Internet Appendix for Deep Learning in Asset Pricing*

Luyang Chen† Markus Pelger‡ Jason Zhu§

April 6, 2022

Abstract

The Internet Appendix collects multiple results that support the results in the main text. The Internet Appendix provides an overview of conditional SDF models, applies the adversarial perspective to conditional multi-factor models, collects the empirical robustness results, illustrates the method in a simulation and collects additional empirical results.

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

†Stanford University, Institute for Computational and Mathematical Engineering, Email: lych@alumni.stanford.edu.
‡Stanford University, Department of Management Science & Engineering, Email: mpelger@stanford.edu.
§Stanford University, Department of Management Science & Engineering, Email: jzhu121@stanford.edu.

Contents

The Internet Appendix collects multiple results that support the results in the main text. Section IA.A provides a detailed overview of the various models for conditional SDFs and their relationship to our framework. Section IA.B shows that our conceptional framework is complementary to multi-factor models. We combine the conditional multi-factor model of Kelly, Pruitt, and Su (2019) with our model, and show that this can further improve asset pricing models.

Section IA.C shows that our findings are robust to small cap stocks, the choice of the tuning parameters, the time period under consideration and are not exploiting limits to arbitrage. When, allowing for a time-varying functional form by estimating the SDF on a rolling window, we find that it is highly correlated with the benchmark SDF and only leads to minor improvements. The estimation is robust to the choice of the tuning parameters. All of the best performing models selected on the validation data capture essentially the same asset pricing model. Our asset pricing model also performs well after excluding small and illiquid stocks from the test assets or the SDF. Thus, our GAN is not simply targeting pricing information that is subject to limits to arbitrage.

Section IA.D illustrates the different methods in a simulation example. We show in simulations that, first, the no-arbitrage condition in GAN is necessary to find the SDF in a low signal-to-noise setup, second, the flexible form of GAN is necessary to correctly capture the interactions between characteristics, and, third, the RNN with LSTM is necessary to correctly incorporate macroeconomic dynamics in the pricing kernel. We have designed a simple illustrative simulation setup to convey these points and to show that the forecasting approach or the simple linear model formulations cannot achieve these goals.

Section IA.E shows the performance for multiple GAN iterations. While additional GAN iterations improve the in-sample results on the training data, the results on the validation and test data do not improve further after one GAN iteration. Section IA.F collects the additional results for the illustrative GAN example based on size, value and investment. In Section IA.G we confirm that SDF loadings are predictive for future returns and that the results for β sorted deciles portfolios hold for equally and value weighted portfolios. Section IA.H reports the structure of the SDF as a function of characteristics for FFN, EN and LS. We show that the forecasting approach FFN seems to capture less interaction effects. By construction these interaction effects are ruled out for the linear models. Section IA.I provides further results for the effect of market capitalization on the performance of machine learning investment portfolios. In Section IA.J we provide the asset pricing results and functional form of the SDF for additional characteristics. Last but not least, Section IA.K describes the macroeconomic and firm-specific variables in detail.

IA.A. Overview of Conditional SDF Models

IA.A.1. A General Conditional Stochastic Discount Factor

We survey the most recent advances of using machine learning methods in asset pricing and explain their differences. We start with a brief review of the general no-arbitrage framework. 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 strictly positive stochastic discount factor (SDF) M_{t+1} 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

$$\mathbb{E}_t\left[M_{t+1}R^e_{t+1,i}\right] = 0 \quad\Leftrightarrow\quad \mathbb{E}_t[R^e_{t+1,i}] = \underbrace{\left(-\frac{\mathrm{Cov}_t(R^e_{t+1,i}, M_{t+1})}{\mathrm{Var}_t(M_{t+1})}\right)}_{\beta^{\mathrm{SDF}}_{t,i}} \cdot \underbrace{\frac{\mathrm{Var}_t(M_{t+1})}{\mathbb{E}_t[M_{t+1}]}}_{\lambda_t},$$

where $\beta^{\mathrm{SDF}}_{t,i}$ 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 SDF is an affine transformation of the tangency portfolio. Without loss of generality we consider the SDF formulation

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

This is the projection on the return space and corresponds to an affine transformation of the SDF that has the smallest conditional variance. The fundamental pricing equation $\mathbb{E}_t[R^e_{t+1}M_{t+1}] = 0$ implies the SDF weights

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

which are the portfolio weights of the conditional mean-variance efficient portfolio.[1] While this provides a theoretical solution to the SDF problem, it is practically not feasible as it requires knowledge of a large dimensional conditional covariance matrix. The estimation of an unconditional large dimensional covariance matrix is possible under additional assumptions (typically an approximate factor model), but it is an open question how to obtain a feasible estimator in the conditional case. Second, the estimation of the conditional mean is a very challenging problem with extremely noisy estimates. As we work with individual stock returns, we cannot resort to unconditional estimates.

We define the tangency portfolio as $F_{t+1} = \omega_t^\top R^e_{t+1}$ and will refer to this traded factor as the SDF. The asset pricing equation can now be formulated as

$$\mathbb{E}_t[R^e_{t+1,i}] = \frac{\mathrm{Cov}_t(R^e_{t+1,i}, F_{t+1})}{\mathrm{Var}_t(F_{t+1})} \cdot \mathbb{E}_t[F_{t+1}] = \beta^{\mathrm{SDF}}_{t,i}\mathbb{E}_t[F_{t+1}].$$

Hence, no-arbitrage implies a one-factor model

$$R^e_{t+1,i} = \beta^{\mathrm{SDF}}_{t,i} F_{t+1} + \epsilon_{t+1,i}$$

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.

Different asset pricing model impose different structures on the SDF weights ω and SDF loadings β^{SDF}. The estimation challenge arises from modeling the conditional expectation $\mathbb{E}_t[.]$ which can

[1] Any portfolio on the globally efficient frontier achieves the maximum Sharpe ratio. These portfolio weights represent one possible efficient portfolio. An alternative formulation would be $M_{t+1} = 1 - \sum_{i=1}^{N} \omega_{t,i}(R^e_{t+1,i} - \mathbb{E}_t[R^e_{t+1,i}])$ which results in the conventional conditional tangency portfolio weights $\omega_t = \mathrm{Cov}_t(R^e_{t+1})^{-1}\mathbb{E}_t[R^e_{t+1}]$.

depend in a complex way on a large number of asset-specific and macroeconomic variables. This is where machine learning tools are essential to deal in a flexible way with the large dimensionality of the problem. Importantly, we need a model and estimator for both, the SDF weights ω and SDF loadings β^{SDF}, to explain individual stock returns.

