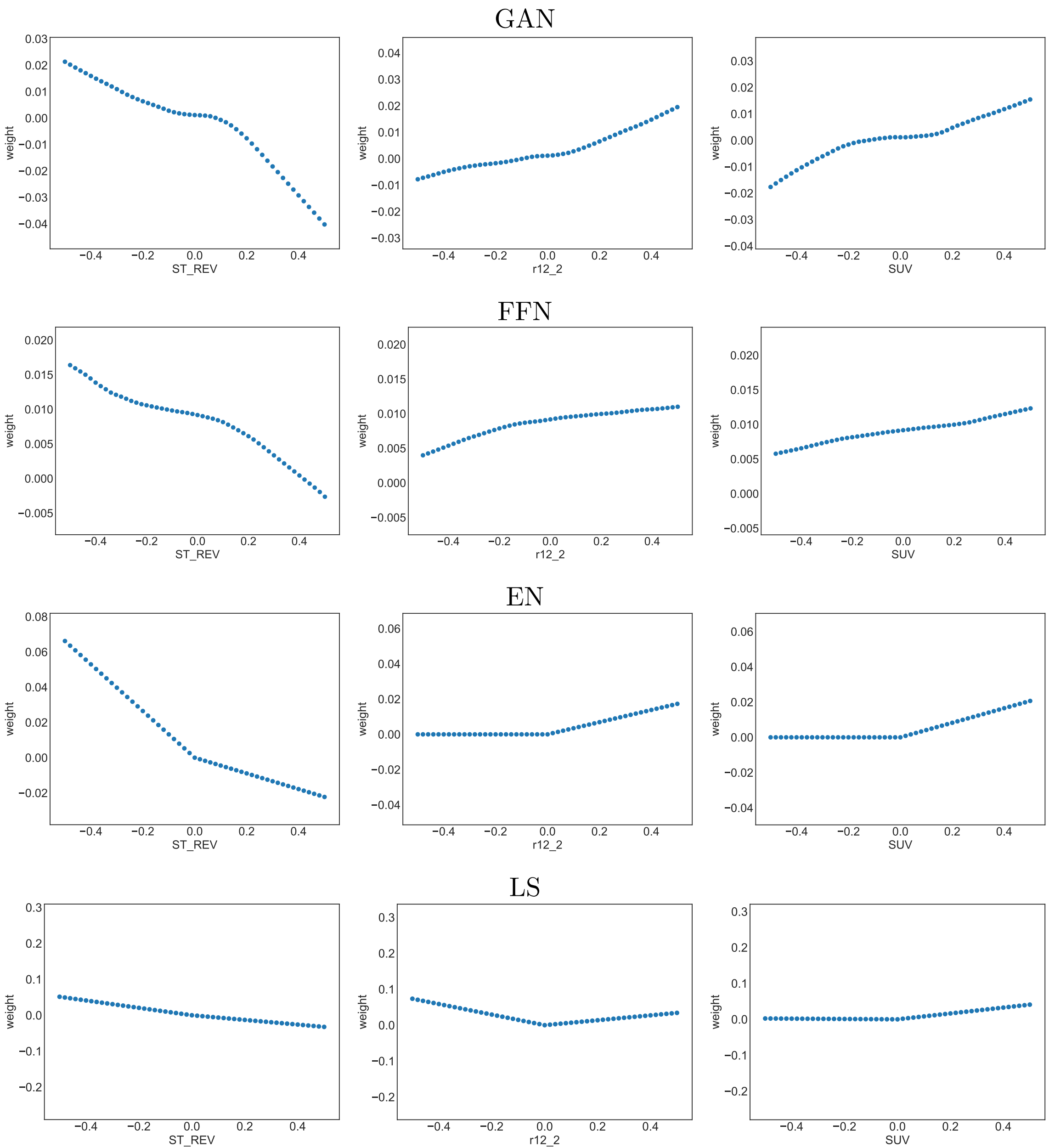
G. *SDF Structure*

We study the structure of the SDF weights and betas as a function of the characteristics. Our main findings are two-fold: Surprisingly, individual characteristics have an almost linear effect on

[47] NBER based Recession Indicators for the United States from the Peak through the Trough are taken from https://fred.stlouisfed.org/series/USRECM. Table A.VI provides a more detailed description of the recessions.

the pricing kernel and the risk loadings, i.e. non-linearities matter less than expected for individual characteristics. Second, the better performance of GAN is explained by non-linear interaction effects, i.e. the general functional form of our model is necessary for capturing the dependency between multiple characteristics.[48]

Figure 17. SDF weight ω as a Function of Covariates for Different Models

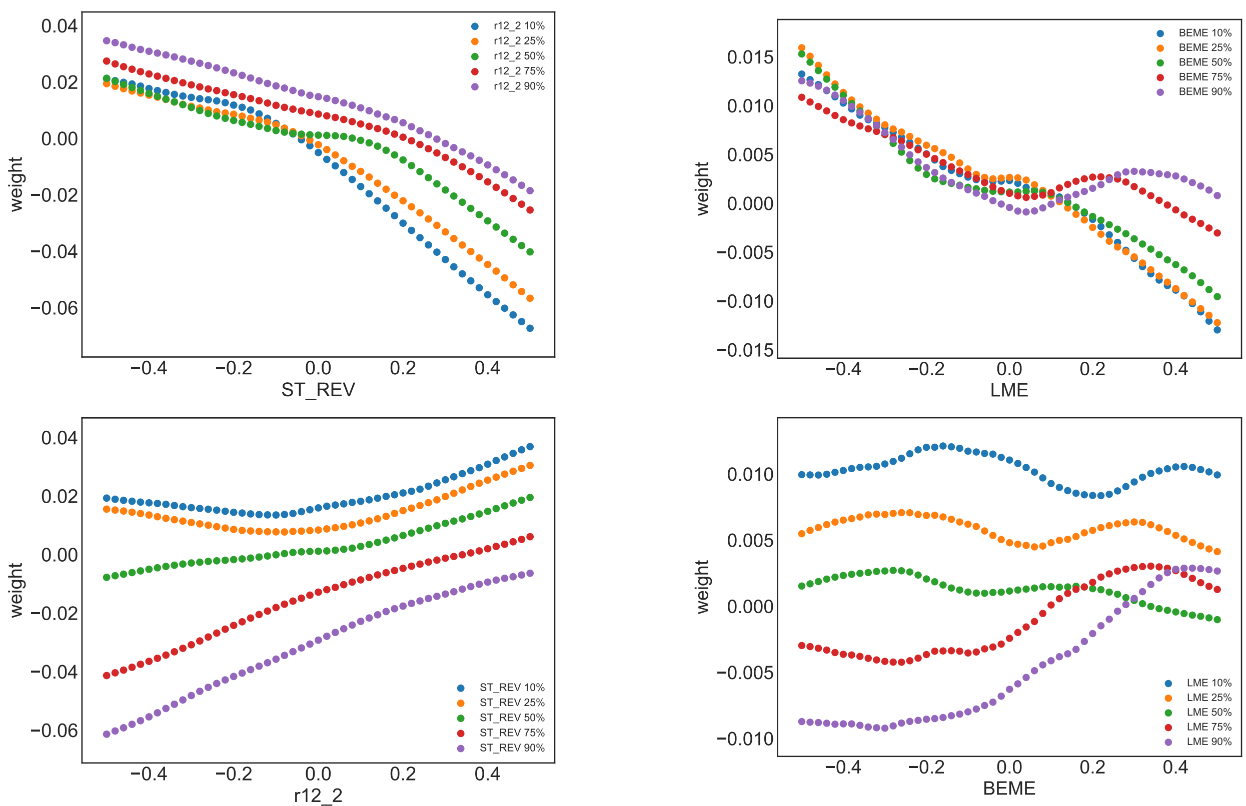


SDF weight ω as a one-dimensional function of covariates keeping the other covariates at their mean level. The covariates are Short-Term Reversal (ST_REV), Momentum (r12_2) and Standard Unexplained Volume (SUV).

[48]The Internet Appendix collects the results for additional characteristics with the same findings.

Figure 17 plots the one-dimensional relationship between the SDF weights ω and one specific characteristic. The other variables are fixed at their mean values. In the case of a linear model these plots simply show the slope of a linear regression coefficient. As have include a separate long and short leg for the linear model, we allow for a kink at the median value. Otherwise the linear model would simply be a straight line. For the non-linear GAN and FFN the one-dimensional relationship can take any functional form. We show the univariate functional form for the three most relevant characteristics in Figure 17, while the Internet Appendix collects the results for the other characteristics. It is striking how close the functional form of the SDF for GAN and FFN is to a linear function. This explains why linear models are actually so successful in explaining single-sorted characteristics. For a small number of characteristics, mainly short-term reversal and momentum, GAN has some non-linearities that allow for a higher slope at slope at the extreme end. These are exactly the decile sorted portfolios for which GAN performs better than FFN and EN. However, for most characteristics the pricing kernel depends linearly on the characteristics as long as we consider a one-dimensional relationship. However, it seems to be relevant to allow the low and high quantiles to have different linear slopes. The linear model without regularization obtains a relationship for some characteristics that is completely out of line with the other models. Given the worse overall performance of LS, this suggests that LS suffers from severe over-fitting.

Figure 18. SDF weight ω as a Function of Covariates for GAN



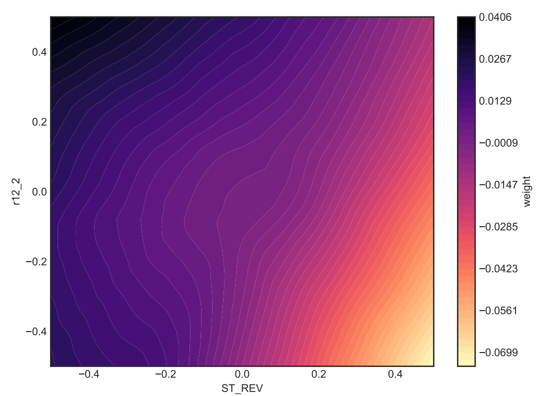
(a) Interaction between Short-Term Reversal (ST_REV) and Momentum (r12_2)

(b) Interaction between Size (LME) and Book to Market Ratio (BEME)

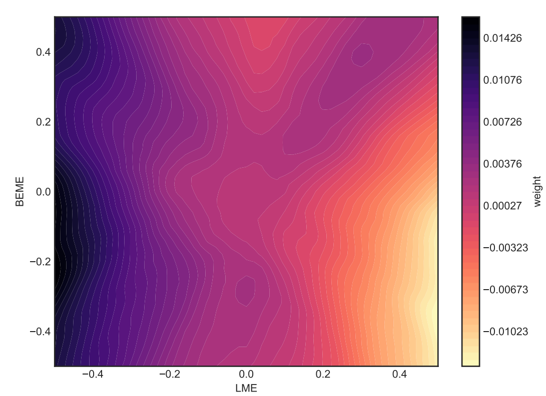
Figures 18, 19 and 20 show the crucial finding for this section. Non-linearities matter for interactions. Here we plot the 2- respectively 3-dimensional functional form of ω when we fix all but two or three variables at their mean. A linear model like EN also assumes an additive effect of different characteristics on the pricing kernel, i.e. small-value stocks cannot have a different exposure to the size characteristics than small stocks. Both, GAN and FNN, relax this condition and allow for general interaction effects. However, the simulation already suggested that the no-arbitrage condition of GAN helps in identifying relevant interaction effects that are not captured by FFN. Indeed, the line plots and heatmaps for GAN reveal more complex interaction patterns than for the other models.

Figure 18 plots the SDF weight of one characteristic conditioned on a quantile of a second characteristic. In an additive model without interaction all lines would be parallel shifts. This is exactly what we see for the two linear models in Figures A.8 and A.9.[49] Interestingly, for size and value, the FFN model in Figure A.7 also has almost parallel shifts in the SDF weights, implying that it does not capture interactions. However, for GAN small stocks have a very different exposure to value than large cap stocks. Note, that while fixing the second characteristic at the median the lines are close to linear, the shape can become non-linear when conditioning the second characteristic on an extreme quantile.

Figure 19. SDF weight ω as a Function of Covariates for GAN



(a) Interaction between Short-Term Reversal (ST_REV) and Momentum (r12_2)



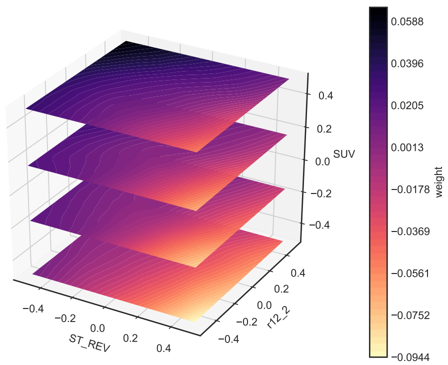
(b) Interaction between Size (LME) and Book to Market Ratio (BEME)

Instead of conditioning on only five quantiles, we plot the whole two-dimensional pricing kernel for GAN in Figure 19. It confirms that the combined size and book-to-market characteristics have a highly non-linear effect on the GAN pricing kernel. The triple interaction in Figure 20 shows that low short-term reversal, high momentum and high explained volume has the highest positive weight while high reversal, low momentum and low unexplained volume has the largest negative weight in

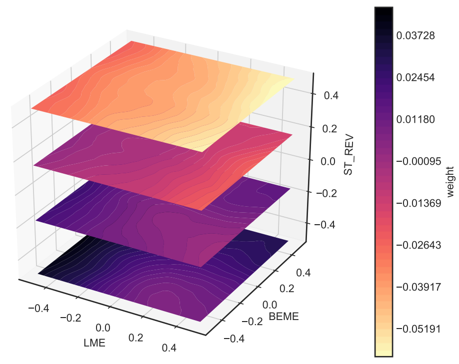
[49] As the linear model with regularization removes variables, it is possible that the SDF weights for one characteristic conditioned on different quantiles of the second characteristic collapse to one line.

the kernel when conditioning on these three characteristics. Low reversal and low momentum or high reversal and high momentum have an almost neutral effect independent of unexplained volume. The interaction effect for size, book-to-market and short-term reversal is even more complicated.

Figure 20. SDF weight ω as a Function of Covariates for GAN



(a) Interaction between Short-Term Reversal (ST_REV), Momentum (r12_2) and Standard Unexplained Volume (SUV)



(b) Interaction between Size (LME), Book to Market Ratio (BEME) and Short-Term Reversal (ST_REV)

VI. Conclusion

We propose a new way to estimate asset pricing models for individual stock returns that can take advantage of the vast amount of conditioning information, while keeping a fully flexible form and accounting for time-variation. For this purpose, we combine three different deep neural network structures in a novel way: A feedforward network to capture non-linearities, a recurrent (LSTM) network to find a small set of economic state processes, and a generative adversarial network to identify the portfolio strategies with the most unexplained pricing information. Our crucial innovation is the use of the no-arbitrage condition as part of the neural network algorithm. We estimate the stochastic discount factor that explains all stock returns from the conditional moment constraints implied by no-arbitrage. Our SDF is a portfolio of all traded assets with time-varying portfolio weights which are general functions of the observable firm-specific and macroeconomic variables. Our model allows us to understand what are the key factors that drive asset prices, identify mis-pricing of stocks and generate the mean-variance efficient portfolio.

Our primary conclusions are four-fold. First, we demonstrate the potential of machine learning methods in asset pricing. We are able to identify the key factors that drive asset prices and the functional form of this relationship on a level of generality and with an accuracy that was not possible with traditional econometric methods. Second, we show and quantify the importance of including a no-arbitrage condition in the estimation of machine learning asset pricing models. The "kitchen-sink" prediction approach with deep learning does not outperform a linear model with no-arbitrage constraints. This illustrates that a successful use of machine learning methods in finance requires both subject specific domain knowledge and a state-of-the-art technical implementation. Third, financial data have a time dimension which has to be taken into account accordingly. Even the most flexible model cannot compensate for the problem if the data is inputted in the wrong format. Standard econometrics techniques of differencing the data to ensure stationarity might lose the information that is essential for the asset pricing model. We show that macroeconomic conditions matter for asset pricing and can be summarized by a small number of economic state variables, which depend on the whole dynamics of all time series. Fourth, asset pricing is actually surprisingly "linear". As long as we consider anomalies in isolation the linear factor models provide a good approximation. However, the multi-dimensional challenge of asset pricing cannot be solved with linear models and requires a different set of tools.

Our results have direct practical benefits for asset pricing researchers that go beyond our empirical findings. First, we provide a new set of benchmark test assets. New asset pricing models can be tested on explaining our SDF factor portfolio respectively the portfolios sorted according to the risk exposure in our model. These test assets incorporate the information of all characteristics and macroeconomic information in a small number of assets. Explaining portfolios sorted on a single characteristic is not a high hurdle to pass. Second, we provide a set of macroeconomic time series of hidden states that encapsulate the relevant macroeconomic information for asset pricing. These time series can also be used as an input for new asset pricing models.[50]

Last but not least, our model is directly valuable for investors and portfolio managers. The main output of our model is the risk measure β and the SDF factor weight ω as a function of characteristics and macroeconomic variables. Given our estimates, the user of our model can assign a risk measure and its portfolio weight to an asset even if it does not have a long time series available.

REFERENCES

Ang, A., Hodrick, R. J., Xing, Y., Zhang, X., 2004. The cross-section of volatility and expected returns. Working Paper 10852, National Bureau of Economic Research.

Arjosvky, M., Chintala, S., Leon, B., 2017. Wasserstein GAN. Working paper.

Balakrishnan, K., Bartov, E., Faurel, L., 2010. Post loss/profit announcement drift. Journal of Accounting and Economics 50 (1), 20–41.

Ball, R., Gerakos, J., Linnainmaa, J. T., Nikolaev, V. V., 2015. Deflating profitability. Journal of Financial Economics 117 (2), 225 – 248.

Bansal, R., Viswanathan, S., 1993. No arbitrage and arbitrage pricing: A new approach. The Journal of Finance 48 (4), 1231–1262.

Basu, S., 1983. The relationship between earnings' yield, market value and return for nyse common stocks: Further evidence. Journal of Financial Economics 12 (1), 129–156.

[50] The data is available on https://mpelger.people.stanford.edu/research.

Bhandari, L. C., 1988. Debt/equity ratio and expected common stock returns: Empirical evidence. The Journal of Finance 43 (2), 507–528.

Bianchi, D., Büchner, M., Tamoni, A., 2019. Bond risk premia with machine learning. Working paper.

Blanchet, J., Kang, Y., Murthy, K., 2016. Robust Wasserstein Profile Inference and Applications to Machine Learning. Working paper.

Bondt, W. F. M. D., Thaler, R., 1985. Does the stock market overreact? The Journal of Finance 40 (3), 793–805.

Bustamante, M. C., Donangelo, A., 2017. Product market competition and industry returns. The Review of Financial Studies 30 (12), 4216–4266.

Chamberlain, G., 1987. Asymptotic efficiency in estimation with conditional moment restrictions. Journal of Econometrics 34 (3), 305–334.

Chung, K. H., Zhang, H., 2014. A simple approximation of intraday spreads using daily data. Journal of Financial Markets 17 (C), 94–120.

Cooper, M. J., Gulen, H., Schill, M. J., 2008. Asset growth and the cross-section of stock returns. Journal of Finance 63 (4), 1609–1651.

D'Acunto, F., Liu, R., Pflueger, C., Weber, M., 2017. Flexible prices and leverage. Working Paper 23066, National Bureau of Economic Research.

Datar, V. T., Naik, N. Y., Radcliffe, R., 1998. Liquidity and stock returns: An alternative test. Journal of Financial Markets 1 (2), 203 – 219.

Desai, H., Rajgopal, S., Venkatachalam, M., 2004. Value-glamour and accruals mispricing: One anomaly or two? The Accounting Review 79 (2), 355–385.

Fama, E. F., French, K. R., 1992. The cross-section of expected stock returns. The Journal of Finance 47 (2), 427–465.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33 (1), 3–56.

Fama, E. F., FrencH, K. R., 1996. Multifactor explanations of asset pricing anomalies. The Journal of Finance 51 (1), 55–84.

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116 (1), 1 – 22.

Fama, E. F., MacBeth, J. D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy 81 (3), 607–636.

Feng, G., He, J., Polson, N. G., 2018a. Deep learning for predicting asset returns. arXiv preprint arXiv:1804.09314.

Feng, G., Polson, N. G., Xu, J., 2018b. Deep learning factor alpha. arXiv preprint arXiv:1805.01104.

Frazzini, A., Pedersen, L. H., 2014. Betting against beta. Journal of Financial Economics 111 (1), 1 – 25.

Freyberger, J., Neuhierl, A., Weber, M., 2017. Dissecting characteristics nonparametrically. Working Paper 23227, National Bureau of Economic Research.

Gandhi, P., Lustig, H., 2015. Size anomalies in u.s. bank stock returns. The Journal of Finance 70 (2), 733–768.

Garfinkel, J. A., 2009. Measuring investors' opinion divergence. Journal of Accounting Research 47 (5), 1317–1348.

George, T. J., Hwang, C.-Y., 2004. The 52-week high and momentum investing. The Journal of Finance 59 (5), 2145–2176.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Advances in neural information processing systems. pp. 2672–2680.

Gorodnichenko, Y., Weber, M., 2016. Are sticky prices costly? evidence from the stock market. American Economic Review 106 (1), 165–99.

Gu, S., Kelly, B., Xiu, D., 2019. Autoencoder asset pricing models. Working paper.

Gu, S., Kelly, B. T., Xiu, D., 2018. Empirical asset pricing via machine learning. Working Paper 25398, National Bureau of Economic Research.

Hansen, L. P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50 (4), 1029–1054.

Haugen, R. A., Baker, N. L., 1996. Commonality in the determinants of expected stock returns. Journal of Financial Economics 41 (3), 401 – 439.

Heaton, J., Polson, N., Witte, J. H., 2017. Deep learning for finance: deep portfolios. Applied Stochastic Models in Business and Industry 33 (1), 3–12.

Hirshleifer, D., Hou, K., Teoh, S. H., Zhang, Y., 2004. Do investors overvalue firms with bloated balance sheets? Journal of Accounting and Economics 38, 297 – 331, conference Issue on Research on Market Efficiency, Valuation, and Mispricing: Risk, Behavioral, and Alternative Explanations.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Computation 9 (8), 1735–1780.

Horel, E., Giesecke, K., 2019. Towards explainable ai: Significance tests for neural networks. arXiv preprint arXiv:1902.06021.

Hou, K., Karolyi, G. A., Kho, B.-C., 2011. What factors drive global stock returns? The Review of Financial Studies 24 (8), 2527–2574.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. The Journal of Finance 48 (1), 65–91.

Jegadeesh, N., Titman, S., 2001. Profitability of momentum strategies: An evaluation of alternative explanations. The Journal of Finance 56 (2), 699–720.

Kaldor, N., 1966. Marginal productivity and the macro-economic theories of distribution: Comment on samuelson and modigliani. The Review of Economic Studies 33 (4), 309–319.

Kelly, B., Pruitt, S., Su, Y., 2018. Characteristics are covariances: A unified model of risk and return. Journal of Financial Economics, forthcoming.

Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. CoRR abs/1412.6980.

Kozak, S., Nagel, S., Santosh, S., 2018. Shrinking the cross section. Journal of Financial Economics, forthcoming.

Lettau, M., Pelger, M., 2018. Factors that fit the time-series and cross-section of stock returns. Working paper.

Lewellen, J., 2015. The Cross-section of Expected Stock Returns. Critical Finance Review 4 (1), 1–44.

Lewis, G., Syrgkanis, V., 2018. Adversarial generalized method of moments. arXiv preprint arXiv:1803.07164.

Litzenberger, R. H., Ramaswamy, K., 1979. The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. Journal of Financial Economics 7 (2), 163–195.

Ludvigson, S., Ng, S., 2007. The empirical risk return relation: A factor analysis approach. Journal of Financial Economics 83 (1), 171–222.

Ludvigson, S., Ng, S., 2009. Macro factors in bond risk premia. Review of Financial Studies 22 (12), 5027–5067.

Lyandres, E., Sun, L., Zhang, L., 2008. The new issues puzzle: Testing the investment-based explanation. The Review of Financial Studies 21 (6), 2825–2855.

McCracken, M. W., Ng, S., 2016. Fred-md: A monthly database for macroeconomic research. Journal of Business & Economic Statistics 34 (4), 574–589.

Messmer, M., 2017. Deep learning and the cross-section of expected returns. Working paper.

Moritz, B., Zimmerman, T., 2016. Tree-based conditional portfolio sorts: The relation between past and future stock returns. Working paper.

Novy-Marx, R., 2011. Operating leverage. Review of Finance 15 (1), 103–134.

Novy-Marx, R., 2012. Is momentum really momentum? Journal of Financial Economics 103 (3), 429–453.

Palazzo, B., 2012. Cash holdings, risk, and expected returns. Journal of Financial Economics 104 (1), 162–185.

Pelger, M., 2019. Understanding systematic risk: A high-frequency approach. Journal of Finance, forthcoming.

Pelger, M., Xiong, R., 2018a. Interpretable sparse proximate factors for large dimensions. Working paper.

Pelger, M., Xiong, R., 2018b. State-varying factor models of large dimensions. Working paper.

Pontiff, J., Woodgate, A., 2008. Share issuance and cross-sectional returns. The Journal of Finance 63 (2), 921–945.

Rossi, A. G., 2018. Predicting stock market returns with machine learning. Working paper.

Ruder, S., 2016. An overview of gradient descent optimization algorithms. CoRR abs/1609.04747.

Sirignano, J., Sadhwani, A., Giesecke, K., 2016. Deep learning for mortgage risk. Working paper.

Sloan, R. G., 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? The Accounting Review 71 (3), 289–315.

Soliman, M. T., 2008. The use of dupont analysis by market participants. The Accounting Review 83 (3), 823–853.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research 15, 1929–1958.

Wager, S., Wang, S., Liang, P. S., 2013. Dropout training as adaptive regularization. Advances in Neural Information Processing Systems 26, 351–359.

Welch, I., Goyal, A., 2007. A comprehensive look at the empirical performance of equity premium prediction. The Review of Financial Studies 21 (4), 1455–1508.

Appendix A. Estimation Method

Figure A.1. Feedforward Network with 3 Hidden Layers and Dropout

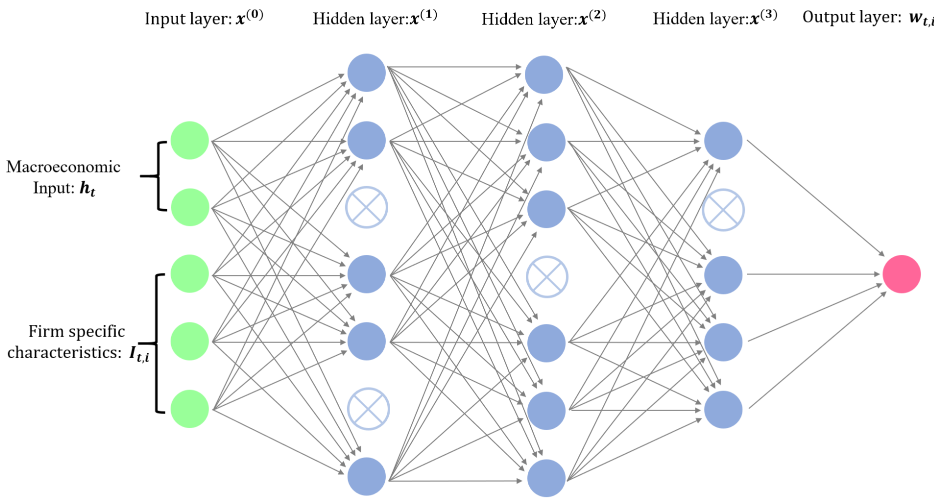
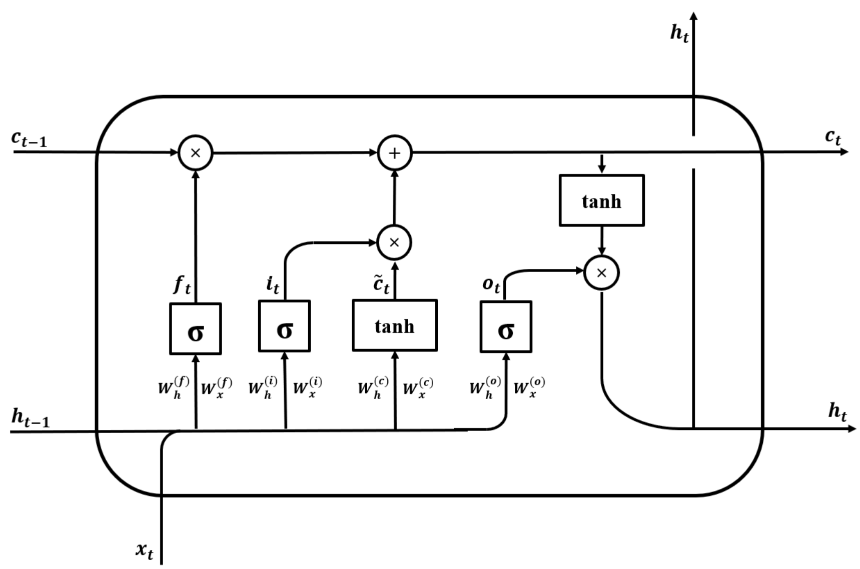


Figure A.2. Structure of an LSTM Unit



Appendix B. Cross-section of Individual Stock Returns

Table A.I Sharpe Ratio of Long-Short Portfolios with FFN

Quantile	SR (Train)	SR (Valid)	SR (Test)
	(i) Equally-Weighted		
1%	1.24	0.65	0.66
5%	1.36	1.10	0.71
10%	1.30	1.31	0.67
25%	1.19	1.20	0.57
50%	1.09	1.26	0.52
	(ii) Value-Weighted		
1%	0.98	0.35	0.39
5%	0.89	0.71	0.42
10%	0.70	0.45	0.32
25%	0.55	0.28	0.17
50%	0.43	0.20	0.15

Monthly Sharpe Ratios of long-short portfolios based on the extreme deciles of returns predicted by FFN. The model is a 3-layer feedforward network, and the hidden layers have 32, 16 and 8 neurons. The predictors are 46 firm-specific characteristics. The stocks are sorted into quantiles (1%, 5%, 10%, 25% and 50%) based on model's forecasts. A zero-net-investment portfolio is constructed by buying the highest expected return stocks and selling the lowest with equal weights or value-weighted by market capitalization.

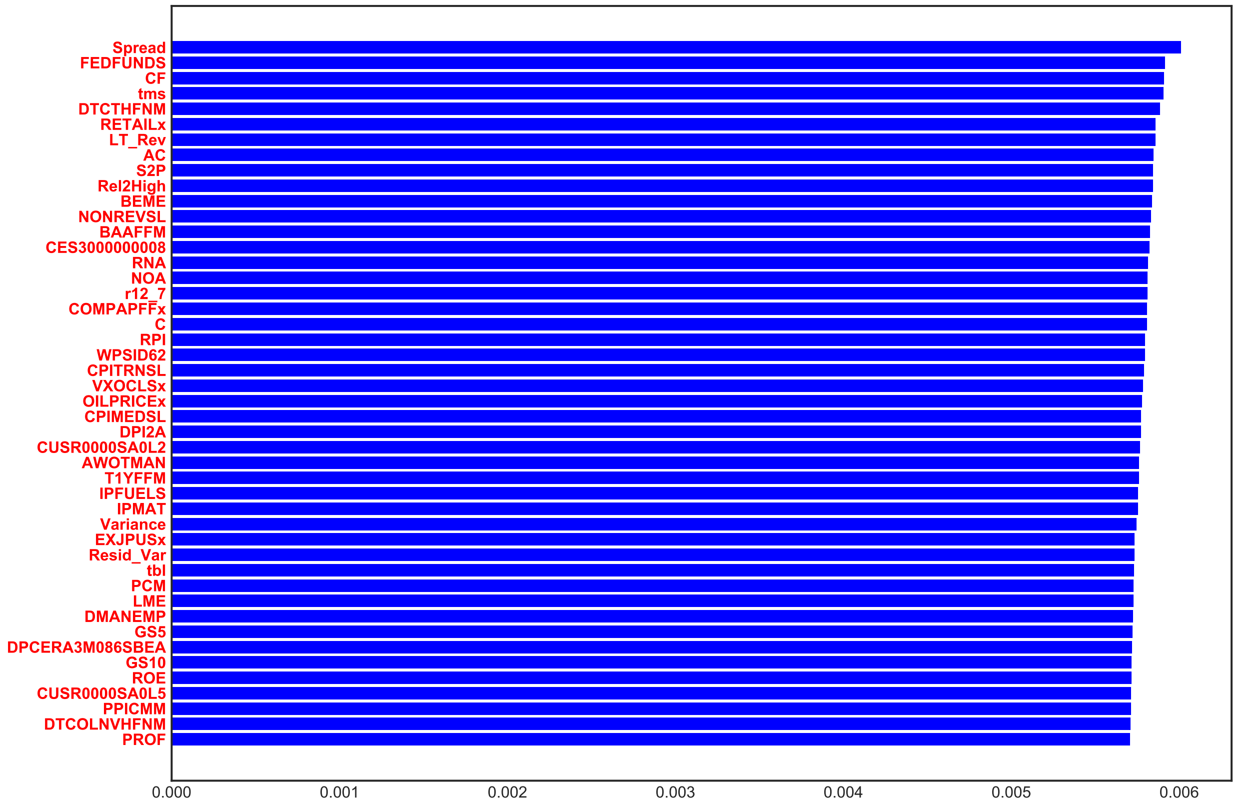
Table A.II Turnover by Models

	Long Position			Short Position		
Model	Train	Valid	Test	Train	Valid	Test
LS	0.25	0.22	0.24	0.64	0.55	0.61
EN	0.36	0.35	0.35	0.83	0.83	0.84
FFN	0.69	0.63	0.65	1.38	1.29	1.27
GAN	0.47	0.40	0.40	1.05	1.04	1.02

Turnover for positions with positive and negative weights for the SDF factor portfolio. It is defined as $\frac{1}{T}\sum_{t=1}^{T}(\sum_i |(1+R_{P,t+1})w_{i,t+1} - (1+R_{i,t+1})w_{i,t}|)$, where $w_{i,t}$ is the portfolio weight of stock i at time t, and $R_{P,t+1} = \sum_i R_{i,t+1}w_{i,t}$ is the corresponding portfolio return. Long and short positions are calculated separately, and the portfolio weights are normalized to $\|w_t\|_1 = 1$.