IA.A.2. Characteristic Projection and Unconditional Models

The most common way is to translate the problem into an unconditional asset pricing model on sorted portfolios. Under additional assumptions one could obtain a valid SDF M_{t+1} conditional on a set of asset-specific characteristics $I_{t,i}$ by its projection on the return space:

$$M_{t+1} = 1 - \omega_t^\top R_t^e \qquad \text{with } \omega_{t,i} = f\left(I_{t,i}\right),$$

where $I_{t,i}$ is a vector of q characteristics observed for N stocks and $f(\cdot)$ is a general, potentially nonlinear and non-separable function. Most of the reduced-form asset pricing models approximate this function by a (potentially very large) set of simple managed portfolios $f_j(\cdot)$, such that $f\left(I_{t,i}\right) \approx \sum_{j=1}^{\tilde{N}^{\text{basis}}} f_j\left(I_{t,i}\right) \tilde{w}_j$. The SDF then becomes a linear combination of these managed portfolios with constant weights $\tilde{\omega}_j$:

$$M_{t+1} = 1 - \sum_{j=1}^{\tilde{N}^{\text{basis}}} \tilde{w}_j \tilde{R}_{t+1,j} \qquad \text{with } \tilde{R}_{t+1,j} = \sum_{i=1}^{N} f_j\left(I_{t,i}\right) R_{t+1,i}^e, \tag{IA.2}$$

where $\tilde{R}_{t+1}$ are the returns of $\tilde{N}^{\text{basis}}$ managed portfolios that correspond to different basis functions in the characteristic space. The number of basis portfolios increases by the complexity of the basis functions and the number of characteristics. The most common managed portfolios are sorted on characteristic quantiles, that is, they use indicator functions based on characteristic quantiles to approximate $f(I_{t,i})$. Popular sorts are the size and value double-sorted portfolios of Fama and French (1992), that are also used to construct their long-short factors. Note, that these characteristic managed portfolios do not include macroeconomic variables which are not asset specific, in contrast to our GAN model. In order to simplify the exposition, we will focus now on the information set, which only conditions on firm-specific characteristics.

Importantly, the translation of the conditional model into an unconditional model implicitly imposes the following assumptions: First, the asset pricing modeler includes all basis functions $f_j\left(I_{t,i}\right)$ that are necessary to approximate the SDF weights $f\left(I_{t,i}\right) \approx \sum_{j=1}^{\tilde{N}^{\text{basis}}} f_j\left(I_{t,i}\right) \tilde{w}_j$. If the asset pricing modeler omits relevant basis functions, for example interactions between characteristics, the SDF will be misspecified. Second, the returns of the characteristics managed portfolios $\tilde{R}_{t+1,j}$ are mean and variance stationary, i.e. in particular they have constant mean and variance.

In this case the SDF can be obtained by solving an unconditional mean-variance optimization problem based on the characteristic managed portfolios:[2]

$$\tilde{w} = \text{Var}\left(\tilde{R}_t\right)^{-1} \mathbb{E}[\tilde{R}_t] \qquad \tilde{w} \in \mathbb{R}^{\tilde{N}^{\text{basis}}}. \tag{IA.3}$$

The SDF weights for individual stocks are then given by the SDF weights for the characteristic managed portfolios and the weights of individual stocks in those portfolios, that is $w_{t,i} = \sum_{j=1}^{\tilde{N}^{\text{basis}}} f_j(I_{t,i})\tilde{w}_j$.

Most asset pricing models fall into this category. Importantly, these models do not provide guidance on how to obtain the loadings for individual stock returns β_t^{SDF}, which is a separate problem. These unconditional models can only readily be used to explain the expected returns of the specific characteristic managed portfolios that have been used for their construction, but require an additional model to obtain the individual stock loadings β_t^{SDF}.

While Equation IA.3 describes the population solution, we need additional assumptions to obtain a feasible estimator. If the number of characteristic managed portfolios $\tilde{N}^{\text{basis}}$ is large, either because the number of characteristics is large and/or because the functional form that should be approximated is complex and requires many basis functions, we cannot apply simple sample estimators and require regularization. The naive sample estimator for $\tilde{w}$ would be $\tilde{w}_{\text{MV}} = \hat{\Sigma}^{-1}\hat{\mu}$, where $\hat{\Sigma}^{-1}$ is the sample estimator of the covariance matrix and $\hat{\mu}$ is the estimator of the mean of the of characteristic managed portfolios $\tilde{R}_t$. Kozak, Nagel, and Santosh (2020) (KNS) propose to apply a modified ridge penalty and lasso penalty to this regression problem to estimate $\tilde{w}$:

$$\tilde{w}_{\text{KNS}} = \arg\min_{\tilde{w}} \left[\left(\hat{\mu} - \hat{\Sigma}\tilde{w}\right)^\top \hat{\Sigma}^{-1} \left(\hat{\mu} - \hat{\Sigma}\tilde{w}\right) + 2\nu_1 \sum_{j=1}^{\tilde{N}} |\tilde{w}_i| + \nu_2 \tilde{w}^\top \tilde{w} \right],$$

where ν_1 is a lasso penalty and ν_2 a ridge-type penalty. Kozak, Nagel, and Santosh (2020) advocate to use their estimator in the PCA space (which implies a diagonal matrix for $\hat{\Sigma}$), which as shown by Lettau and Pelger (2020) yields a closed-form solution:

$$\hat{w}_{\text{KNS},i} = \begin{cases} \frac{\hat{\mu}_i - \nu_1}{\hat{\sigma}_i^2 + \nu_2} & \text{if } \hat{\mu}_i \geq \nu_1 \\ 0 & \text{if } \hat{\mu}_i < \nu_1, \end{cases}$$

where $\hat{\sigma}_i^2$ are the variances of the PCA factors and $\hat{\mu}_i$ their mean returns. The choice of the lasso penalty maps into a sparse representation of a small number of basis assets, which are selected based on their means. Importantly, the KNS model does not readily provide a framework to explain the mean return for individual stocks or assets, which are not used as basis assets.

The Asset-Pricing Trees (AP-Trees) of Bryzgalova, Pelger, and Zhu (2019) generalize the SDF approach of KNS among several dimensions. First, given a choice of basis assets, they deal with the large uncertainty in mean estimation, by introducing mean-shrinkage in addition to variance shrinkage. Second and more importantly, they use recursive tree basis functions to generate the characteristic

[2] In order to keep the notation closer to the literature we consider here the SDF $M_{t+1} = 1 - \sum_{j=1}^{J} \tilde{w}_j (\tilde{R}_{t+1,j} - \mathbb{E}[\tilde{R}_{t+1,j}])$ which yields the covariance matrix instead of the uncentered second moment matrix in the mean-variance problem. However, this normalization is inconsequential for the results.

projected portfolios. These more flexible basis functions include the conventional univariate sorts as special cases, but allow also to deal with interaction effects between multiple characteristics. A particularly appealing element is to leverage the overlapping structure in trees and to use the robust SDF recovery to select the tree basis functions to span the SDF. This has the crucial advantage of obtaining a low dimensional representation similar to PCA factors, while retaining interpretability and capturing more general patterns. The primary focus of the basis functions obtained by Bryzgalova, Pelger, and Zhu (2019) is to use them as informative test assets for unconditional asset pricing models.