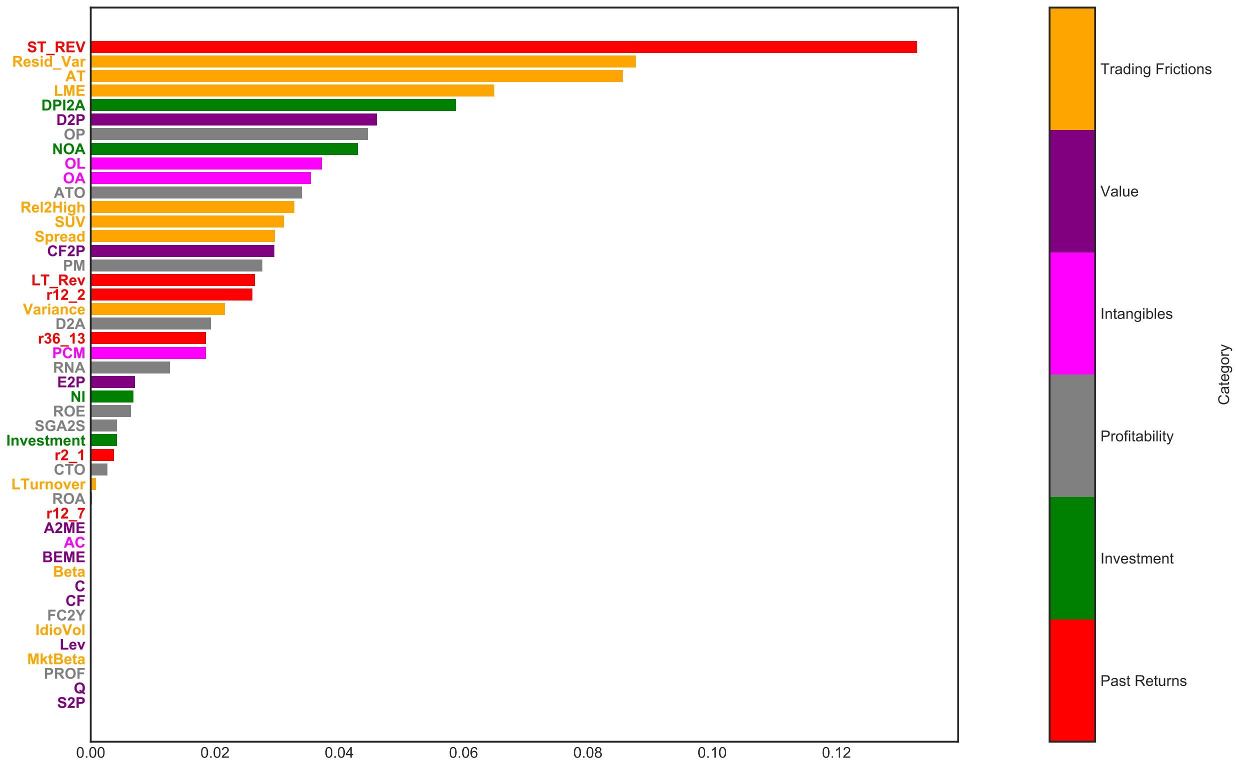
Appendix C. Variable Importance

Figure A.3. Macroeconomic Variable Importance for GAN



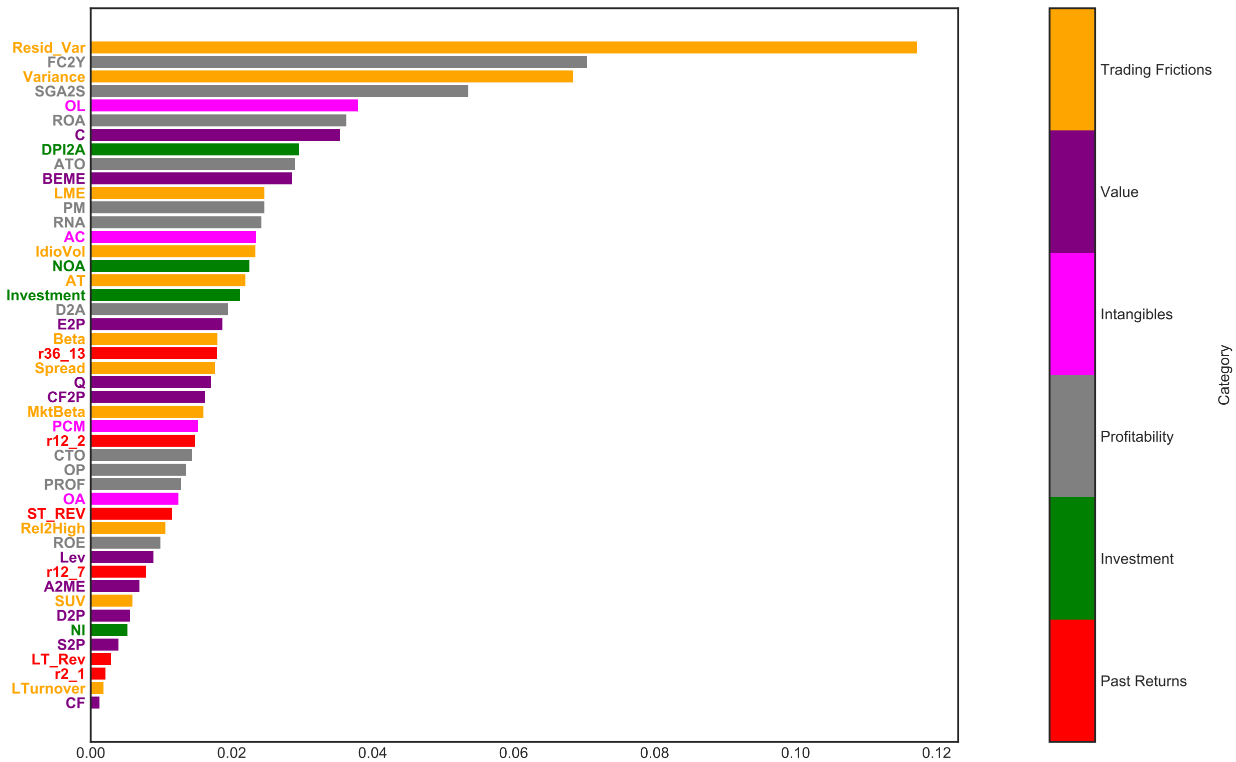
GAN variable importance ranking of the 178 macroeconomic variables in terms of average absolute gradient (VI) in the test data. The values are normalized to sum up to one.

Figure A.4. Characteristic Importance for Elastic Net



Elastic Net variable importance ranking of the 46 firm-specific characteristics in terms of average absolute gradient (VI) in the test data. The values are normalized to sum up to one.

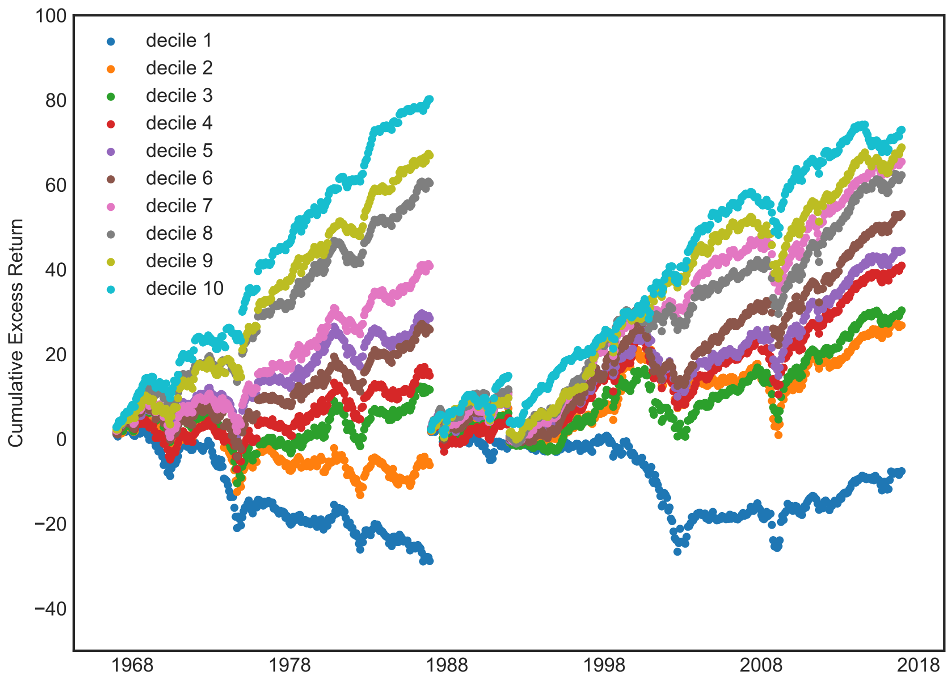
Figure A.5. Characteristic Importance for LS



LS variable importance ranking of the 46 firm-specific characteristics in terms of average absolute gradient (VI) in the test data. The values are normalized to sum up to one.

Appendix D. Predictive Performance

Figure A.6. Cumulative Excess Return of Value Weighted Decile β Portfolios with GAN



Cumulative excess return of decile sorted portfolios based on the risk loadings β for GAN. The first portfolio is based on the smallest decile of risk loadings, while the last decile portfolio is constructed with the largest loading decile. Within each decile the stocks are value weighted.

Table A.III Time Series Pricing Errors for Value Weighted β-Sorted Portfolios

	Average Returns		Market-Rf				Fama-French 3				Fama-French 5			
	Whole	Test	Whole		Test		Whole		Test		Whole		Test	
Decile			α	t	α	t	α	t	α	t	α	t	α	t
1	-0.04	-0.02	-0.11	-6.10	-0.12	-3.87	-0.11	-5.99	-0.12	-3.90	-0.10	-5.14	-0.10	-3.28
2	0.03	0.05	-0.03	-2.87	-0.02	-1.19	-0.03	-2.28	-0.02	-0.91	-0.02	-2.07	-0.01	-0.72
3	0.05	0.06	-0.01	-1.43	-0.02	-1.01	-0.00	-0.48	-0.01	-0.34	-0.00	-0.12	-0.00	-0.05
4	0.06	0.07	-0.00	-0.50	0.00	0.13	0.00	0.49	0.01	0.92	0.00	0.27	0.01	0.89
5	0.08	0.08	0.02	2.04	0.01	0.52	0.02	2.63	0.01	1.08	0.02	2.07	0.01	0.43
6	0.09	0.10	0.02	2.62	0.02	1.69	0.03	2.86	0.03	2.11	0.02	2.32	0.03	1.67
7	0.12	0.12	0.05	5.23	0.05	3.27	0.05	4.87	0.05	3.24	0.04	3.52	0.03	2.10
8	0.14	0.11	0.08	6.37	0.04	2.71	0.07	5.52	0.04	2.32	0.05	4.10	0.02	1.11
9	0.18	0.15	0.11	6.56	0.07	3.24	0.08	5.47	0.05	2.52	0.06	4.32	0.03	1.39
10	0.29	0.24	0.20	7.20	0.13	3.38	0.15	6.72	0.10	2.88	0.16	6.88	0.11	3.01
10-1	0.33	0.26	0.31	10.00	0.25	5.61	0.26	9.68	0.22	5.23	0.25	9.19	0.22	4.90
GRS Asset Pricing Test			GRS	p	GRS	p	GRS	p	GRS	p	GRS	p	GRS	p
			11.15	0.00	3.94	0.00	10.29	0.00	3.76	0.00	8.80	0.00	2.87	0.00

Average returns, time series pricing errors and corresponding t-statistics for value weighted β-sorted decile portfolios based on GAN. The pricing errors are based on the CAPM and Fama-French 3 and 5 factors models. Returns are annualized. The GRS-test is under the null hypothesis of correctly pricing all decile portfolios and includes the p-values. We consider the whole time period and the test period. Within each decile the stocks are value weighted.

Appendix E. Portfolio Results

Table A.IV Explained Variation and Pricing Errors for Double-Sorted Portfolios based on Short-Term Reversal and Momentum

		EN	FFN	GAN	EN	FFN	GAN
ST_REV	r12_2	Explained Variation			Alpha		
1	1	0.35	0.32	0.62	0.16	0.13	0.08
1	2	0.55	0.48	0.72	-0.02	-0.04	-0.05
1	3	0.66	0.61	0.74	-0.06	-0.07	-0.05
1	4	0.74	0.62	0.67	-0.06	-0.05	-0.02
1	5	0.69	0.58	0.58	-0.10	-0.06	-0.03
2	1	0.17	0.16	0.53	0.22	0.19	0.11
2	2	0.32	0.39	0.67	0.18	0.11	0.08
2	3	0.59	0.61	0.71	0.08	0.03	0.01
2	4	0.72	0.74	0.59	0.00	-0.03	-0.02
2	5	0.56	0.61	0.54	0.08	0.05	0.06
3	1	-0.02	-0.01	0.48	0.18	0.16	0.01
3	2	0.13	0.33	0.65	0.12	0.02	-0.03
3	3	0.41	0.62	0.66	0.13	0.02	-0.00
3	4	0.46	0.60	0.48	0.03	-0.06	-0.07
3	5	0.39	0.53	0.42	0.08	-0.01	-0.02
4	1	-0.24	-0.27	0.31	0.26	0.24	0.06
4	2	-0.24	0.15	0.58	0.14	0.05	-0.04
4	3	0.02	0.51	0.68	0.11	-0.02	-0.06
4	4	0.19	0.53	0.51	0.11	-0.01	-0.04
4	5	0.17	0.47	0.51	0.14	0.02	-0.01
5	1	-0.58	-0.88	0.08	0.13	0.17	-0.08
5	2	-0.41	-0.12	0.42	0.14	0.06	-0.06
5	3	-0.28	0.23	0.53	0.16	0.03	-0.03
5	4	-0.06	0.31	0.44	0.12	-0.00	-0.05
5	5	-0.01	0.27	0.36	0.29	0.16	0.11
		Explained Variation			Cross-Sectional R^2		
All		0.20	0.26	0.53	0.50	0.77	0.92

Out-of-sample explained variation and pricing errors for double sorted portfolios based on Short-Term Reversal (ST_REV) and Momentum (r12_2).

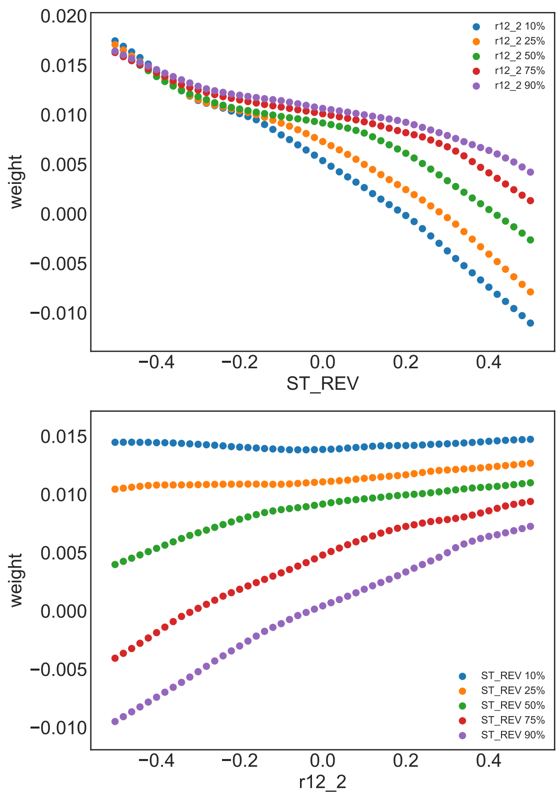
Table A.V Explained Variation and Pricing Errors Double Sorted Portfolios based on Size and Book to Market Ratio

		EN	FFN	GAN	EN	FFN	GAN
LME	BEME	Explained Variation			Alpha		
1	1	0.55	0.47	0.63	-0.01	-0.00	-0.06
1	2	0.66	0.62	0.74	0.01	0.00	-0.04
1	3	0.74	0.70	0.76	0.04	0.01	0.01
1	4	0.77	0.69	0.75	0.01	-0.02	0.01
1	5	0.70	0.66	0.76	-0.01	-0.03	0.02
2	1	0.58	0.20	0.68	0.01	0.11	-0.02
2	2	0.68	0.48	0.81	0.02	0.07	-0.01
2	3	0.82	0.74	0.86	0.04	0.06	0.03
2	4	0.81	0.75	0.85	-0.03	-0.00	-0.01
2	5	0.77	0.79	0.85	-0.04	0.00	0.02
3	1	0.53	0.25	0.73	0.08	0.12	0.02
3	2	0.70	0.59	0.85	0.10	0.11	0.05
3	3	0.86	0.82	0.90	0.06	0.08	0.05
3	4	0.86	0.82	0.88	0.01	0.05	0.02
3	5	0.79	0.76	0.81	-0.04	0.02	0.01
4	1	0.53	0.50	0.79	0.12	0.09	0.01
4	2	0.74	0.78	0.85	0.07	0.04	-0.00
4	3	0.80	0.84	0.83	0.05	0.02	0.00
4	4	0.83	0.81	0.85	0.02	0.03	0.01
4	5	0.73	0.77	0.79	-0.05	-0.02	-0.01
5	1	0.28	0.29	0.44	0.01	-0.09	-0.06
5	2	0.54	0.53	0.58	0.00	-0.08	-0.05
5	3	0.51	0.56	0.57	-0.01	-0.04	-0.04
5	4	0.54	0.60	0.67	-0.01	-0.00	-0.02
5	5	0.37	0.52	0.56	-0.04	-0.03	-0.03
		Explained Variation			Cross-Sectional R^2		
All		0.67	0.60	0.76	0.94	0.91	0.98

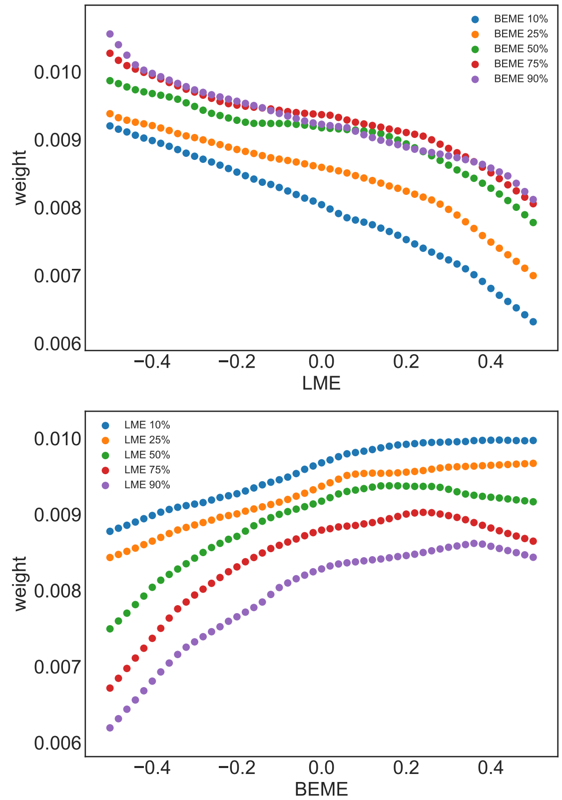
Out-of-sample explained variation and pricing errors for double sorted portfolios based on Size (LME) and Book to Market Ratio (BEME).