IA.A.3. Inversion of Unconditional Models

In order to use an SDF, which is estimated as an unconditional model from conditional portfolio sorts or projections, we need to invert the conditional projection. In other words, the unconditional SDF model on the projected returns $\tilde{R}_{t+1}$ needs to be mapped back into the individual stock returns R^e_{t+1}. The fundamental challenge is that this requires the modeling of conditional covariances of the individual stocks with the unconditional SDF model. Hence, specifying a set of portfolio sorts and estimating an unconditional SDF model on those portfolio sorts, is not sufficient. We need to estimate the conditional covariance of each stock with the SDF portfolios:

$$\beta^{\text{SDF}}_{t,i} = \frac{\text{Cov}_t(R^e_{t+1,i}, F_{t+1})}{\text{Var}_t(F_{t+1})} = \frac{\text{Cov}_t(R^e_{t+1,i}, \tilde{w}^\top \tilde{R}_{t+1})}{\text{Var}_t(\tilde{w}^\top \tilde{R}_{t+1})}.$$

Obviously, using an unconditional covariance is not internally consistent if stocks have time-varying characteristics. In other words, the assumption of a mean and covariance stationary model on portfolio sorts, implies a conditional model for individual stock returns. An internally consistent model would use the same basis functions $f_j(I_{t,i})$, which are used to span the conditional SDF weights $w_{t,i}$ to span the conditional SDF loadings $\beta^{\text{SDF}}_{t,i}$. This requires an additional non-parametric regression with the corresponding challenges in a high-dimensional setup.

IA.A.4. Unconditional Factor Models

The linear factor model literature imposes the additional assumption that a small number of risk factors based on characteristic managed portfolios should span the SDF. The majority of the literature studies unconditional factor models on characteristic managed portfolios. This requires the same assumptions as in the unconditional SDF models in Section IA.A.2. In addition, it imposes that the excess returns of characteristic managed portfolios follow a factor structure

$$\tilde{R}_{t,i} = \tilde{F}_t \tilde{\beta}_i^\top + e_{t,i} \qquad i = 1, ..., \tilde{N}^{\text{basis}}, \;\; t = 1, ..., T \tag{IA.4}$$

The factors can be observed fundamental factors, for example the Fama-French factor model of Fama and French (2015), or latent asset pricing factors estimated from the unconditional moments of $\tilde{R}_t$ by PCA or its improvement RP-PCA (Lettau and Pelger (2020)). The tangency portfolio, that is spanned by $\tilde{F}_t$, has the factor weights

$$\tilde{\omega}_{\tilde{F}} = \Sigma_{\tilde{F}}^{-1} \mu_{\tilde{F}}, \tag{IA.5}$$

where $\mu_{\tilde{F}}$ and $\Sigma_{\tilde{F}}$ are the mean and variance-covariance matrix of $\tilde{F}_t$. The implied SDF is given by $M_t = 1 - \tilde{\omega}_{\tilde{F}}^\top \left(\tilde{F}_t - \mathbb{E}_t[\tilde{F}_t] \right)$. The SDF weights of the factors in Equation IA.5 follow from pricing the factors with the factors themselves, that is the factors serve as basis and test assets to obtain the SDF weight. Importantly, the assumption of constant loadings $\tilde{\beta}$ is only reasonable on the characteristic managed portfolios, but individual stock returns must have time-varying loadings if characteristics are time-varying. Hence, estimating any factor model on portfolio sorts does not readily imply a factor model and an SDF on individual stock returns. The loadings for individual stock returns are given by the conditional loadings $\beta_{t,i}^{\text{SDF}}$:

$$\beta_{t,i}^{\text{SDF}} = \frac{\text{Cov}_t(R^e_{t+1,i}, F_{t+1})}{\text{Var}_t(F_{t+1})} = \frac{\text{Cov}_t(R^e_{t+1,i}, \tilde{\omega}_{\tilde{F}}^\top \tilde{F}_{t+1})}{\text{Var}_t(\tilde{\omega}_{\tilde{F}}^\top \tilde{F}_{t+1})}.$$

If for example one of the risk factors is a “size” factor, which is a long-short factor based on size sorted quantiles, such a risk factor should have a larger loading $\tilde{\beta}$ for a portfolio that includes only small stocks compared to a portfolio that includes only large stocks. As the market capitalization of stocks is time-varying, this mechanically implies that the loadings of a size factor have to be time-varying for individual stocks. More specifically, as the weights to construct the size factor are a function of the market capitalization, it mechanically implies that the loadings for individual stocks have to be a function of the market capitalization as well.

In practice, unconditional factors, which have been constructed based on a specific choice of portfolio sorts, e.g. the 3 Fama-French factors in Fama and French (1992) based on double-sorted size and book-to-market sorted quantiles or PCA factors estimated from a large set of univariate quantile sorts, have been applied to test assets which are not the basis assets or to individual stock returns. The usual application runs unconditional time-series regressions on the set of factors. This is only valid if the loadings to the factors and the moments of the factors are constant over time, which by construction cannot be the case for individual stocks with time-varying characteristics. A common practice is to run rolling window regressions to obtain local loadings on individual stocks. While this can potentially address the time-variation, it does not readily result in a conditional factor model which can be evaluated on different stocks out-of-sample.

The extreme unbalancedness in individual stock returns severely limits the use of local window regressions with unconditional factors. The stocks in the first part of the data are too a large extent not available in the second part, which restricts the number of stocks that can be used for an out-of-sample evaluation. In a conditional model we estimate the conditional SDF weights ω and loadings β^{SDF} as a function of characteristics on different stocks for the estimation and evaluation. Therefore, we can evaluate the model on stocks that are not available in the first part of the data. Importantly, the conditional model also allows us to directly study the economic sources of risk in terms of firm characteristics.