Appendix F. SDF Structure

Figure A.7. SDF weight ω as a Function of Covariates for FFN

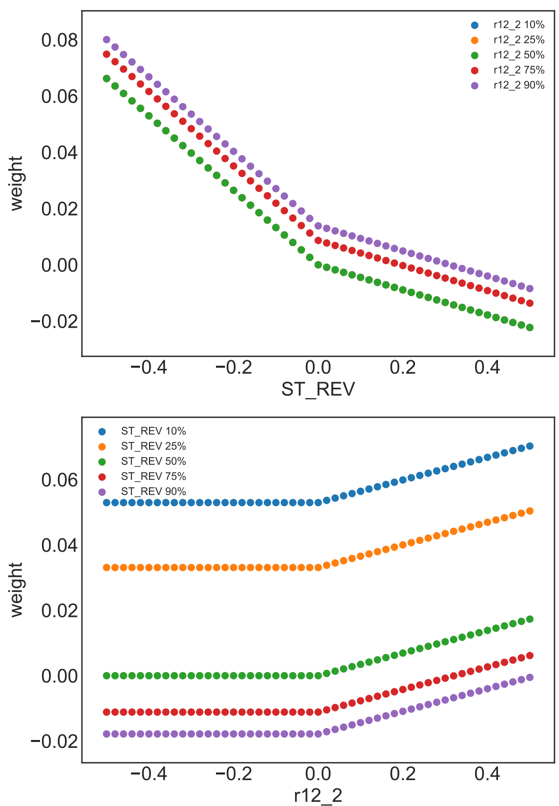


(a) Interaction between Short-Term Reversal (ST_REV) and Momentum (r12_2)

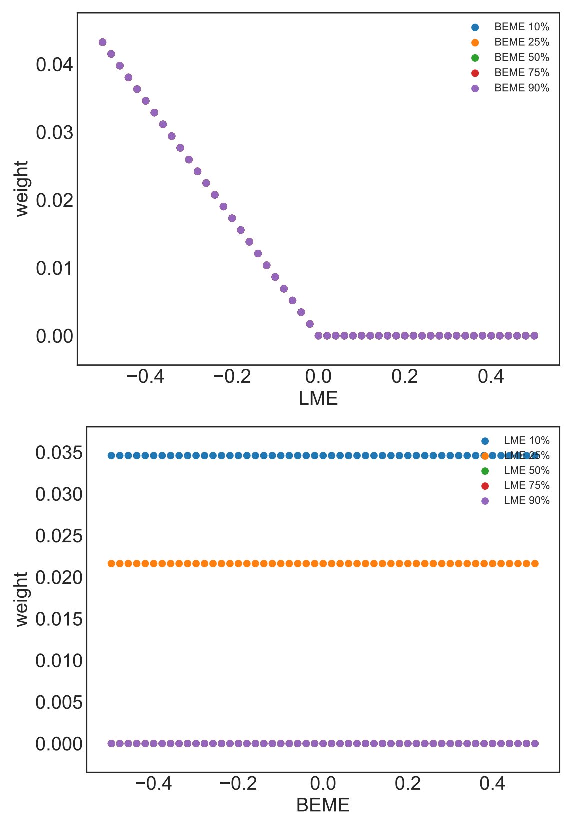


(b) Interaction between Size (LME) and Book to Market Ratio (BEME)

Figure A.8. SDF weight ω as a Function of Covariates for EN

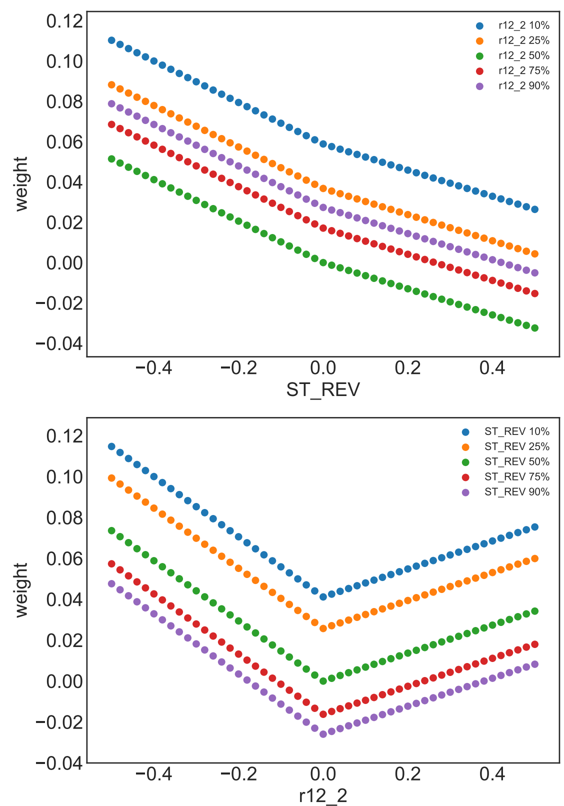


(a) Interaction between Short-Term Reversal (ST_REV) and Momentum (r12_2)

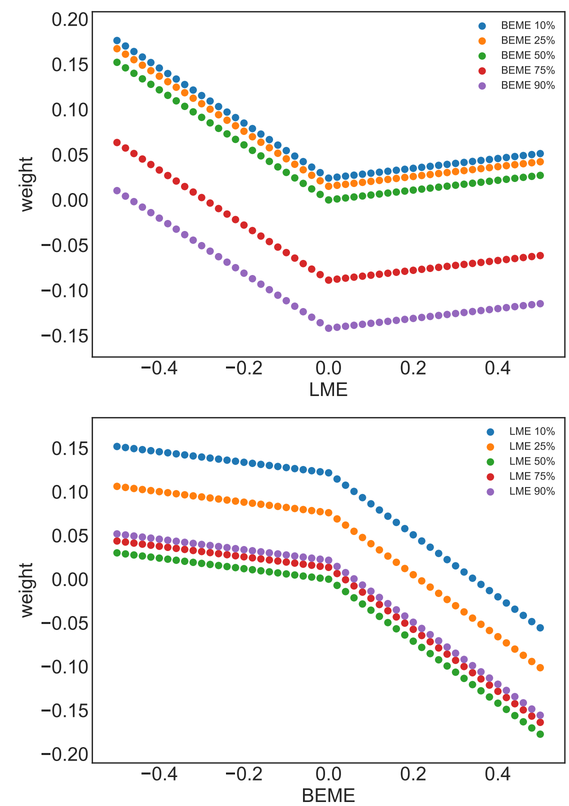


(b) Interaction between Size (LME) and Book to Market Ratio (BEME)

Figure A.9. SDF weight ω as a Function of Covariates for LS



(a) Interaction between Short-Term Reversal (ST_REV) and Momentum (r12_2)



(b) Interaction between Size (LME) and Book to Market Ratio (BEME)

Appendix G. Data

Appendix A. List of Macroeconomic Variables

Variable Name	Description	Source	tCode
RPI	Real Personal Income	Fred-MD	5
W875RX1	Real personal income ex transfer receipts	Fred-MD	5
DPCERA3M086SBEA	Real personal consumption expenditures	Fred-MD	5
CMRMTSPLx	Real Manu. and Trade Industries Sales	Fred-MD	5
RETAILx	Retail and Food Services Sales	Fred-MD	5
INDPRO	IP Index	Fred-MD	5
IPFPNSS	IP: Final Products and Nonindustrial Supplies	Fred-MD	5
IPFINAL	IP: Final Products (Market Group)	Fred-MD	5
IPCONGD	IP: Consumer Goods	Fred-MD	5
IPDCONGD	IP: Durable Consumer Goods	Fred-MD	5
IPNCONGD	IP: Nondurable Consumer Goods	Fred-MD	5
IPBUSEQ	IP: Business Equipment	Fred-MD	5
IPMAT	IP: Materials	Fred-MD	5
IPDMAT	IP: Durable Materials	Fred-MD	5
IPNMAT	IP: Nondurable Materials	Fred-MD	5
IPMANSICS	IP: Manufacturing (SIC)	Fred-MD	5
IPB51222S	IP: Residential Utilities	Fred-MD	5
IPFUELS	IP: Fuels	Fred-MD	5
CUMFNS	Capacity Utilization: Manufacturing	Fred-MD	2
HWI	Help-Wanted Index for United States	Fred-MD	2
HWIURATIO	Ratio of Help Wanted/No. Unemployed	Fred-MD	2
CLF16OV	Civilian Labor Force	Fred-MD	5
CE16OV	Civilian Employment	Fred-MD	5
UNRATE	Civilian Unemployment Rate	Fred-MD	2
UEMPMEAN	Average Duration of Unemployment (Weeks)	Fred-MD	2
UEMPLT5	Civilians Unemployed - Less Than 5 Weeks	Fred-MD	5
UEMP5TO14	Civilians Unemployed for 5-14 Weeks	Fred-MD	5
UEMP15OV	Civilians Unemployed - 15 Weeks & Over	Fred-MD	5
UEMP15T26	Civilians Unemployed for 15-26 Weeks	Fred-MD	5
UEMP27OV	Civilians Unemployed for 27 Weeks and Over	Fred-MD	5
CLAIMSx	Initial Claims	Fred-MD	5
PAYEMS	All Employees: Total nonfarm	Fred-MD	5
USGOOD	All Employees: Goods-Producing Industries	Fred-MD	5
CES1021000001	All Employees: Mining and Logging: Mining	Fred-MD	5
USCONS	All Employees: Construction	Fred-MD	5
MANEMP	All Employees: Manufacturing	Fred-MD	5
DMANEMP	All Employees: Durable goods	Fred-MD	5
NDMANEMP	All Employees: Nondurable goods	Fred-MD	5
SRVPRD	All Employees: Service-Providing Industries	Fred-MD	5
USTPU	All Employees: Trade, Transportation & Utilities	Fred-MD	5
USWTRADE	All Employees: Wholesale Trade	Fred-MD	5
USTRADE	All Employees: Retail Trade	Fred-MD	5
USFIRE	All Employees: Financial Activities	Fred-MD	5
USGOVT	All Employees: Government	Fred-MD	5
CES0600000007	Avg Weekly Hours : Goods-Producing	Fred-MD	1
AWOTMAN	Avg Weekly Overtime Hours : Manufacturing	Fred-MD	2
AWHMAN	Avg Weekly Hours : Manufacturing	Fred-MD	1
HOUST	Housing Starts: Total New Privately Owned	Fred-MD	4
HOUSTNE	Housing Starts, Northeast	Fred-MD	4
HOUSTMW	Housing Starts, Midwest	Fred-MD	4
HOUSTS	Housing Starts, South	Fred-MD	4
HOUSTW	Housing Starts, West	Fred-MD	4
PERMIT	New Private Housing Permits (SAAR)	Fred-MD	4
PERMITNE	New Private Housing Permits, Northeast (SAAR)	Fred-MD	4
PERMITMW	New Private Housing Permits, Midwest (SAAR)	Fred-MD	4
PERMITS	New Private Housing Permits, South (SAAR)	Fred-MD	4
PERMITW	New Private Housing Permits, West (SAAR)	Fred-MD	4
AMDMNOx	New Orders for Durable Goods	Fred-MD	5
AMDMUOx	Unfilled Orders for Durable Goods	Fred-MD	5
BUSINVx	Total Business Inventories	Fred-MD	5
ISRATIOx	Total Business: Inventories to Sales Ratio	Fred-MD	2
M1SL	M1 Money Stock	Fred-MD	6
M2SL	M2 Money Stock	Fred-MD	6
M2REAL	Real M2 Money Stock	Fred-MD	5
AMBSL	St. Louis Adjusted Monetary Base	Fred-MD	6
TOTRESNS	Total Reserves of Depository Institutions	Fred-MD	6
NONBORRES	Reserves Of Depository Institutions	Fred-MD	7
BUSLOANS	Commercial and Industrial Loans	Fred-MD	6
REALLN	Real Estate Loans at All Commercial Banks	Fred-MD	6
NONREVSL	Total Nonrevolving Credit	Fred-MD	6
CONSPI	Nonrevolving consumer credit to Personal Income	Fred-MD	2

Variable Name	Description	Source	tCode
S&P 500	S&P's Common Stock Price Index: Composite	Fred-MD	5
S&P: indust	S&P's Common Stock Price Index: Industrials	Fred-MD	5
S&P div yield	S&P's Composite Common Stock: Dividend Yield	Fred-MD	2
S&P PE ratio	S&P's Composite Common Stock: Price-Earnings Ratio	Fred-MD	5
FEDFUNDS	Effective Federal Funds Rate	Fred-MD	2
CP3Mx	3-Month AA Financial Commercial Paper Rate	Fred-MD	2
TB3MS	3-Month Treasury Bill	Fred-MD	2
TB6MS	6-Month Treasury Bill	Fred-MD	2
GS1	1-Year Treasury Rate	Fred-MD	2
GS5	5-Year Treasury Rate	Fred-MD	2
GS10	10-Year Treasury Rate	Fred-MD	2
AAA	Moody's Seasoned Aaa Corporate Bond Yield	Fred-MD	2
BAA	Moody's Seasoned Baa Corporate Bond Yield	Fred-MD	2
COMPAPFFx	3-Month Commercial Paper Minus FEDFUNDS	Fred-MD	1
TB3SMFFM	3-Month Treasury C Minus FEDFUNDS	Fred-MD	1
TB6SMFFM	6-Month Treasury C Minus FEDFUNDS	Fred-MD	1
T1YFFM	1-Year Treasury C Minus FEDFUNDS	Fred-MD	1
T5YFFM	5-Year Treasury C Minus FEDFUNDS	Fred-MD	1
T10YFFM	10-Year Treasury C Minus FEDFUNDS	Fred-MD	1
AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS	Fred-MD	1
BAAFFM	Moody's Baa Corporate Bond Minus FEDFUNDS	Fred-MD	1
EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	Fred-MD	5
EXJPUSx	Japan / U.S. Foreign Exchange Rate	Fred-MD	5
EXUSUKx	U.S. / U.K. Foreign Exchange Rate	Fred-MD	5
EXCAUSx	Canada / U.S. Foreign Exchange Rate	Fred-MD	5
WPSFD49207	PPI: Finished Goods	Fred-MD	6
WPSFD49502	PPI: Finished Consumer Goods	Fred-MD	6
WPSID61	PPI: Intermediate Materials	Fred-MD	6
WPSID62	PPI: Crude Materials	Fred-MD	6
OILPRICEx	Crude Oil, spliced WTI and Cushing	Fred-MD	6
PPICMM	PPI: Metals and metal products	Fred-MD	6
CPIAUCSL	CPI : All Items	Fred-MD	6
CPIAPPSL	CPI : Apparel	Fred-MD	6
CPITRNSL	CPI : Transportation	Fred-MD	6
CPIMEDSL	CPI : Medical Care	Fred-MD	6
CUSR0000SAC	CPI : Commodities	Fred-MD	6
CUSR0000SAD	CPI : Durables	Fred-MD	6
CUSR0000SAS	CPI : Services	Fred-MD	6
CPIULFSL	CPI : All Items Less Food	Fred-MD	6
CUSR0000SA0L2	CPI : All items less shelter	Fred-MD	6
CUSR0000SA0L5	CPI : All items less medical care	Fred-MD	6
PCEPI	Personal Cons. Expend.: Chain Index	Fred-MD	6
DDURRG3M086SBEA	Personal Cons. Exp: Durable goods	Fred-MD	6
DNDGRG3M086SBEA	Personal Cons. Exp: Nondurable goods	Fred-MD	6
DSERRG3M086SBEA	Personal Cons. Exp: Services	Fred-MD	6
CES0600000008	Avg Hourly Earnings : Goods-Producing	Fred-MD	6
CES2000000008	Avg Hourly Earnings : Construction	Fred-MD	6
CES3000000008	Avg Hourly Earnings : Manufacturing	Fred-MD	6
MZMSL	MZM Money Stock	Fred-MD	6
DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding	Fred-MD	6
DTCTHFNM	Total Consumer Loans and Leases Outstanding	Fred-MD	6
INVEST	Securities in Bank Credit at All Commercial Banks	Fred-MD	6
VXOCLSx	CBOE S&P 100 Volatility Index: VXO	Fred-MD	1
A2ME	Cross sectional Median of A2ME	Calculated from Characteristics	5
AC	Cross sectional Median of AC	Calculated from Characteristics	2
AT	Cross sectional Median of AT	Calculated from Characteristics	6
ATO	Cross sectional Median of ATO	Calculated from Characteristics	5
BEME	Cross sectional Median of BEME	Calculated from Characteristics	5
Beta	Cross sectional Median of Beta	Calculated from Characteristics	1
C	Cross sectional Median of C	Calculated from Characteristics	5
CF	Cross sectional Median of CF	Calculated from Characteristics	2
CF2P	Cross sectional Median of CF2P	Calculated from Characteristics	5
CTO	Cross sectional Median of CTO	Calculated from Characteristics	5
D2A	Cross sectional Median of D2A	Calculated from Characteristics	5
D2P	Cross sectional Median of D2P	Calculated from Characteristics	2
DPI2A	Cross sectional Median of DPI2A	Calculated from Characteristics	5
E2P	Cross sectional Median of E2P	Calculated from Characteristics	5
FC2Y	Cross sectional Median of FC2Y	Calculated from Characteristics	5
IdioVol	Cross sectional Median of IdioVol	Calculated from Characteristics	5
Investment	Cross sectional Median of Investment	Calculated from Characteristics	5
Lev	Cross sectional Median of Lev	Calculated from Characteristics	5
LME	Cross sectional Median of LME	Calculated from Characteristics	6
LT_Rev	Cross sectional Median of LT_Rev	Calculated from Characteristics	2
LTurnover	Cross sectional Median of LTurnover	Calculated from Characteristics	5
MktBeta	Cross sectional Median of MktBeta	Calculated from Characteristics	1
NI	Cross sectional Median of NI	Calculated from Characteristics	1
NOA	Cross sectional Median of NOA	Calculated from Characteristics	5
OA	Cross sectional Median of OA	Calculated from Characteristics	2
OL	Cross sectional Median of OL	Calculated from Characteristics	5
OP	Cross sectional Median of OP	Calculated from Characteristics	5
PCM	Cross sectional Median of PCM	Calculated from Characteristics	5
PM	Cross sectional Median of PM	Calculated from Characteristics	5
PROF	Cross sectional Median of PROF	Calculated from Characteristics	5
Q	Cross sectional Median of Q	Calculated from Characteristics	5

Variable Name	Description	Source	tCode
r2_1	Cross sectional Median of r2_1	Calculated from Characteristics	2
r12_2	Cross sectional Median of r12_2	Calculated from Characteristics	2
r12_7	Cross sectional Median of r12_7	Calculated from Characteristics	2
r36_13	Cross sectional Median of r36_13	Calculated from Characteristics	2
Rel2High	Cross sectional Median of Rel2High	Calculated from Characteristics	5
Resid_Var	Cross sectional Median of Resid_Var	Calculated from Characteristics	5
RNA	Cross sectional Median of RNA	Calculated from Characteristics	5
ROA	Cross sectional Median of ROA	Calculated from Characteristics	5
ROE	Cross sectional Median of ROE	Calculated from Characteristics	5
S2P	Cross sectional Median of S2P	Calculated from Characteristics	5
SGA2S	Cross sectional Median of SGA2S	Calculated from Characteristics	5
Spread	Cross sectional Median of Spread	Calculated from Characteristics	5
ST_REV	Cross sectional Median of ST_REV	Calculated from Characteristics	2
SUV	Cross sectional Median of SUV	Calculated from Characteristics	1
Variance	Cross sectional Median of Variance	Calculated from Characteristics	5
dp	Divident-price ratio	Welch and Goyal (2008)	2
ep	Earnings-price ratio	Welch and Goyal (2008)	2
bm	Book-to-market ratio	Welch and Goyal (2008)	5
ntis	Net equity expansion	Welch and Goyal (2008)	2
tbl	Treasury-bill rate	Welch and Goyal (2008)	2
tms	Term spread	Welch and Goyal (2008)	1
dfy	Default spread	Welch and Goyal (2008)	2
svar	Stock variance	Welch and Goyal (2008)	5

Table A.VI List of Recessions in the United States (1967-2016)

Period Range	Duration	Description
Dec 1969 - Nov 1970	11 months	fiscal tightening, monetary tightening
Nov 1973 - Mar 1975	16 months	oil crisis (1973), stock market crash (1973-1974)
Jan 1980 - July 1980	6 months	monetary tightening
July 1981 - Nov 1982	16 months	energy crisis (1979), monetary tightening
July 1990 - Mar 1991	8 months	oil price shock (1990), debt accumulation, consumer pessimism
Mar 2001 - Nov 2001	8 months	dot-com bubble, 9/11 attacks
Dec 2007 - June 2009	18 months	subprime mortgage crisis

Description of NBER Recessions.