IA.A.5. Conditional Factor Models

A conditional factor model assumes that the SDF is spanned by a linear combination of conditional risk factors, and hence it restricts the basis assets that span the SDF. In contrast to the unconditional model, the SDF weights and loadings of the conditional risk factors are a function of the characteristics and hence time-varying. We use the Instrumented Principal Component Analysis (IPCA) of Kelly, Pruitt, and Su (2019) to illustrate this setup. This conditional factor model directly models individual stock returns as a function of characteristics given by $R^e_{t+1,i} = b_{t,i}^\top f_{t+1}^{\text{IPCA}} + \epsilon_{t+1,i}$ with $b_{t,i} = I_{i,t}^\top \Gamma_b$. Instead of allowing the SDF to be $M_{t+1} = 1 - \sum_{i=1}^N \omega_{t,i} R^e_{t+1,i}$, it is restricted to $M_{t+1} = 1 - \sum_{k=1}^K w_{t,k}^{\text{IPCA}} f_{t+1,k}^{\text{IPCA}}$, where the IPCA factors are estimate with IPCA. Even after estimating the conditional factors, we still need to solve a conditional GMM problem. The fundamental moment equation becomes

$$\mathbb{E}_t \left[\left(1 - \sum_{k=1}^K w_{t,k}^{\text{IPCA}} f_{t+1,k}^{\text{IPCA}} \right) R^e_{t+1,i} \right] = 0. \tag{IA.6}$$

If we assume no misspecification and that the $\epsilon_{t+1,i}$ are independent of the IPCA factors with a conditional mean of zero, then population solution for the SDF weights to the equation IA.13 becomes

$$w_t^{\text{IPCA}} = \text{Cov}_t \left(f_{t+1}^{\text{IPCA}} \right)^{-1} \mathbb{E}_t[f_{t+1}^{\text{IPCA}}],$$

which requires the estimation of conditional moments. These conditional moments can be estimated by a prediction of f_{t+1}^{IPCA} and $f_{t+1}^{\text{IPCA}^2}$ using the information set at time t. Given the SDF weights w_t^{IPCA}, the SDF loadings are a linear combination of the conditional factor loadings $\beta_t^{\text{SDF}} = b_{t,i}^\top \nu_t^{\text{IPCA}}$ such that

$$R^e_{t+1,i} = b_{t,i}^\top \nu_t^{\text{IPCA}} w_t^{\text{IPCA}^\top} f_{t+1}^{\text{IPCA}} + \epsilon_{t+1,i}.$$

Under additional assumptions it holds that $\nu_t^{\text{IPCA}} = w_t^{\text{IPCA}} \cdot c_t$ up to a proportionality constant c_t, but this does not need to hold in the general case. However, the choice of w_t^{IPCA} uniquely pins down ν_t^{IPCA}, that is, selecting a combination of the conditional factors results in a specific combination of the conditional loadings.

Several papers with conditional factors assume unconditional SDF weights, that is, they set

$$w^{\text{IPCA}} = \text{Cov} \left(f_{t+1}^{\text{IPCA}} \right)^{-1} \mathbb{E}[f_{t+1}^{\text{IPCA}}].$$

However, this choice of SDF weights also restrict the SDF loadings to be a specific linear combination of the conditional factors loadings, that is, w^{IPCA} determines ν_t^{IPCA}. The residuals relative to a conditional multivariate cross-sectional regression on b_t are not the same residuals as those relative to the SDF loading β_t^{SDF}, when the SDF weights are set to the unconditional tangency portfolio weights $w^{\text{IPCA}} = \text{Cov} \left(f_{t+1}^{\text{IPCA}} \right)^{-1} \mathbb{E}[f_{t+1}^{\text{IPCA}}]$. A coherent asset pricing model needs to impose the same constraints on how it combines the conditional factors and conditional loadings into a one-factor representation. In Section IA.B, we show that for the IPCA model the unconditional mean-variance combination of the conditional factors does not explain the mean returns of stocks, while the conditional combination of the conditional factors estimated with the GAN framework results in a better asset pricing model.

IA.A.6. *Adversarial Estimation and Mean-Variance Optimization*

We compare the adversarial estimation with the mean-variance estimation of the SDF. In order to provide intuition, we start with the special case of pre-specified basis functions to estimate the conditional moments. Intuitively, we assume that a pre-specified set of portfolios sorts, for example univariate sorting based on characteristics, is sufficiently rich to span the SDF. Our starting point is the conditional moment equation $\mathbb{E}_t\left[M_{t+1}R^e_{t+1,i}\right]=0$, which implies the unconditional moments

$$\mathbb{E}[M_{t+1}R^e_{t+1,i}g(I_{t,i})]=0 \tag{IA.7}$$

for any function $g(.)$. In the case of pre-specified basis functions, we assume that the SDF weight and the conditioning functions, that are sufficient to identify the SDF weights, are spanned by those basis functions:

$$w(I_{t,i})=\sum_{j=1}^{N^{\text{basis}}} f_j(I_{t,i})\tilde{w}_j, \qquad g(I_{i,t})=\sum_{j=1}^{N^{\text{test}}} g_j(I_{t,i})\tilde{w}^{\text{test}}_j.$$

The N^{basis} characteristic managed portfolios $\tilde{R}^{\text{basis}}_{t+1,j}=\sum_{i=1}^N f_j(I_{t,i})R^e_{t+1,i}$ can be interpreted as the basis assets to span the SDF. The N^{test} characteristic managed portfolios $\tilde{R}^{\text{test}}_{t+1,j}=\sum_{i=1}^N g_j(I_{t,i})R^e_{t+1,i}$ correspond to the test assets, which are used to identify and estimate the constant weights $\tilde{w}$. For the pre-specified basis functions, we impose the additional assumption that the basis and test assets have stationary first and second moments.

First, we consider exact identification in the GMM problem. If we set the basis assets identical to the test assets, that is $\tilde{R}^{\text{basis}}_t=\tilde{R}^{\text{test}}_t$, then the unconditional moment equation

$$\mathbb{E}\left[\left(1-\tilde{w}^\top\tilde{R}^{\text{basis}}_t\right)\tilde{R}^{\text{test}}_t\right]=0 \tag{IA.8}$$

has the solution

$$\tilde{w}=\mathbb{E}\left[\tilde{R}^{\text{basis}}_t\tilde{R}^{\text{basis}\top}_t\right]^{-1}\mathbb{E}\left[\tilde{R}^{\text{basis}}_t\right]. \tag{IA.9}$$

If we approach this problem from an adversarial perspective, the solution would be the same, that is, the problem

$$\min_{\tilde{w}}\max_{\tilde{R}^{\text{test}}_t}\left\|\mathbb{E}\left[\left(1-\tilde{w}^\top\tilde{R}^{\text{basis}}_t\right)\tilde{R}^{\text{test}}_t\right]\right\|_2^2 \tag{IA.10}$$

has the solution in Equation IA.9. This is because we have a GMM problem with exact identification, where the number of parameters is exactly the same as the number of moments. While the discussion so far is based on population moments, it becomes more complex when we estimate it empirically.

When the number of basis and test assets $N^{\text{basis}} = N^{\text{test}}$ is large relative to the number of time-series observations T, then we need to regularize the second (and potentially also first) sample moment of $\tilde{R}_t^{\text{basis}}$ by using some form of regularization. The possible regularizations include using the PCA factors based on the covariance matrix of $\tilde{R}_t^{\text{basis}}$, or a ridge, lasso or elastic net penalty in the mean-variance optimization problem. These regularizations have in common that they either hard-threshold (in the case of PCA) or soft-threshold (for ridge penalties) the components that explain less variation in the relatively arbitrarily chosen set of basis and test assets. One problem of such an approach is that it misses weak factors, which are relevant for explaining mean returns, but would be down-weighted by the regularization.

In general, we have an overidentified GMM problem, where the basis assets for the SDF and the test assets do not need to be the same. This means that the space of test assets can be larger than the set of basis assets and $N^{\text{test}} > N^{\text{basis}}$. Under appropriate assumptions, the solution to the unconditional moments in IA.8 becomes

$$\tilde{w} = \left(\mathbb{E}\left[\tilde{R}_t^{\text{basis}}\tilde{R}_t^{\text{test}\top}\right]\Omega_{\text{GMM}}\mathbb{E}\left[\tilde{R}_t^{\text{test}}\tilde{R}_t^{\text{basis}\top}\right]\right)^{-1}\mathbb{E}\left[\tilde{R}_t^{\text{basis}}\tilde{R}_t^{\text{test}\top}\right]\Omega_{\text{GMM}}\mathbb{E}\left[\tilde{R}_t^{\text{test}}\right], \qquad \text{(IA.11)}$$

where Ω_{GMM} is the GMM weighting matrix for the moments. In the special case of an identity matrix, that is $\Omega_{\text{GMM}} = I_{N^{\text{test}}}$, all test assets receive the same weight. The sample solution obviously requires some form of regularization similar to the case with exact identification.

If the number of test assets is larger than the number of basis assets, then solution of the adversarial GMM in IA.10 is more complex and can in general not be mapped into the form of Equation IA.11. The adversarial approach has a number of advantages relative to the conventional regularized GMM framework:

1. It is more robust to misspecification as shown in Hansen and Jagannathan (1997). More specifically, if the set of basis functions is chosen too restrictive or the form or regularization is not appropriate, the adversarial approach provides a more robust pricing kernel.
2. The adversarial GMM can address the issue of weak asset pricing factors. The test assets in most asset pricing applications are to some degree arbitrarily chosen (usually by using the same set of double-sorted portfolios). A weak asset pricing factor might only explain a small amount of variation for the specific choice of test assets, but might be important to explain a certain component of the SDF, that is relevant for the risk premium. The adversarial GMM will upweight this component, while a conventional regularization based on ridge or PCA will neglect these weak factors.
3. The adversarial GMM will ensure identification of all parameters of the SDF. If certain components of the SDF are based on weak factors, the conventional GMM with regularization might not identify those components.

The general problem considered in this paper is substantially harder as we do not use pre-specified basis functions for the basis and test assets, but learn them from the data. The set of potential non-parametric basis functions that can be generated by the neural networks is extremely large and goes into many millions, when considering all possible non-linear interactions. Under this level of generality and when we allow for this degree of flexibility, there is no meaningful way to solve this problem in a conventional regularized mean-variance framework. Here, the adversarial GMM framework is essential to obtain a feasible solution. Hence, in addition to the benefits outlined above, the adversarial approach provides a feasible solution to the general problem that cannot be solved otherwise. Furthermore, the non-parametric adversarial GMM approach is based on weaker assumptions as it does not require the stationarity of the first and second moments of a large number of pre-specified sorted portfolios, which is assumed in IA.11. The non-parametric adversarial GMM only requires the orthogonality of $\left(1 - \sum_{i=1}^{N} w(I_{t,i}) R^e_{t+1,i}\right)$ from $R^e_{t+1,i} g(I_{t,i})$, which allows for time-varying moments.

IA.B. SDF of Multi-Factor Models

IA.B.1. Main Results

Our GAN framework is *complementary* to conditional and unconditional multi-factor models. Multi-factor models are based on the assumption that the SDF is a linear combination of the multiple factors. In an unconditional multi-factor model with constant factor loadings and pricing errors, the SDF can easily be constructed as the unconditional mean-variance efficient combination of the factors. The time series regression pricing error in such a multi-factor model is identical to that of a one-factor regression on the SDF. However, this relationship does not hold out-of-sample for unconditional models and breaks down in-sample and out-of-sample for conditional multi-factor models. So far, the factor literature has mainly focused on extracting the factors and their loadings, but has been largely silent on the construction of a coherent conditional SDF framework based on a conditional factor structure. Our GAN framework can help to close this gap. The fundamental moment condition of our GAN framework to construct the SDF is

$$\min_{\omega} \max_{g} \frac{1}{N} \sum_{j=1}^{N} \left\| \mathbb{E}\left[\left(1 - \sum_{i=1}^{N} \omega(I_t, I_{t,i}) R^e_{t+1,i} \right) R^e_{t+1,j} g(I_t, I_{t,j}) \right] \right\|^2, \tag{IA.12}$$

and can incorporate the restriction that the SDF is a linear combination of factors.

We use one of the most important conditional multi-factor models and combine it with the GAN framework to estimate its SDF. Instrumented Principal Component Analysis (IPCA) developed by Kelly, Pruitt, and Su (2019) allows for latent factors and time-varying loadings. Both elements are key as we need a conditional factor model with time-varying loadings to explain individual stock returns and want to estimate the best performing factors without taking an a prior stand on what the factors are. Note that this model includes simple unconditional factor models as a special case. As most of the literature Kelly, Pruitt, and Su (2019) use the multi-factor framework to calculate pricing errors and report Sharpe ratios for the unconditional mean-variance efficient combination of the factors, but they do not estimate pricing errors and Sharpe ratios for the same one-factor model of the SDF. Here we show that using the additional economic structure of spanning the SDF with IPCA factors and combining it with the GAN framework can lead to an even better asset pricing model. We condition only on firm characteristics to make the results more comparable with the original IPCA framework.