Appendix B. List of Firm-Specific Character Variables

Acronym	Name	Definition	Reference
A2ME	Assets to market cap	Total assets (AT) over market capitalization (PRC*SHROUT) as of December t-1	Bhandari (1988)
AC	Accrual	Change in operating working capital per split-adjusted share from the fiscal year end t-2 to t-1 divided by book equity (defined in BEME) per share in t-1. Operating working capital per split-adjusted share is defined as current assets (ACT) minus cash and short-term investments (CHE) minus current liabilities (LCT) minus debt in current liabilities (DLC) minus income taxes payable (TXP).	Sloan (1996)
AT	Total Assets	Total Assets (AT)	Gandhi and Lustig (2015)
ATO	Net sales over lagged net operating assets	Net sales (SALE) over lagged net operating assets. Net operating assets are the difference between operating assets and operating liabilities (defined in NOA)	Soliman (2008)
BEME	Book to Market Ratio	Book equity is shareholder equity (SH) plus deferred taxes and investment tax credit (TXDITC), minus preferred stock (PS). SH is shareholders equity (SEQ). If missing, SH is the sum of common equity (CEQ) and preferred stock (PS). If missing, SH is the difference between total assets (AT) and total liabilities (LT). Depending on availability, we use the redemption (item PSTKRV), liquidating (item PSTKL), or par value (item PSTK) for PS. The market value of equity (PRC*SHROUT) is as of December t-1.	Fama and French (1992)
Beta	CAPM Beta	Product of correlations between the excess return of stock i and the market excess return and the ratio of volatilities. We calculate volatilities from the standard deviations of daily log excess returns over a one-year horizon requiring at least 120 observations. We estimate correlations using overlapping three-day log excess returns over a five-year period requiring at least 750 non-missing observations.	Frazzini and Pedersen (2014)
C	Ratio of cash and short-term investments to total assets	Ratio of cash and short-term investments (CHE) to total assets (AT)	Palazzo (2012)
CF	Free Cash Flow to Book Value	Cash flow to book value of equity is the ratio of net income (NI), depreciation and amortization (DP), less change in working capital (WCAPCH), and capital expenditure (CAPX) over the book-value of equity (defined in BEME)	Hou et al. (2011)
CF2P	Cashflow to price	Cashflow over market capitalization (PRC*SHROUT) as of December t-1. Cashflow is defined as income before extraordinary items (IB) plus depreciation and amortization (DP) plus deferred taxes (TXDB).	Desai et al. (2004)
CTO	Capital turnover	Ratio of net sales (SALE) to lagged total assets (AT)	Haugen and Baker (1996)
D2A	Capital intensity	Ratio of depreciation and amortization (DP) to total assets (AT)	Gorodnichenko and Weber (2016)
D2P	Dividend Yield	Total dividends (DIVAMT) paid from July of t-1 to June of t per dollar of equity (LME) in June of t	Litzenberger and Ramaswamy (1979)
DPI2A	Change in property, plants, and equipment	Changes in property, plants, and equipment (PPEGT) and inventory (INVT) over lagged total assets (TA)	Lyandres et al. (2008)
E2P	Earnings to price	The earnings used in June of year t are total earnings before extraordinary items for the last fiscal year end in t-1. P (actually ME) is price times shares outstanding at the end of December of t-1.	Basu (1983)
FC2Y	Fixed costs to sales	Ratio of selling, general, and administrative expenses (XSGS), research and development expenses (XRD), and advertising expenses (XAD) to net sales (SALE)	D'Acunto et al. (2017)
IdioVol	Idiosyncratic volatility	Standard deviation of the residuals from aregression of excess returns on the Fama and French three-factor model	Ang et al. (2004)
Investment	Investment	Change in total assets (AT) from the fiscal year ending in year t-2 to the fiscal year ending in t-1, divided by t-2 total assets	Cooper et al. (2008)
Lev	Leverage	Ratio of long-term debt (DLTT) and debt in current liabilities (DLC) to the sum of long-term debt, debt in current liabilities, and stockholders' equity (SEQ)	Lewellen (2015)
LME	Size	Total market capitalization at the end of the previous month defined as price times shares outstanding	Fama and French (1992)
LT_Rev	Long-term reversal	Cumulative return from 60 months before the return prediction to 13 months before	Jegadeesh and Titman (2001)

continue on the next page

Acronym	Name	Definition	Reference
Lturnover	Turnover	Turnover is last month's volume (VOL) over shares outstanding (SHROUT)	Datar et al. (1998)
MktBeta	Market Beta	Coefficient of the market excess return from the regression on excess returns in the past 60 months (24 months minimum)	Fama and MacBeth (1973)
NI	Net Share Issues	The change in the natural log of split-adjusted shares outstanding (CSHO*AJEX) from the fiscal yearend in t-2 to the fiscal yearend in t-1	Pontiff and Woodgate (2008)
NOA	Net operating assets	Difference between operating assets minus operating liabilities scaled by lagged total assets (AT). Operating assets are total assets (AT) minus cash and short-term investments (CHE), minus investment and other advances (IVAO). Operating liabilities are total assets (AT), minus debt in current liabilities (DLC), minus long-term debt (DLTT), minus minority interest (MIB), minus preferred stock (PSTK), minus common equity (CEQ).	Hirshleifer et al. (2004)
OA	Operating accruals	Changes in non-cash working capital minus depreciation (DP) scaled by lagged total assets (TA). Non-cash working capital is defined in Accrual (AC)	Sloan (1996)
OL	Operating leverage	Sum of cost of goods sold (COGS) and selling, general, and administrative expenses (XSGA) over total assets (AT)	Novy-Marx (2011)
OP	Operating profitability	Annual revenues (REVT) minus cost of goods sold (COGS), interest expense (TIE), and selling, general, and administrative expenses (XSGA) divided by book equity (defined in BEME)	Fama and French (2015)
PCM	Price to cost margin	Difference between net sales (SALE) and costs of goods sold (COGS) divided by net sales (SALE)	Bustamante and Donangelo (2017)
PM	Profit margin	Operating income after depreciation (OIADP) over net sales (SALE)	Soliman (2008)
PROF	Profitability	Gross profitability (GP) divided by the book value of equity (defined in BEME)	Ball et al. (2015)
Q	Tobin's Q	Tobin's Q is total assets (AT), the market value of equity (SHROUT times PRC)minus cash and short-term investments (CEQ), minus deferred taxes (TXDB) scaled by total assets (AT)	Kaldor (1966)
r2_1	Short-term momentum	Lagged one-month return	Jegadeesh and Titman (1993)
r12_2	Momentum	To be included in a portfolio for month t (formed at the end of month t-1), a stock must have a price for the end of month t-13 and a good return for t-2. In addition, any missing returns from t-12 to t-3 must be -99.0, CRSP's code for a missing price. Each included stock also must have ME for the end of month t-1.	Fama and FrencH (1996)
r12_7	Intermediate momentum	Cumulative return from 12 months before the return prediction to seven months before	Novy-Marx (2012)
r36_13	Long-term momentum	Cumulative return from 36 months before the return prediction to 13 months before	Bondt and Thaler (1985)
Rel2High	Closeness to past year high	The ratio of stock price at the end of the previous calendar month and the highest daily price in the past year	George and Hwang (2004)
Resid_Var	Residual Variance	Variance of the residuals from a regression of excess returns in the past two months on the Fama and French three-factor model	Ang et al. (2004)
RNA	Return on net operating assets	Ratio of operating income after depreciation (OIADP) to lagged net operating assets. Net operating assets are the difference between operating assets minus operating liabilities. (defined in NOA)	Soliman (2008)
ROA	Return on assets	Income before extraordinary items (IB) to lagged total assets (AT)	Balakrishnan et al. (2010)
ROE	Return on equity	Income before extraordinary items (IB) to lagged book-value of equity (defined in BEME)	Haugen and Baker (1996)
S2P	Sales to price	Ratio of net sales (SALE) to the market capitalization (LME)	Lewellen (2015)
SGA2S	Selling, general and administrative expenses to sales	Ratio of selling, general and administrative expenses (XSGA) to net sales (SALE)	Freyberger et al. (2017)
Spread	Bid-ask spread	The average daily bid-ask spread in the previous month	Chung and Zhang (2014)
ST_Rev	Short-term reversal	Prior month return	Jegadeesh and Titman (1993)

continue on the next page

Acronym	Name	Definition	Reference
SUV	Standard unexplained volume	Difference between actual volume and predicted volume in the previous month. Predicted volume comes from a regression of daily volume on a constant and the absolute values of positive and negative returns. Unexplained volume is standardized by the standard deviation of the residuals from the regression	Garfinkel (2009)
Variance	Variance	Variance of daily returns in the past two months	Ang et al. (2004)

Note: Most Characteristic Variables in this table are summarized by Freyberger et al. (2017). We augment this list by adding extra variables listed on the Kenneth French Data Library.

Table A.VIII Firm Characteristics by Category

	Past Returns			**Value**	
(1)	r2_1	Short-term momentum	(26)	A2ME	Assets to market cap
(2)	r12_2	Momentum	(27)	BEME	Book to Market Ratio
(3)	r12_7	Intermediate momentum	(28)	C	Ratio of cash and short-term investments to total assets
(4)	r36_13	Long-term momentum	(29)	CF	Free Cash Flow to Book Value
(5)	ST_Rev	Short-term reversal	(30)	CF2P	Cashflow to price
(6)	LT_Rev	Long-term reversal	(31)	D2P	Dividend Yield
			(32)	E2P	Earnings to price
	Investment		(33)	Q	Tobin's Q
(7)	Investment	Investment	(34)	S2P	Sales to price
(8)	NOA	Net operating assets	(35)	Lev	Leverage
(9)	DPI2A	Change in property, plants, and equipment			
(10)	NI	Net Share Issues		**Trading Frictions**	
			(36)	AT	Total Assets
	Profitability		(37)	Beta	CAPM Beta
(11)	PROF	Profitability	(38)	IdioVol	Idiosyncratic volatility
(12)	ATO	Net sales over lagged net operating assets	(39)	LME	Size
(13)	CTO	Capital turnover	(40)	LTurnover	Turnover
(14)	FC2Y	Fixed costs to sales	(41)	MktBeta	Market Beta
(15)	OP	Operating profitability	(42)	Rel2High	Closeness to past year high
(16)	PM	Profit margin	(43)	Resid_Var	Residual Variance
(17)	RNA	Return on net operating assets	(44)	Spread	Bid-ask spread
(18)	ROA	Return on assets	(45)	SUV	Standard unexplained volume
(19)	ROE	Return on equity	(46)	Variance	Variance
(20)	SGA2S	Selling, general and administrative expenses to sales			
(21)	D2A	Capital intensity			
	Intangibles				
(22)	AC	Accrual			
(23)	OA	Operating accruals			
(24)	OL	Operating leverage			
(25)	PCM	Price to cost margin			

Internet Appendix for
"Deep Learning in Asset Pricing"

Luyang Chen, Markus Pelger and Jason Zhu[1]

Abstract

This Internet Appendix provides additional tables and figures supporting the main text. Among others it includes the SDF structure as a function of additional characteristics and pricing results for additional portfolio sorts.

IA.A. Overview

The Internet Appendix collects multiple robustness results that support the results in the main text. Section IA.B shows that the results are robust to the normalization of the SDF and the model performance for large cap stocks. Section IA.C collects the predictive performance results for the FFN, EN and LS model. In Section IA.D we provide the asset pricing results for additional characteristic sorted portfolios. Section IA.E shows the functional form of different SDFs conditioned on univariate or multiple characteristics.

[1]Citation format: Chen, Luyang, Pelger, Markus and Zhu, Jason, Internet Appendix for "Deep Learning in Asset Pricing".

IA.B. Robustness Results

Table IA.I SDF Factor Portfolio Performance without Normalized Weights

	SR			Max Loss			Max Drawdown		
Model	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
LS	1.82	0.57	0.41	-1.54	-1.96	-5.13	1	3	4
EN	1.28	1.13	0.47	-2.20	-1.82	-5.77	1	3	5
FFN	0.48	0.42	0.47	-4.60	-4.72	-3.28	6	3	5
GAN	3.21	1.45	0.72	0.18	-0.27	-5.95	0	1	5

Sharpe Ratio, maximum 1-month loss and maximum drawdown of the SDF factor portfolios.

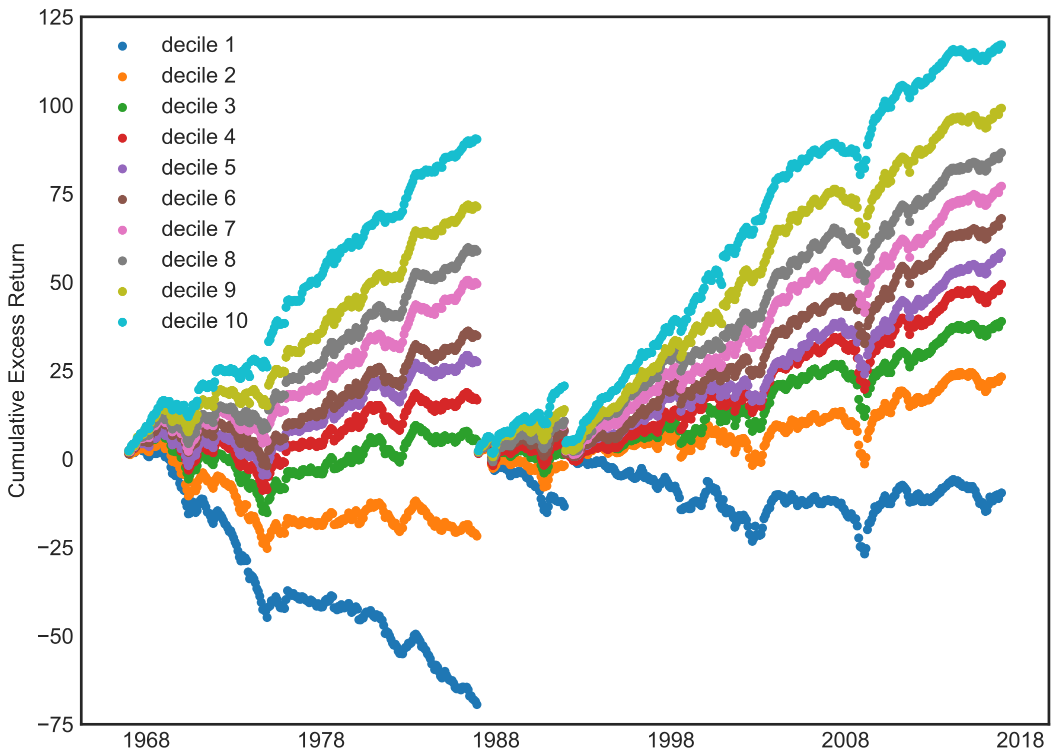
Table IA.II SDF Factor Risk Measures for Large Market Cap Stocks

	SR			Max Loss			Max Drawdown		
Model	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
Size $\geq$ 0.001% of total market cap									
LS	1.44	0.31	0.13	-3.07	-2.19	-4.59	1	3	7
EN	0.93	0.56	0.15	-3.00	-2.45	-4.82	2	3	5
FFN	0.42	0.20	0.30	-3.89	-4.66	-4.33	6	4	5
GAN	2.32	1.09	0.41	-1.17	-1.14	-4.84	1	1	5
Size $\geq$ 0.01% of total market cap									
LS	0.32	-0.11	-0.06	-3.11	-1.82	-3.67	4	5	7
EN	0.37	0.26	0.23	-4.44	-2.67	-4.66	4	3	7
FFN	0.32	0.17	0.24	-3.30	-4.53	-5.08	7	5	5
GAN	0.97	0.54	0.26	-6.91	-1.36	-5.01	2	2	7

Sharpe Ratio, maximum 1-month loss and maximum drawdown of the SDF factor portfolios. The model is evaluated on stocks with market capitalization larger than 0.01% or 0.001% of the total market capitalization.

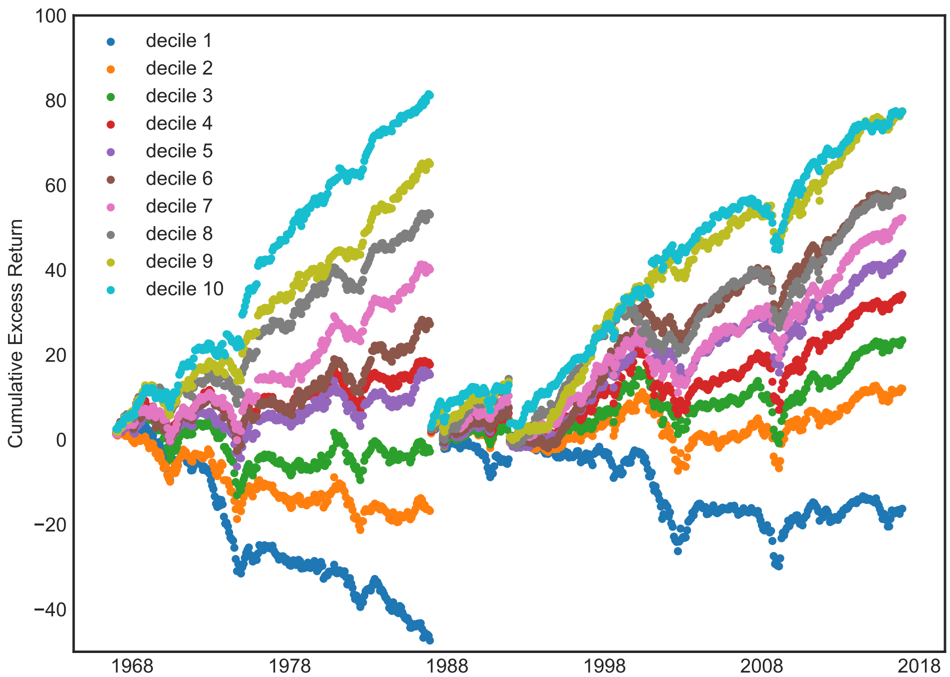
IA.C. Predictive Portfolios

Figure IA.1. Equally Weighted β Sorted Decile Portfolios with FFN



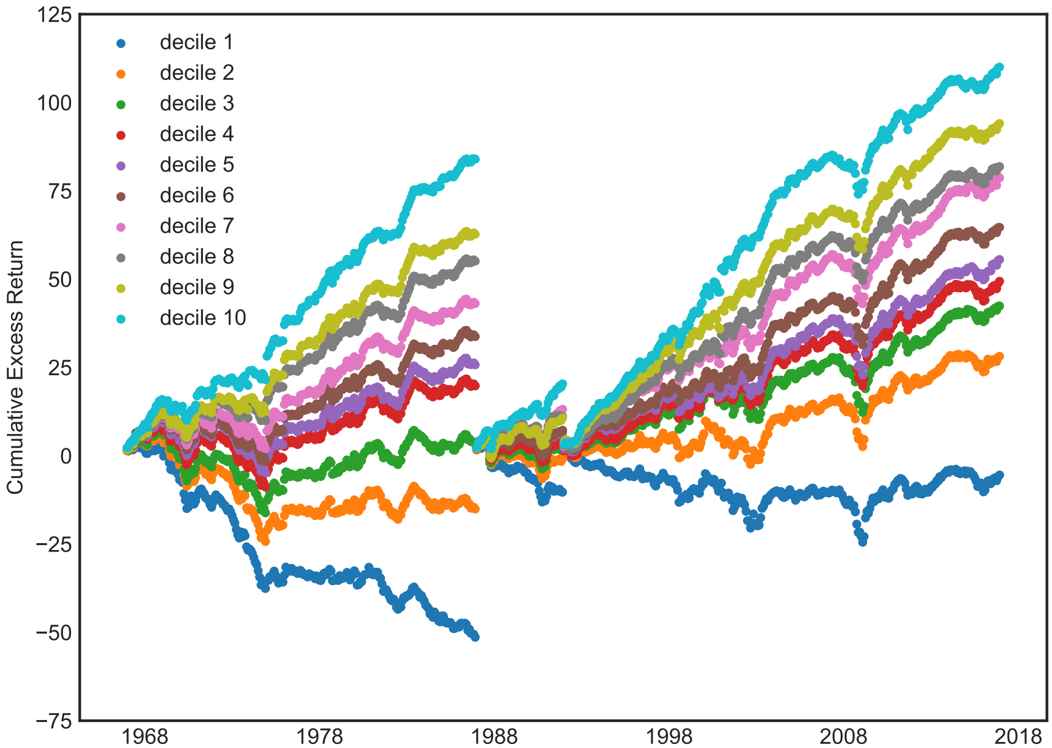
Cumulative excess return of decile sorted portfolios based on the risk loadings β for FFN. The first portfolio is based on the smallest decile of risk loadings, while the last decile portfolio is constructed with the largest loading decile. Within each decile the stocks are equally weighted.

Figure IA.2. Value Weighted β Sorted Decile Portfolios with FFN



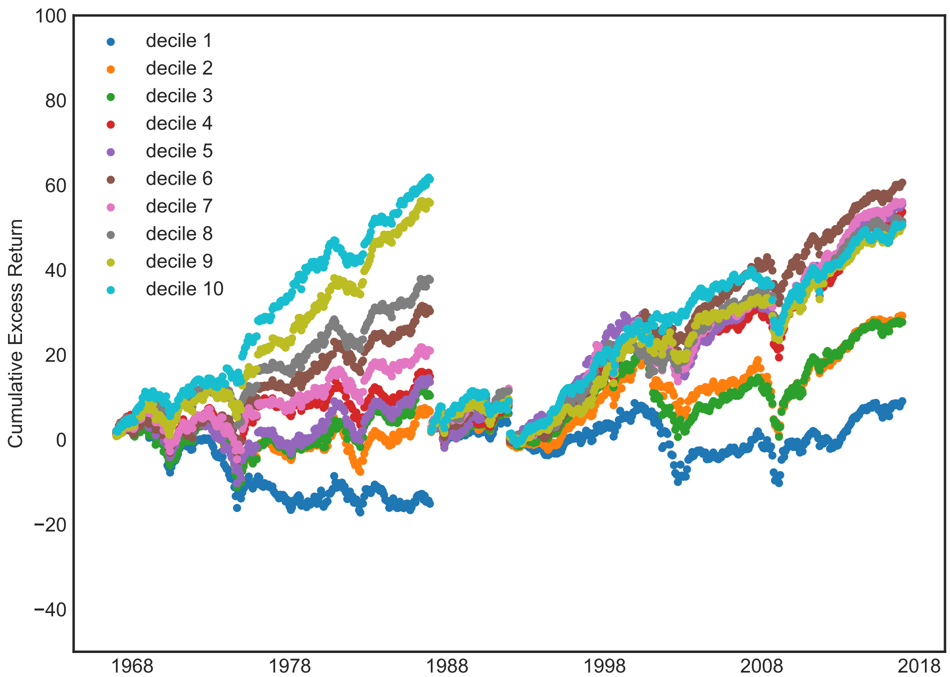
Cumulative excess return of decile sorted portfolios based on the risk loadings β for FFN. The first portfolio is based on the smallest decile of risk loadings, while the last decile portfolio is constructed with the largest loading decile. Within each decile the stocks are value weighted.

Figure IA.3. Equally Weighted β Sorted Decile Portfolios with EN



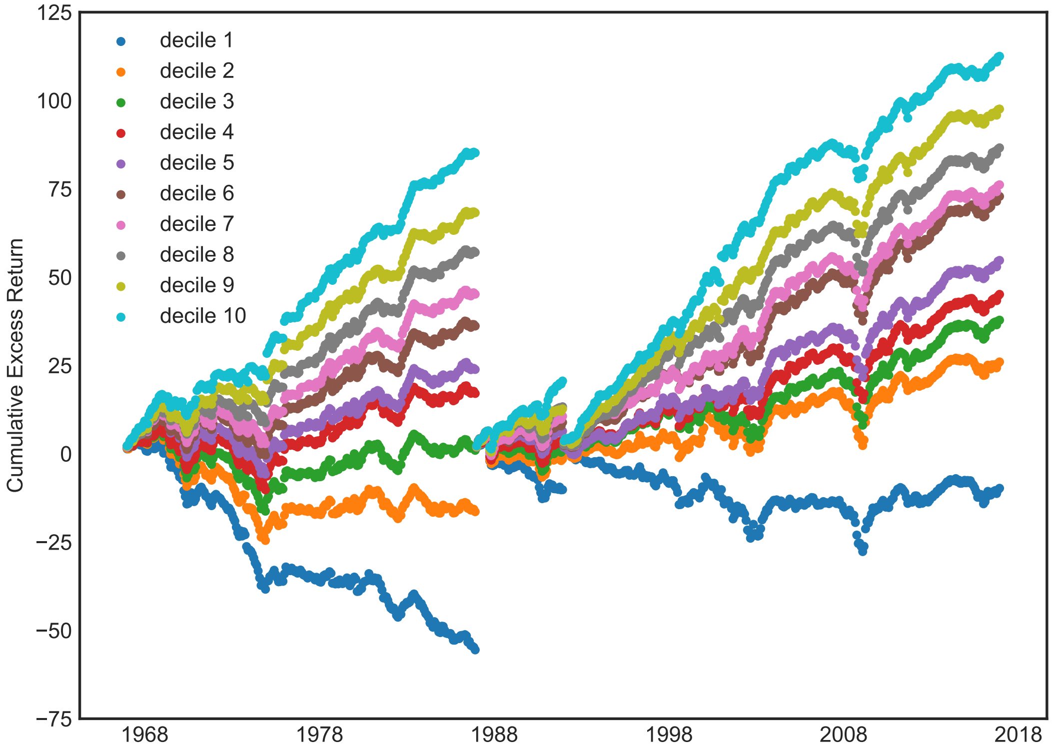
Cumulative excess return of decile sorted portfolios based on the risk loadings β for EN. The first portfolio is based on the smallest decile of risk loadings, while the last decile portfolio is constructed with the largest loading decile. Within each decile the stocks are equally weighted.

Figure IA.4. Value Weighted β Sorted Decile Portfolios with EN



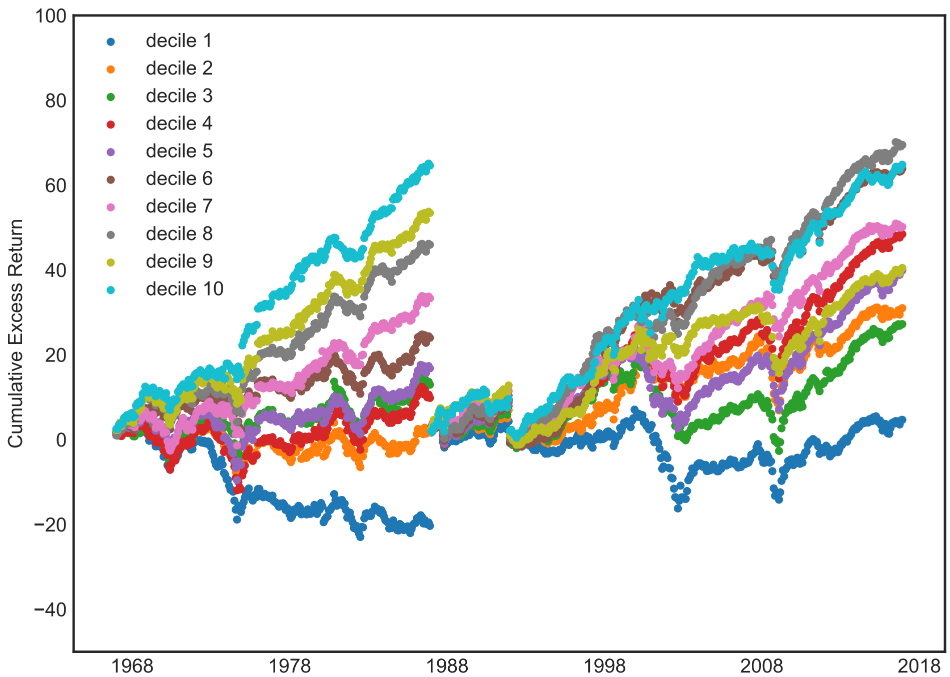
Cumulative excess return of decile sorted portfolios based on the risk loadings β for EN. The first portfolio is based on the smallest decile of risk loadings, while the last decile portfolio is constructed with the largest loading decile. Within each decile the stocks are value weighted.

Figure IA.5. Equally Weighted β Sorted Decile Portfolios with LS



Cumulative excess return of decile sorted portfolios based on the risk loadings β for LS. The first portfolio is based on the smallest decile of risk loadings, while the last decile portfolio is constructed with the largest loading decile. Within each decile the stocks are equally weighted.

Figure IA.6. Value Weighted β Sorted Decile Portfolios with LS



Cumulative excess return of decile sorted portfolios based on the risk loadings β for LS. The first portfolio is based on the smallest decile of risk loadings, while the last decile portfolio is constructed with the largest loading decile. Within each decile the stocks are value weighted.

IA.D. Portfolio Pricing

Table IA.III Explained Variation and Pricing Errors for Size and Dividend Yield Sorted Portfolios

		EN	FFN	GAN	EN	FFN	GAN
LME	D2P	Explained Variation			Alpha		
1	1	0.82	0.78	0.83	-0.01	-0.01	-0.02
1	2	0.79	0.72	0.78	0.01	0.01	-0.01
1	3	0.74	0.71	0.77	0.04	0.02	-0.00
1	4	0.29	0.30	0.31	0.09	0.04	0.07
1	5	0.21	0.13	0.44	-0.10	-0.11	-0.04
2	1	0.82	0.51	0.83	-0.03	0.06	-0.01
2	2	0.81	0.56	0.85	0.01	0.08	0.01
2	3	0.72	0.54	0.78	-0.01	0.05	-0.01
2	4	0.61	0.52	0.60	-0.04	-0.03	-0.07
2	5	0.51	0.58	0.67	-0.07	-0.06	-0.03
3	1	0.73	0.46	0.81	0.09	0.15	0.06
3	2	0.76	0.54	0.84	0.04	0.11	0.02
3	3	0.70	0.51	0.83	0.09	0.15	0.06
3	4	0.77	0.69	0.83	0.05	0.07	0.02
3	5	0.67	0.70	0.70	-0.05	-0.04	-0.03
4	1	0.62	0.47	0.80	0.12	0.14	0.04
4	2	0.67	0.58	0.83	0.08	0.09	0.01
4	3	0.59	0.52	0.79	0.10	0.10	0.02
4	4	0.77	0.78	0.78	0.03	-0.00	-0.02
4	5	0.56	0.54	0.54	-0.07	-0.09	-0.07
5	1	0.15	0.35	0.53	0.11	0.07	0.01
5	2	0.23	0.39	0.60	0.09	0.05	0.00
5	3	0.23	0.21	0.51	0.03	-0.05	-0.06
5	4	0.40	0.22	0.36	-0.03	-0.09	-0.06
5	5	0.36	0.38	0.43	-0.00	-0.06	-0.03
		Explained Variation			Cross-Sectional R^2		
All		0.58	0.50	0.67	0.89	0.84	0.96

Out-of-sample explained variation and pricing errors for double sorted portfolios based on Size (LME) and Dividend Yield (D2P).

Table IA.IV Explained Variation and Pricing Errors for Standard Unexplained Volume Sorted Portfolios

SUV	EN	FFN	GAN	EN	FFN	GAN
Decile	Explained Variation			Alpha		
1	-0.22	0.50	0.78	0.28	0.00	-0.06
2	-0.03	0.64	0.82	0.33	0.10	0.03
3	0.11	0.69	0.80	0.26	0.06	0.02
4	0.28	0.71	0.80	0.21	0.03	-0.01
5	0.49	0.79	0.83	0.16	0.02	0.01
6	0.58	0.84	0.87	0.10	-0.04	-0.04
7	0.72	0.84	0.86	0.11	0.00	0.03
8	0.78	0.82	0.85	0.03	-0.01	0.01
9	0.76	0.78	0.83	-0.03	-0.09	-0.02
10	0.76	0.83	0.85	-0.13	-0.06	-0.00
	Explained Variation			Cross-Sectional R^2		
All	0.42	0.75	0.83	0.64	0.97	0.99

Out-of-sample explained variation and pricing errors for decile sorted portfolios based on Standard Unexplained Volume (SUV).

Table IA.V Explained Variation and Pricing Errors for Net Operating Assets Sorted Portfolios

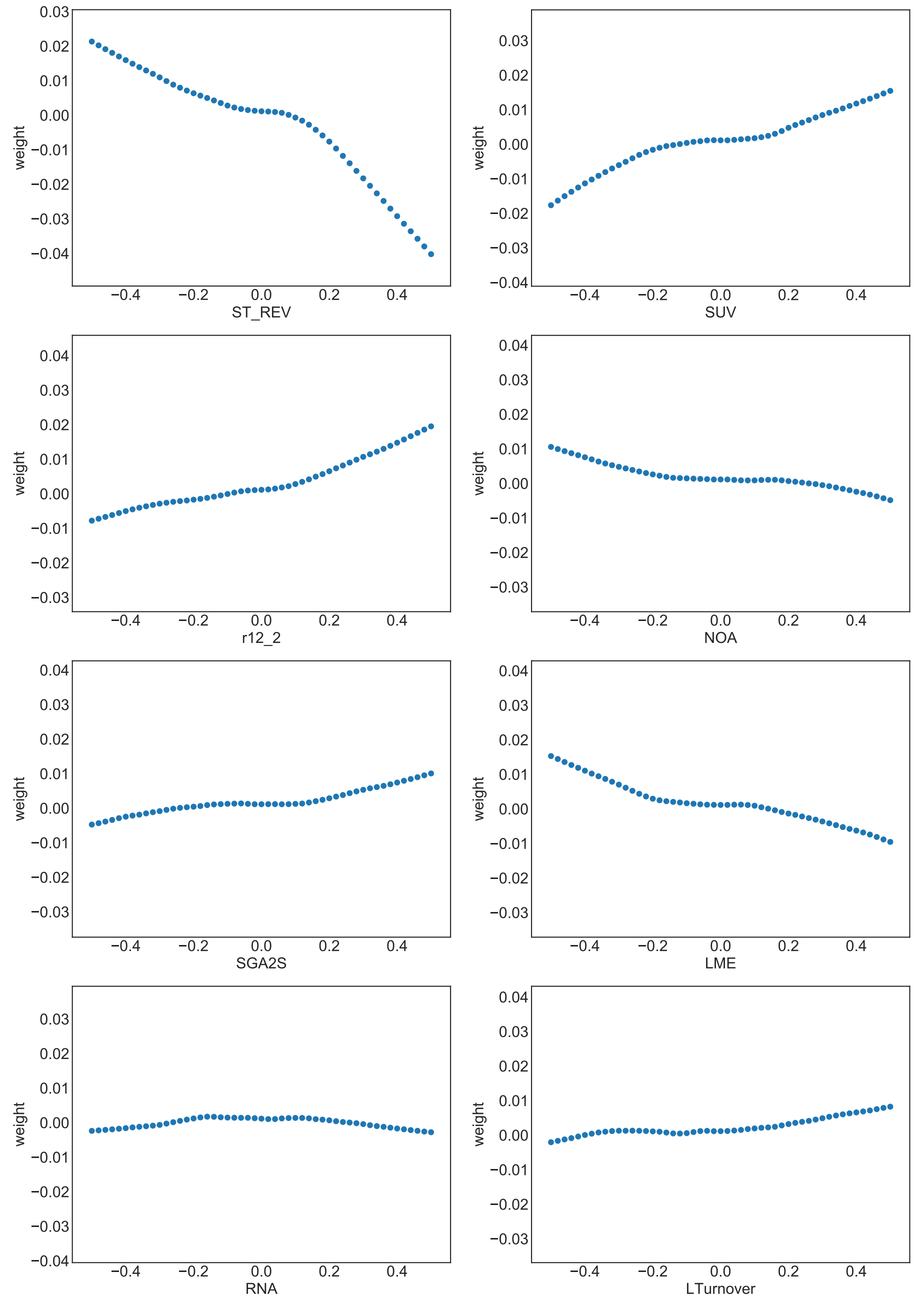
NOA	EN	FFN	GAN	EN	FFN	GAN
Decile	Explained Variation			Alpha		
1	0.41	0.55	0.66	0.17	0.10	0.09
2	0.57	0.72	0.80	0.05	-0.01	0.04
3	0.58	0.79	0.84	-0.06	-0.07	-0.03
4	0.69	0.76	0.78	0.02	0.01	0.05
5	0.73	0.75	0.77	-0.03	-0.04	0.00
6	0.64	0.75	0.75	0.06	0.03	0.05
7	0.72	0.82	0.83	0.02	-0.01	-0.00
8	0.67	0.75	0.84	-0.08	-0.12	-0.13
9	0.66	0.79	0.85	0.10	0.07	0.02
10	0.43	0.47	0.75	-0.04	-0.06	-0.15
	Explained Variation			Cross-Sectional R^2		
All	0.58	0.69	0.78	0.94	0.96	0.95

Out-of-sample explained variation and pricing errors for decile sorted portfolios based on Net Operating Assets (NOA).