IPCA assumes a K-factor model where the loadings are a linear function of the characteristics:[3]

$$R^e_{t+1,i} = a_{t,i} + b_{t,i}^\top f^{\text{IPCA}}_{t+1} + \epsilon_{t+1,i} \qquad b_{t,i} = I_{i,t}^\top \Gamma_b.$$

Any multi-factor model assumes that the SDF is spanned by the factors, that is,

$$F = \sum_{k=1}^{K} \omega^f(I_{k,t}, I_t) f^{\text{IPCA}}_{t+1,k}. \tag{IA.13}$$

Appendix IA.A provides further details. Under weak assumptions, the SDF weights are given by the conditional tangency portfolio based on the factors: $\omega^f(I_{k,t}, I_t) = \text{Cov}_t\left(f^{\text{IPCA}}_{t+1}, f^{\text{IPCA}\top}_{t+1}\right)^{-1} \mathbb{E}_t\left[f^{\text{IPCA}}_{t+1}\right]$. However, the usual approach in most papers is to use constant weights, which are set to the unconditional mean-variance efficient portfolio weights, that is, $\omega^{\text{I-SR}} = \text{Cov}\left(f^{\text{IPCA}}_{t+1}, f^{\text{IPCA}\top}_{t+1}\right)^{-1} \mathbb{E}\left[f^{\text{IPCA}}_{t+1}\right]$. An alternative approach to obtain a one-factor representation with constant weights is to find a linear combination of the IPCA factors and loadings that either minimizes the cross-sectional pricing errors or the amount of unexplained variation. This means that we obtain the SDF weights $\omega^{\text{I-XS}} \in \mathbb{R}^K$ and the corresponding loading weights $v^{\text{I-XS}} \in \mathbb{R}^K$ such that the residuals $R^e_{t+1,i} - b_{t,i}^\top\left(v^{\text{I-XS}}\omega^{\text{I-XS}\top}\right) f^{\text{IPCA}}_{t+1} = \hat{e}^{\text{I-XS}}_{t+1,i}$ maximize the XS-R^2. Similarly, we obtain $\omega^{\text{I-EV}}$ and $v^{\text{I-EV}}$ to maximize EV. If the correct one-factor representation has constant weights on the IPCA factors and their loadings, all three criteria would represent valid identification conditions to recover those weights. Note that in contrast to Kelly, Pruitt, and Su (2019) we estimate the total pricing errors and not only the component that is spanned by the characteristics. In the case of $\omega^{\text{I-EV}}$ all the weight will be put on the first IPCA factor which by construction maximizes the amount of explained variation.

The SDF loadings of the one-factor combination that maximizes the Sharpe ratio are not simply the same combination of the IPCA loadings, that is, in the above notation $\omega^{\text{I-SR}}$ does not need to equal $v^{\text{I-SR}}$. The internally consistent way to estimate the loadings of $F^{\text{I-SR}} = \omega^{\text{I-SR}} f^{\text{IPCA}}$ is to run the IPCA loading regression:

$$\Gamma_\beta = \left(\sum_{t=1}^{T-1} I_t I_t^\top \otimes \left(F^{\text{I-SR}}_{t+1}\right)^2\right)^{-1} \left(\sum_{t=1}^{T-1} \left(I_t \otimes F^{\text{I-SR}}_{t+1}\right)^\top R^e_{t+1}\right), \qquad \beta^{\text{I-SR}}_{t,i} = I_{i,t}^\top \Gamma_\beta,$$

where by abuse of notation $I_t \in \mathbb{R}^{q \times N_t}$ denotes here the firm-specific characteristics. This IPCA regression would simply return the IPCA loadings in a multi-factor model. In order to assess the effect of linearity between the characteristics and loadings, we also estimate the IPCA SDF loading $\beta^{\text{I-FFN}}_{i,t}$ with a feedforward neural network, that is, we estimate $\mathbb{E}_t\left[R^e_{t+1,i} F^{\text{I-SR}}_{t+1}\right]$. The combination of GAN and IPCA estimates $\omega^{\text{I-GAN}}$ using Formula IA.12 but restricting the SDF to a linear combination of IPCA factors based on Equation IA.13 and then estimates the loadings with FFN in the prediction $\beta^{\text{I-GAN}}_{i,t} = \mathbb{E}_t\left[R^e_{t+1,i} F^{\text{I-GAN}}_{t+1}\right]$.

Table IA.I summarizes the out-of-sample asset pricing results.[4] We can replicate the high Sharpe

[3] Kelly, Pruitt, and Su (2019) also allow for an error term in the loading equation. However, this does not affect the estimation procedure and our discussion, but only the confidence intervals. In order to simplify notation we leave it out.

[4] Additional results are in Section IA.B.2.

Table IA.I: IPCA Asset Pricing with Different SDFs

Model	Benchmark	3	4	5	6	7	8	9	10
IPCA GAN	SR	0.61	0.71	0.77	0.70	0.79	0.82	0.72	0.81
	EV	0.05	0.04	0.04	0.05	0.05	0.05	0.04	0.05
$(\omega^{\text{I-GAN}}, \beta^{\text{I-GAN}})$	XS-R^2	0.20	0.19	0.17	0.20	0.18	0.20	0.17	0.21
IPCA Max SR FFN Beta	SR	0.69	0.79	0.82	0.84	0.83	0.86	0.86	0.94
	EV	0.04	0.03	0.03	0.04	0.04	0.04	0.06	0.03
$(\omega^{\text{I-SR}}, \beta^{\text{I-FFN}})$	XS-R^2	0.14	0.13	0.11	0.14	0.14	0.15	0.19	0.14
IPCA Max SR	SR	0.69	0.79	0.82	0.84	0.83	0.86	0.86	0.94
	EV	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
$(\omega^{\text{I-SR}}, \beta^{\text{I-SR}})$	XS-R^2	-0.05	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04
IPCA Max EV	SR	0.11	0.11	0.15	0.17	0.15	0.15	0.14	0.16
	EV	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
$(\omega^{\text{I-EV}}, \beta^{\text{I-EV}})$	XS-R^2	-0.02	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03
IPCA Max XS-R^2	SR	-0.06	0.15	0.12	0.41	0.33	0.37	0.34	0.41
	EV	-0.02	-0.01	-0.02	-0.02	-0.02	-0.01	-0.02	-0.02
$(\omega^{\text{I-XS}}, \beta^{\text{I-XS}})$	XS-R^2	-0.03	0.07	0.06	0.12	0.12	0.13	0.13	0.14
IPCA Multifactor	SR	0.69	0.79	0.82	0.84	0.83	0.86	0.86	0.94
	EV	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.07
$(b_{t,i} \in \mathbb{R}^K)$	XS-R^2	-0.04	-0.03	-0.02	-0.01	-0.02	-0.01	-0.02	-0.02