IA.E. SDF Structure

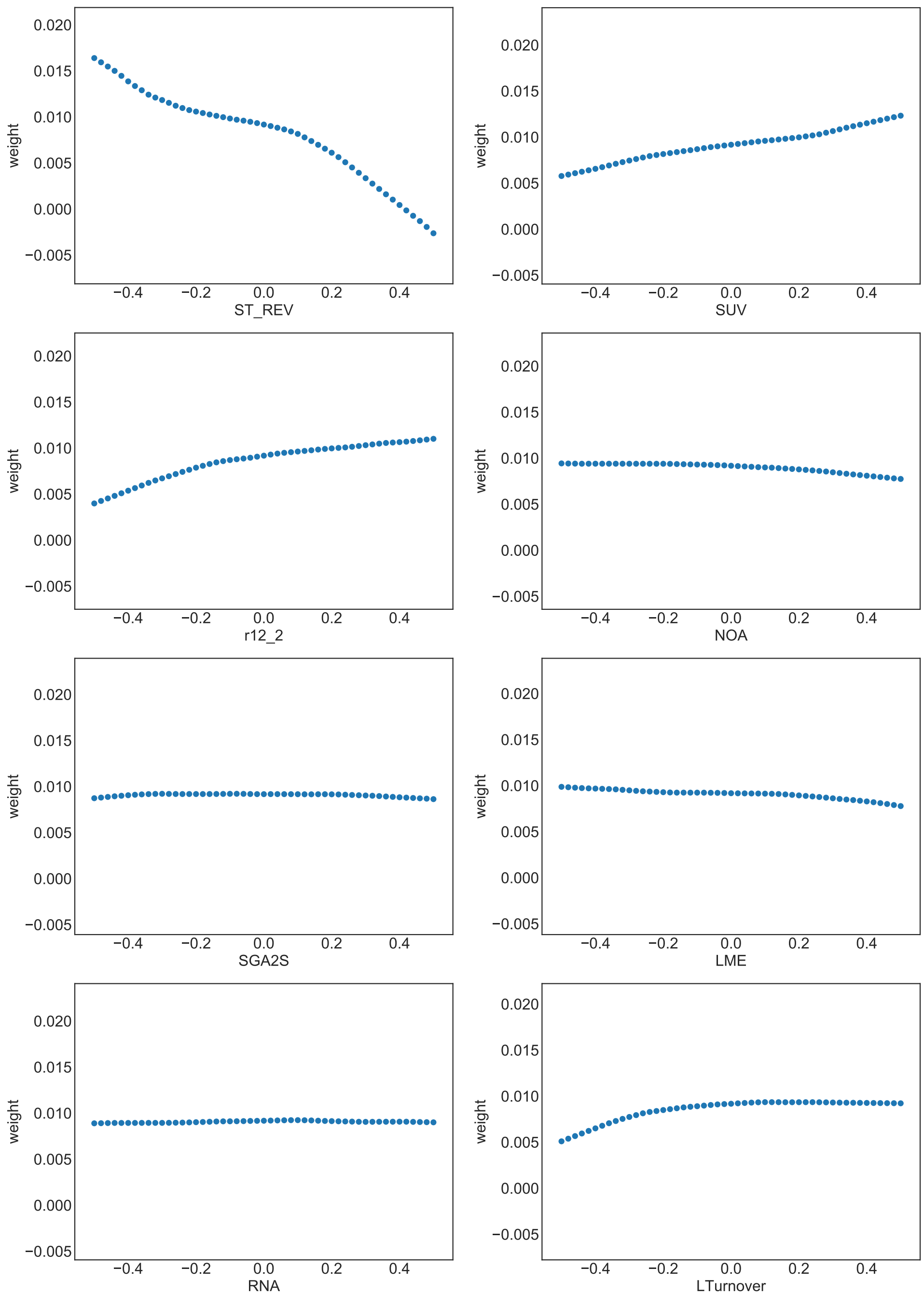
IA.E.1. One-Dimensional Relationship

Figure IA.7. SDF weight ω as a Function of Covariates for GAN



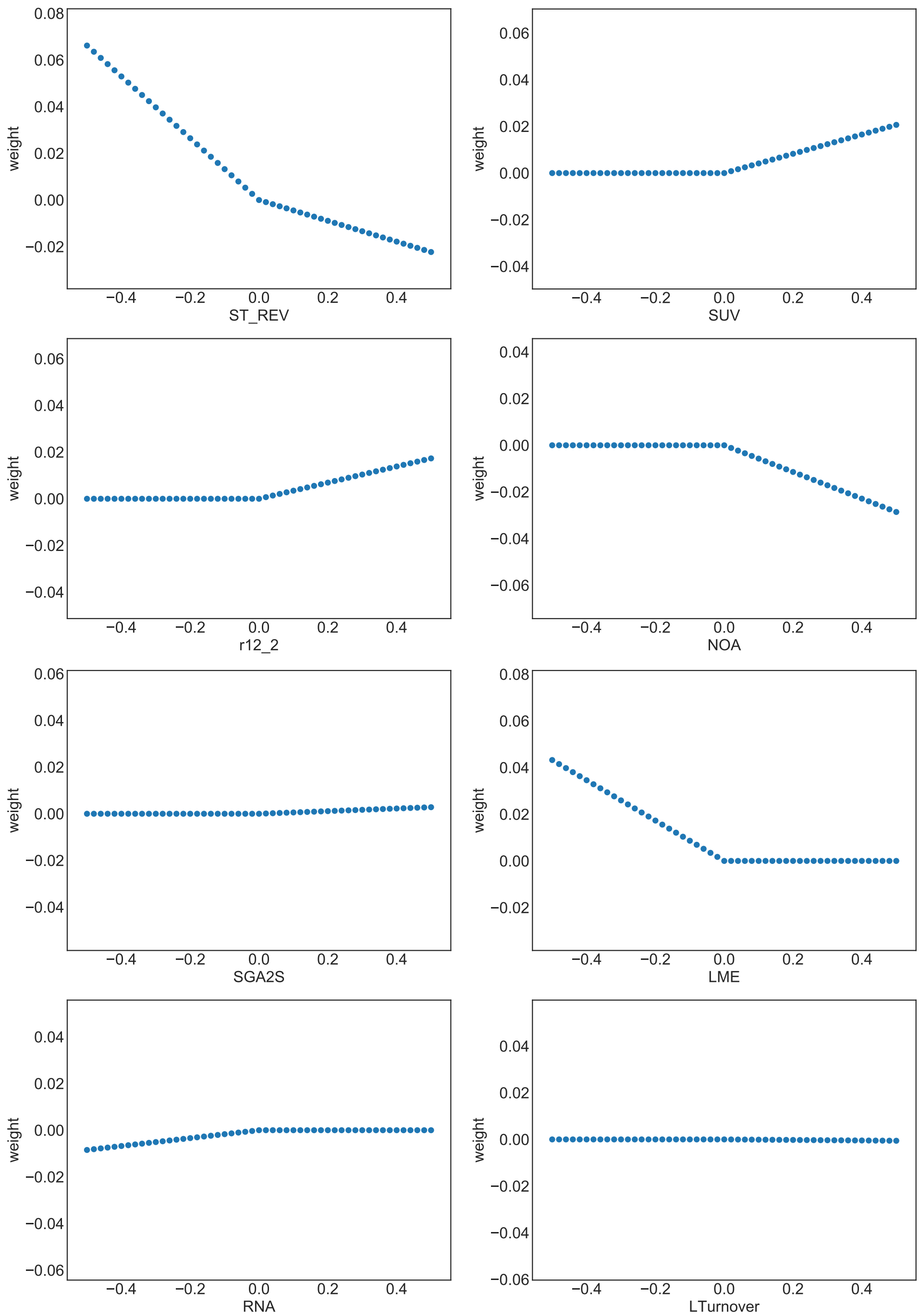
SDF weight ω as a one-dimensional function of covariates keeping the other covariates at their mean level. The covariates are Short-Term Reversal (ST_REV), Standard Unexplained Volume (SUV), Momentum (r12_2), Net Operating Assets (NOA), Selling, General and Administrative Expenses to Sales (SGA2S), Size (LME), Return on Net Operating Assets (RNA) and Turnover (LTurnover).

Figure IA.8. SDF weight ω as a Function of Covariates for FFN



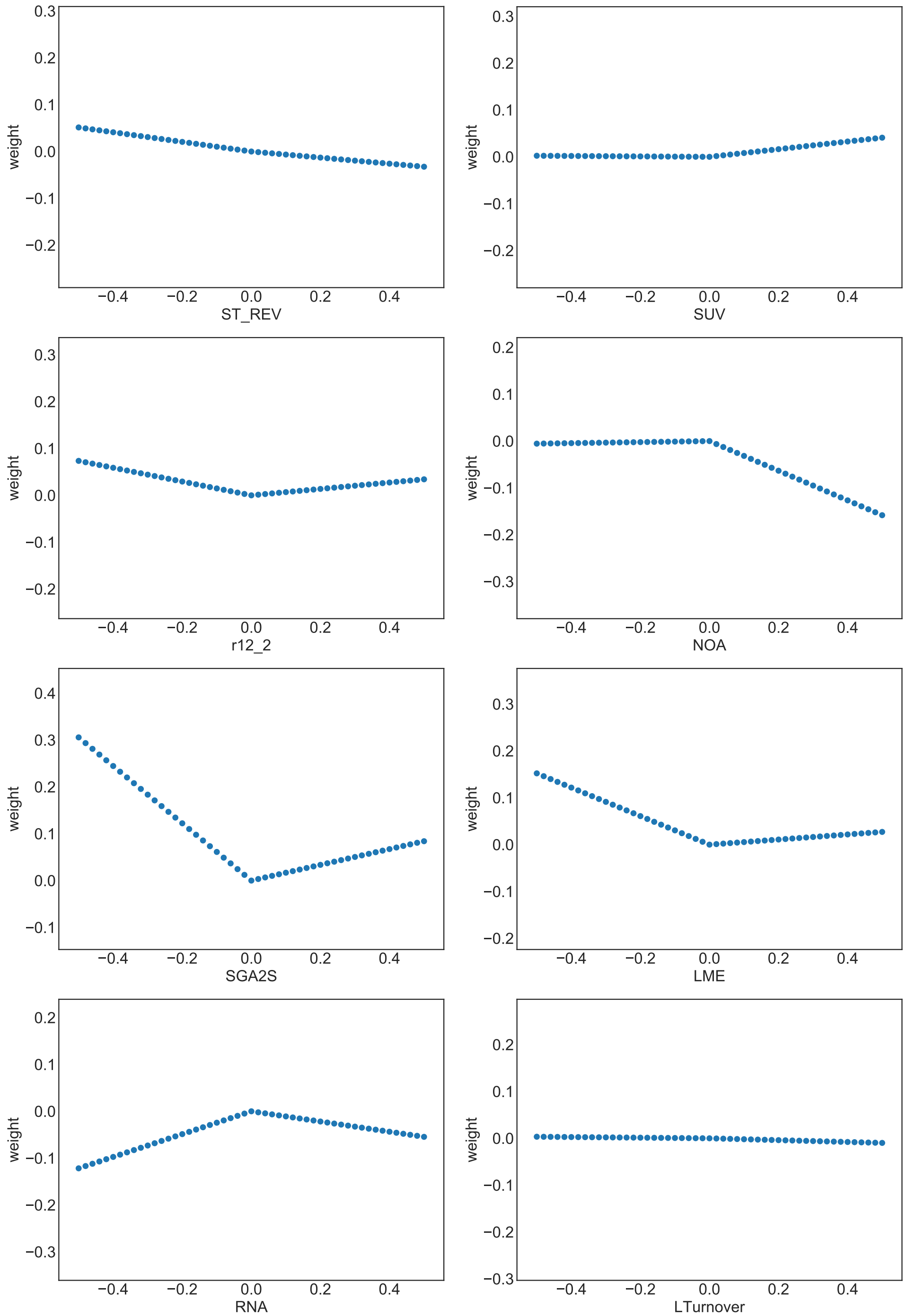
SDF weight ω as a one-dimensional function of covariates keeping the other covariates at their mean level. The covariates are Short-Term Reversal (ST_REV), Standard Unexplained Volume (SUV), Momentum (r12_2), Net Operating Assets (NOA), Selling, General and Administrative Expenses to Sales (SGA2S), Size (LME), Return on Net Operating Assets (RNA) and Turnover (LTurnover).

Figure IA.9. SDF weight ω as a Function of Covariates for EN



SDF weight ω as a one-dimensional function of covariates keeping the other covariates at their mean level. The covariates are Short-Term Reversal (ST_REV), Standard Unexplained Volume (SUV), Momentum (r12_2), Net Operating Assets (NOA), Selling, General and Administrative Expenses to Sales (SGA2S), Size (LME), Return on Net Operating Assets (RNA) and Turnover (LTurnover).

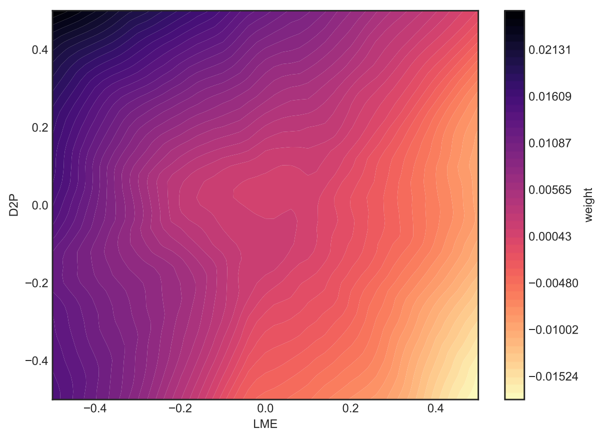
Figure IA.10. SDF weight ω as a Function of Covariates for LS



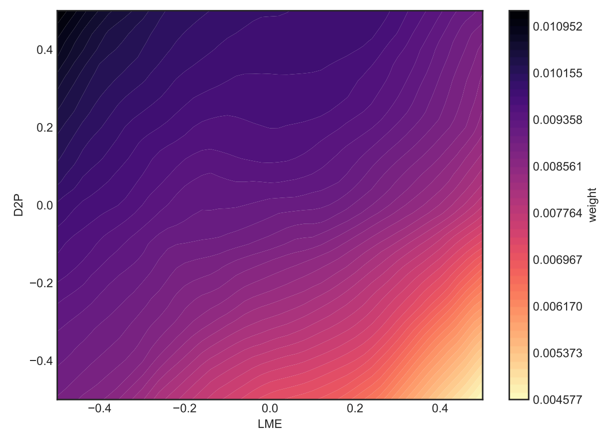
SDF weight ω as a one-dimensional function of covariates keeping the other covariates at their mean level. The covariates are Short-Term Reversal (ST_REV), Standard Unexplained Volume (SUV), Momentum (r12_2), Net Operating Assets (NOA), Selling, General and Administrative Expenses to Sales (SGA2S), Size (LME), Return on Net Operating Assets (RNA) and Turnover (LTurnover).

IA.E.2. Interaction between Characteristics

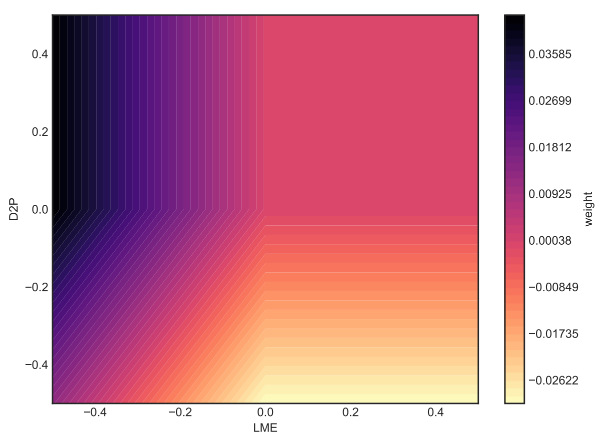
Figure IA.11. Interaction between Size (LME) and Dividend Yield (D2P)



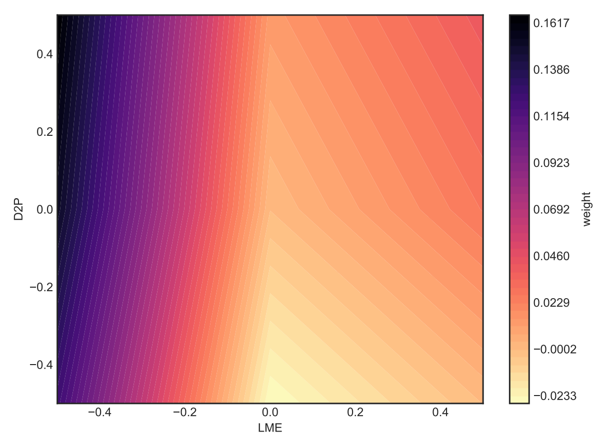
(a) GAN



(b) FFN



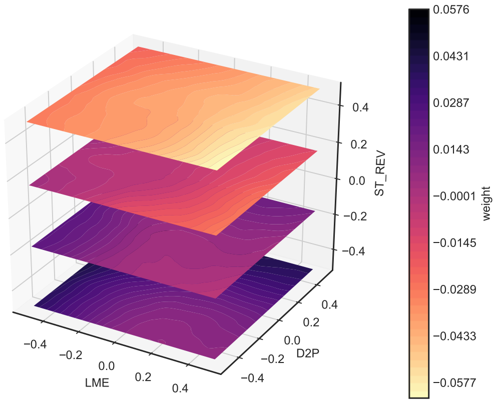
(c) EN



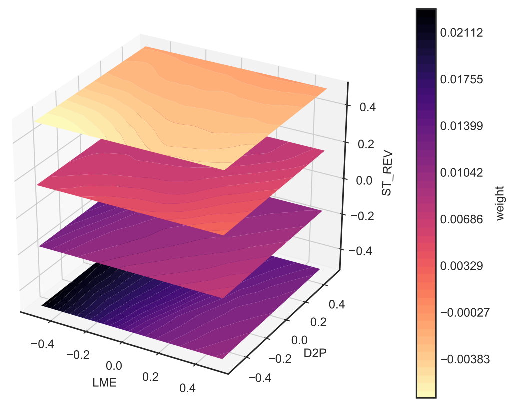
(d) LS

SDF weight ω as a two-dimensional function of covariates keeping the other covariates at their mean level. The two covariates are Size (LME) and Dividend Yield (D2P)

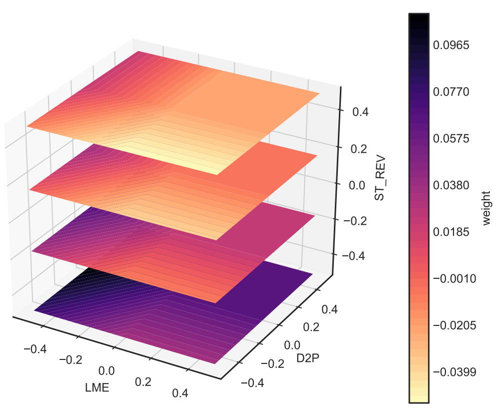
Figure IA.12. Interaction between Size (LME), Book to Dividend Yield (D2P) and Short-Term Reversal (ST_REV)



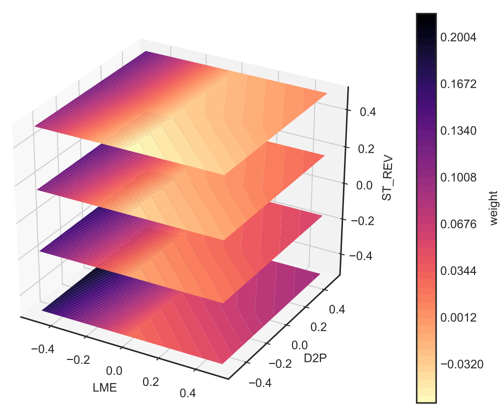
(a) GAN



(b) FFN



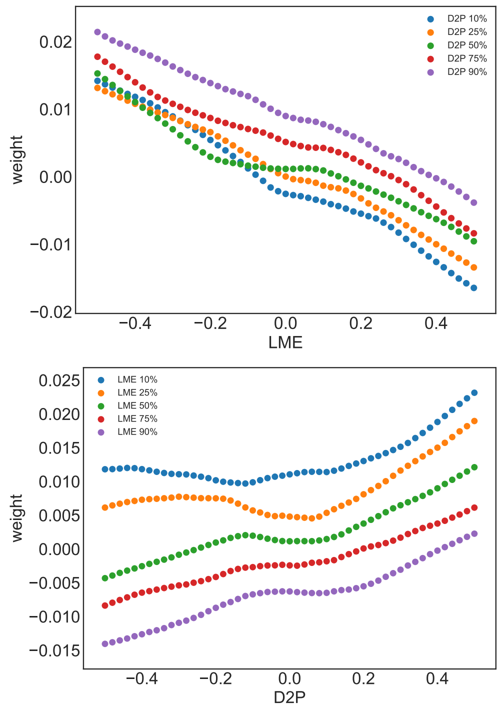
(c) EN



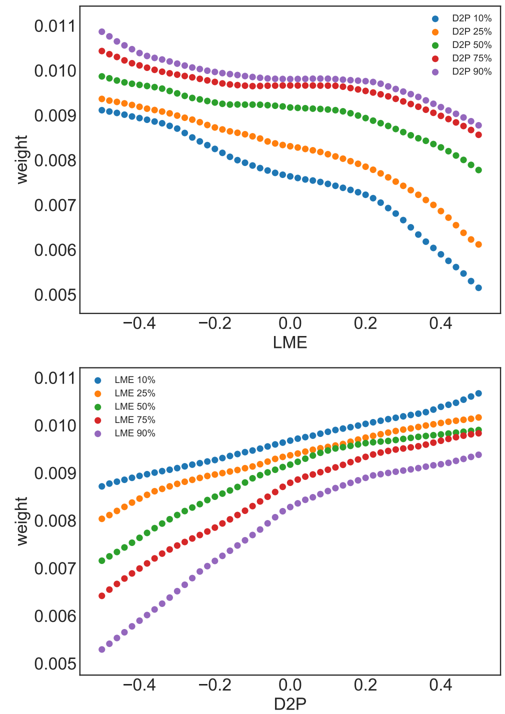
(d) LS

SDF weight ω as a three-dimensional function of covariates keeping the other covariates at their mean level. The three covariates are Size (LME), Book to Dividend Yield (D2P) and Short-Term Reversal (ST_REV).

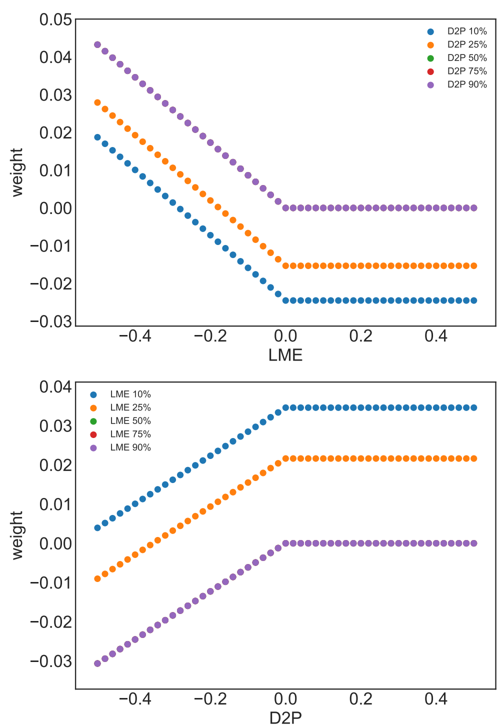
Figure IA.13. Interaction between Size (LME) and Dividend Yield (D2P)



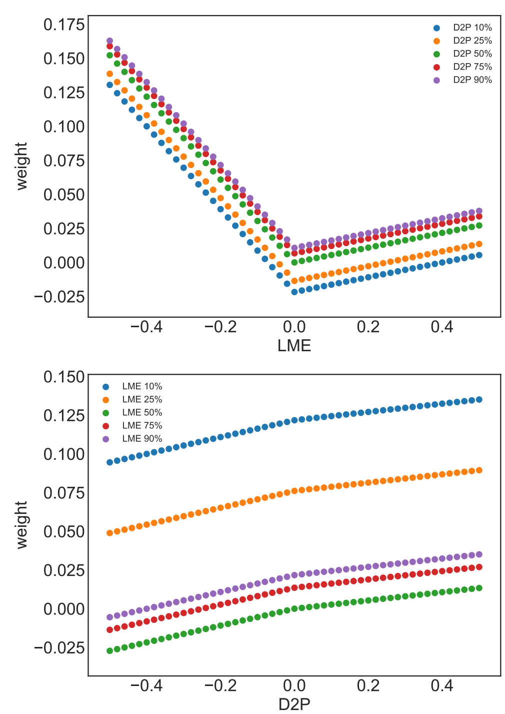
(a) GAN



(b) FFN



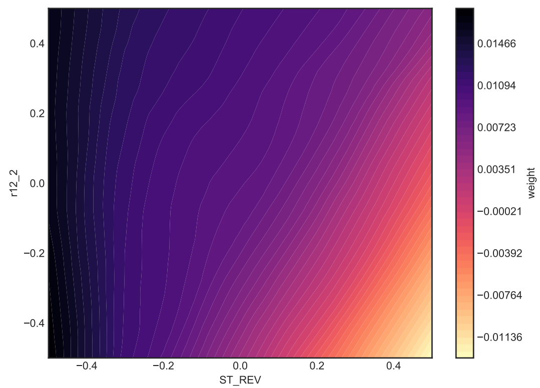
(c) EN



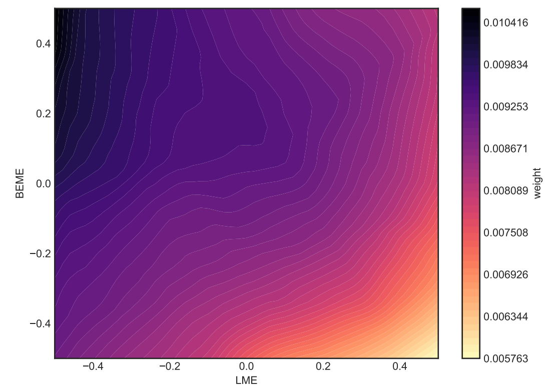
(d) LS

SDF weight ω as a two-dimensional function of covariates keeping the other covariates at their mean level. The two covariates are Size (LME) and Dividend Yield (D2P)

Figure IA.14. SDF weight ω as a Function of Covariates for FFN

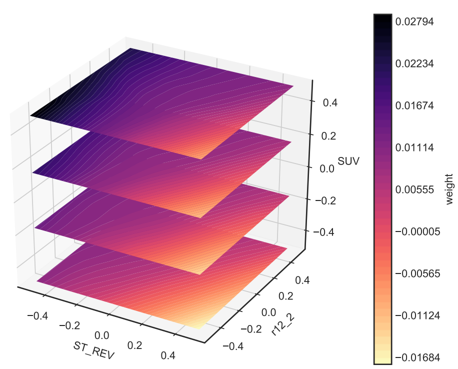


(a) Interaction between Short-Term Reversal (ST_REV) and Momentum (r12_2)

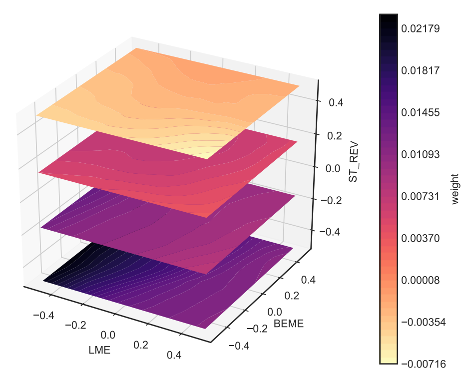


(b) Interaction between Size (LME) and Book to Market Ratio (BEME)

Figure IA.15. SDF weight ω as a Function of Covariates for FFN

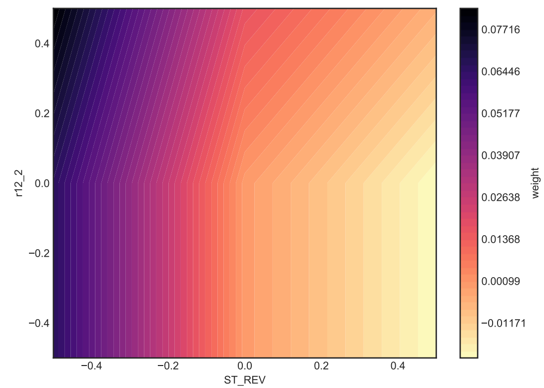


(a) Interaction between Short-Term Reversal (ST_REV), Momentum (r12_2) and Standard Unexplained Volume (SUV)

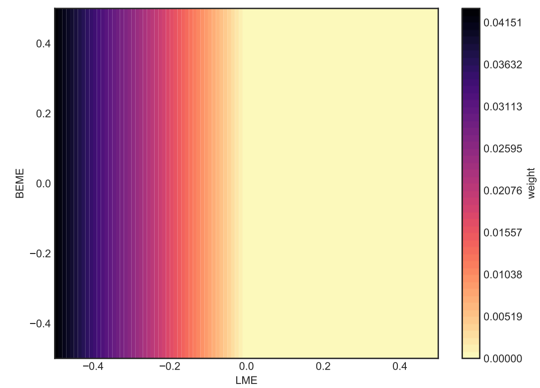


(b) Interaction between Size (LME), Book to Market Ratio (BEME) and Short-Term Reversal (ST_REV)

Figure IA.16. SDF weight ω as a Function of Covariates for EN

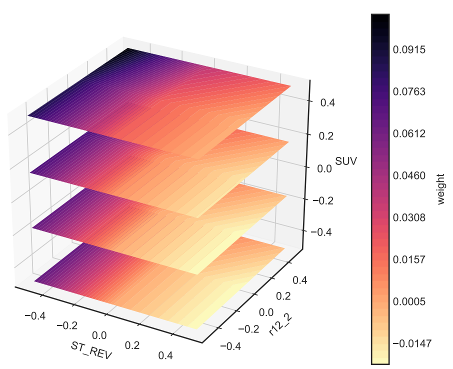


(a) Interaction between Short-Term Reversal (ST_REV) and Momentum (r12_2)

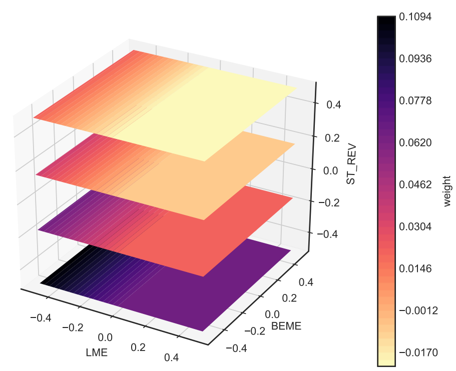


(b) Interaction between Size (LME) and Book to Market Ratio (BEME)

Figure IA.17. SDF weight ω as a Function of Covariates for EN

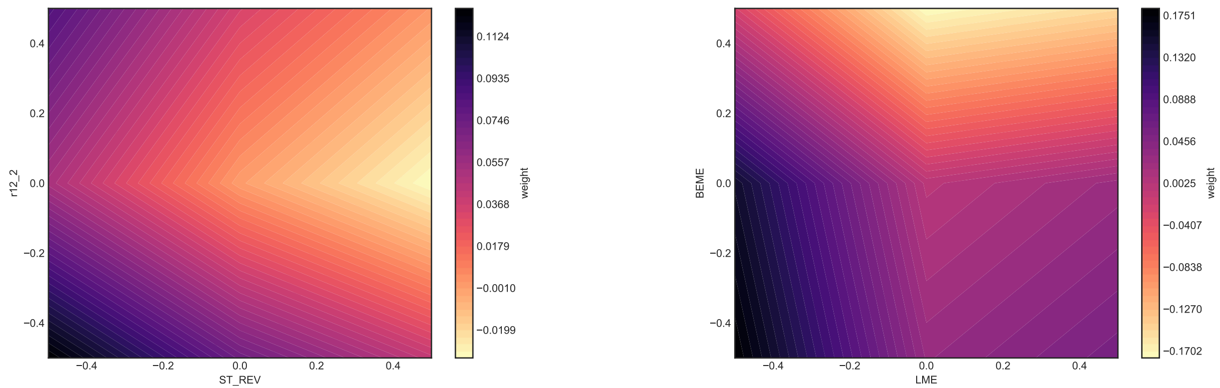


(a) Interaction between Short-Term Reversal (ST_REV), Momentum (r12_2) and Standard Unexplained Volume (SUV)



(b) Interaction between Size (LME), Book to Market Ratio (BEME) and Short-Term Reversal (ST_REV)

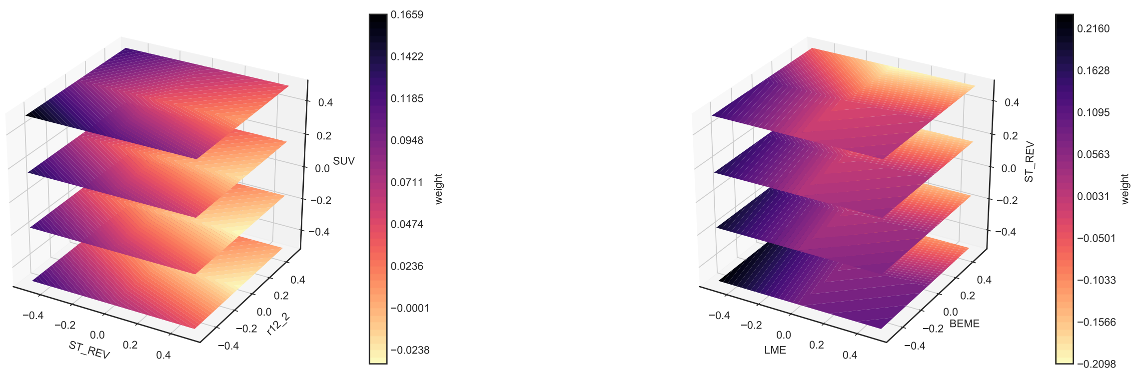
Figure IA.18. SDF weight ω as a Function of Covariates for LS



(a) Interaction between Short-Term Reversal (ST_REV) and Momentum (r12_2)

(b) Interaction between Size (LME) and Book to Market Ratio (BEME)

Figure IA.19. SDF weight ω as a Function of Covariates for LS



(a) Interaction between Short-Term Reversal (ST_REV), Momentum (r12_2) and Standard Unexplained Volume (SUV)

(b) Interaction between Size (LME), Book to Market Ratio (BEME) and Short-Term Reversal (ST_REV)