This table shows the out-of-sample asset pricing results for different SDFs based on IPCA. We consider $K = 3$ to 10 IPCA factors. $(\omega^{\text{I-GAN}}, \beta^{\text{I-GAN}})$ uses the GAN framework to estimate the SDF weights of IPCA factors and the SDF loading. $(\omega^{\text{I-SR}}, \beta^{\text{I-FFN}})$ is the unconditional mean-variance efficient combination of IPCA factors with flexible SDF weights. $(\omega^{\text{I-SR}}, \beta^{\text{I-SR}})$ restricts those weights to be linear. $(\omega^{\text{I-XS}}, \beta^{\text{I-EV}})$ combines the IPCA factors to maximize EV while $(\omega^{\text{I-XS}}, \beta^{\text{I-XS}})$ maximizes XS-R^2. The multi-factor representation obtains the residuals with a cross-sectional regression on the multiple loadings. The SDF weights and loadings are estimated on the training data and tuning parameters are chosen optimally on the validation data set.

ratios of Kelly, Pruitt, and Su (2019) by forming the unconditional mean variance efficient combination of the IPCA factors. Note that IPCA factors use more information than our models in the main text which might explain the high Sharpe ratios. The IPCA regressions use the dependency structure in the characteristic variables over time which makes the factor weights and loadings at each point in time a function of all past characteristics. The last subtable shows the IPCA results in a multi-factor framework, that is, we use a cross-sectional regression with multiple loadings to obtain the residual. Similar to PCA methods, IPCA captures a large amount of the variation in returns, but not their means. In fact, the cross-sectional R^2 can become negative. If we construct a one-factor model that maximizes the Sharpe ratio, the explained variation of IPCA drops to around 1% without explaining more of the mean returns. The one factor that explains the most variation is simply the first IPCA factor. Similar to PCA methods, this first factor does not explain mean returns. Note that IPCA factors are not necessarily orthogonal to each other and hence estimating the factor that maximizes EV can lead to slightly different results when including more factors. As expected, the best combination for explaining mean returns leads to substantially higher XS-R^2, however at the cost of SR and EV. Estimating nonlinear loadings $\beta^{\text{I-FFN}}$ explains more variation and leads to smaller pricing errors. However, the best performing model among all three dimensions is IPCA combined with GAN. The Sharpe ratios are close to those of the mean-variance efficient combination and actually higher than our benchmark GAN for $K \geq 7$, while the XS-R^2 almost reaches the level of our benchmark GAN. The explained variation is among the highest of all models and only better for our fully flexible benchmark GAN. In summary, the GAN framework is complementary to multi-factor models and can optimally make use of the additional structure imposed on the SDF and the additional information incorporated in factors.

IA.B.2. Additional Results

Table IA.II: IPCA GAN

	3	4	5	6	7	8	9	10
SR Train	1.60	1.65	1.82	1.80	1.86	1.82	1.86	1.87
SR Valid	1.00	1.20	1.28	1.26	1.53	1.44	1.22	1.34
SR Test	0.61	0.71	0.77	0.70	0.79	0.82	0.72	0.81
EV Train	0.13	0.12	0.11	0.13	0.13	0.15	0.12	0.14
EV Valid	0.05	0.05	0.04	0.05	0.05	0.06	0.04	0.06
EV Test	0.05	0.04	0.04	0.05	0.05	0.05	0.04	0.05
XS-R^2 Train	0.03	0.02	0.01	0.02	0.01	0.01	0.01	0.02
XS-R^2 Valid	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
XS-R^2 Test	0.20	0.19	0.17	0.20	0.18	0.20	0.17	0.21

This table shows the asset pricing results for different SDFs based on IPCA. We consider $K = 3$ to 10 IPCA factors. $(\omega^{\text{I-GAN}}, \beta^{\text{I-GAN}})$ uses the GAN framework to estimate the SDF weights of IPCA factors and the SDF loading. $(\omega^{\text{I-SR}}, \beta^{\text{I-FFN}})$ is the unconditional mean-variance efficient combination of IPCA factors with flexible SDF weights. $(\omega^{\text{I-SR}}, \beta^{\text{I-SR}})$ restricts those weights to be linear. $(\omega^{\text{I-XS}}, \beta^{\text{I-EV}})$ combines the IPCA factors to maximize EV while $(\omega^{\text{I-XS}}, \beta^{\text{I-XS}})$ maximizes XS-R^2. The multi-factor representation obtains the residuals with a cross-sectional regression on the multiple loadings. The SDF weights and loadings are estimated on the training data and tuning parameters are chosen optimally on the validation data set.

Table IA.III: IPCA Max SR FFN Beta

	3	4	5	6	7	8	9	10
SR Train	1.59	1.67	1.78	1.82	1.85	1.88	1.88	1.95
SR Valid	1.12	1.27	1.37	1.42	1.42	1.42	1.45	1.50
SR Test	0.69	0.79	0.82	0.84	0.83	0.86	0.86	0.94
EV Train	0.10	0.09	0.08	0.10	0.10	0.10	0.16	0.10
EV Valid	0.05	0.04	0.03	0.04	0.04	0.04	0.07	0.04
EV Test	0.04	0.03	0.03	0.04	0.04	0.04	0.06	0.03
XS-R^2 Train	0.12	0.15	0.13	0.13	0.14	0.14	0.13	0.14
XS-R^2 Valid	-0.02	-0.01	-0.01	-0.01	-0.01	-0.01	0.00	-0.01
XS-R^2 Test	0.14	0.13	0.11	0.14	0.14	0.15	0.19	0.14

This table shows the asset pricing results for different SDFs based on IPCA. We consider $K = 3$ to 10 IPCA factors. $(\omega^{\text{I-GAN}}, \beta^{\text{I-GAN}})$ uses the GAN framework to estimate the SDF weights of IPCA factors and the SDF loading. $(\omega^{\text{I-SR}}, \beta^{\text{I-FFN}})$ is the unconditional mean-variance efficient combination of IPCA factors with flexible SDF weights. $(\omega^{\text{I-SR}}, \beta^{\text{I-SR}})$ restricts those weights to be linear. $(\omega^{\text{I-XS}}, \beta^{\text{I-EV}})$ combines the IPCA factors to maximize EV while $(\omega^{\text{I-XS}}, \beta^{\text{I-XS}})$ maximizes XS-R^2. The multi-factor representation obtains the residuals with a cross-sectional regression on the multiple loadings. The SDF weights and loadings are estimated on the training data and tuning parameters are chosen optimally on the validation data set.

Table IA.IV: IPCA Max SR

	3	4	5	6	7	8	9	10
SR Train	1.59	1.67	1.78	1.82	1.85	1.88	1.88	1.95
SR Valid	1.12	1.27	1.37	1.42	1.42	1.42	1.45	1.50
SR Test	0.69	0.79	0.82	0.84	0.83	0.86	0.86	0.94
EV Train	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
EV Valid	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
EV Test	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
XS-R^2 Train	-0.05	-0.04	-0.04	-0.03	-0.03	-0.03	-0.03	-0.04
XS-R^2 Valid	0.03	0.02	0.02	0.04	0.04	0.04	0.03	0.03
XS-R^2 Test	-0.05	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04

This table shows the asset pricing results for different SDFs based on IPCA. We consider $K = 3$ to 10 IPCA factors. $(\omega^{\text{I-GAN}}, \beta^{\text{I-GAN}})$ uses the GAN framework to estimate the SDF weights of IPCA factors and the SDF loading. $(\omega^{\text{I-SR}}, \beta^{\text{I-FFN}})$ is the unconditional mean-variance efficient combination of IPCA factors with flexible SDF weights. $(\omega^{\text{I-SR}}, \beta^{\text{I-SR}})$ restricts those weights to be linear. $(\omega^{\text{I-XS}}, \beta^{\text{I-EV}})$ combines the IPCA factors to maximize EV while $(\omega^{\text{I-XS}}, \beta^{\text{I-XS}})$ maximizes XS-R^2. The multi-factor representation obtains the residuals with a cross-sectional regression on the multiple loadings. The SDF weights and loadings are estimated on the training data and tuning parameters are chosen optimally on the validation data set.

Table IA.V: IPCA Max EV

	3	4	5	6	7	8	9	10
SR Train	0.03	0.05	0.14	0.19	0.10	0.11	0.09	0.10
SR Valid	-0.02	-0.01	0.07	0.12	0.05	0.06	0.02	0.07
SR Test	0.11	0.11	0.15	0.17	0.15	0.15	0.14	0.16
EV Train	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
EV Valid	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
EV Test	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
XS-R^2 Train	0.05	0.04	0.03	0.01	0.03	0.03	0.03	0.03
XS-R^2 Valid	0.00	0.00	-0.01	-0.02	-0.01	-0.01	-0.01	-0.01
XS-R^2 Test	-0.02	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03

This table shows the asset pricing results for different SDFs based on IPCA. We consider $K = 3$ to 10 IPCA factors. $(\omega^{\text{I-GAN}}, \beta^{\text{I-GAN}})$ uses the GAN framework to estimate the SDF weights of IPCA factors and the SDF loading. $(\omega^{\text{I-SR}}, \beta^{\text{I-FFN}})$ is the unconditional mean-variance efficient combination of IPCA factors with flexible SDF weights. $(\omega^{\text{I-SR}}, \beta^{\text{I-SR}})$ restricts those weights to be linear. $(\omega^{\text{I-XS}}, \beta^{\text{I-EV}})$ combines the IPCA factors to maximize EV while $(\omega^{\text{I-XS}}, \beta^{\text{I-XS}})$ maximizes XS-R^2. The multi-factor representation obtains the residuals with a cross-sectional regression on the multiple loadings. The SDF weights and loadings are estimated on the training data and tuning parameters are chosen optimally on the validation data set.

Table IA.VI: IPCA Max XS-R^2

	3	4	5	6	7	8	9	10
SR Train	0.05	0.56	0.47	1.10	1.03	1.07	0.99	0.95
SR Valid	0.00	0.34	0.25	0.70	0.63	0.65	0.64	0.69
SR Test	-0.06	0.15	0.12	0.41	0.33	0.37	0.34	0.41
EV Train	0.01	-0.01	-0.01	-0.04	-0.03	-0.04	-0.04	-0.05
EV Valid	-0.01	0.00	0.00	-0.02	-0.01	-0.02	-0.01	-0.02
EV Test	-0.02	-0.01	-0.02	-0.02	-0.02	-0.01	-0.02	-0.02
XS-R^2 Train	0.13	0.21	0.20	0.29	0.30	0.31	0.30	0.31
XS-R^2 Valid	-0.04	0.12	0.11	0.22	0.20	0.20	0.22	0.24
XS-R^2 Test	-0.03	0.07	0.06	0.12	0.12	0.13	0.13	0.14

This table shows the asset pricing results for different SDFs based on IPCA. We consider $K = 3$ to 10 IPCA factors. $(\omega^{\text{I-GAN}}, \beta^{\text{I-GAN}})$ uses the GAN framework to estimate the SDF weights of IPCA factors and the SDF loading. $(\omega^{\text{I-SR}}, \beta^{\text{I-FFN}})$ is the unconditional mean-variance efficient combination of IPCA factors with flexible SDF weights. $(\omega^{\text{I-SR}}, \beta^{\text{I-SR}})$ restricts those weights to be linear. $(\omega^{\text{I-XS}}, \beta^{\text{I-EV}})$ combines the IPCA factors to maximize EV while $(\omega^{\text{I-XS}}, \beta^{\text{I-XS}})$ maximizes XS-R^2. The multi-factor representation obtains the residuals with a cross-sectional regression on the multiple loadings. The SDF weights and loadings are estimated on the training data and tuning parameters are chosen optimally on the validation data set.

Table IA.VII: IPCA Multi-Factor

	3	4	5	6	7	8	9	10
SR Train	1.59	1.67	1.78	1.82	1.85	1.88	1.88	1.95
SR Valid	1.12	1.27	1.37	1.42	1.42	1.42	1.45	1.50
SR Test	0.69	0.79	0.82	0.84	0.83	0.86	0.86	0.94
EV Train	0.06	0.07	0.08	0.08	0.08	0.09	0.09	0.09
EV Valid	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.04
EV Test	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.07
XS-R^2 Train	0.04	0.04	0.05	0.06	0.06	0.06	0.07	0.07
XS-R^2 Valid	0.04	0.05	0.06	0.07	0.06	0.06	0.06	0.06
XS-R^2 Test	-0.04	-0.03	-0.02	-0.01	-0.02	-0.01	-0.02	-0.02

This table shows the asset pricing results for different SDFs based on IPCA. We consider $K = 3$ to 10 IPCA factors. $(\omega^{\text{I-GAN}}, \beta^{\text{I-GAN}})$ uses the GAN framework to estimate the SDF weights of IPCA factors and the SDF loading. $(\omega^{\text{I-SR}}, \beta^{\text{I-FFN}})$ is the unconditional mean-variance efficient combination of IPCA factors with flexible SDF weights. $(\omega^{\text{I-SR}}, \beta^{\text{I-SR}})$ restricts those weights to be linear. $(\omega^{\text{I-XS}}, \beta^{\text{I-EV}})$ combines the IPCA factors to maximize EV while $(\omega^{\text{I-XS}}, \beta^{\text{I-XS}})$ maximizes XS-R^2. The multi-factor representation obtains the residuals with a cross-sectional regression on the multiple loadings. The SDF weights and loadings are estimated on the training data and tuning parameters are chosen optimally on the validation data set.

IA.C. Robustness Results

IA.C.1. Main Results

Our findings are robust to small cap stocks, the choice of the tuning parameters, the time period under consideration and are not exploiting limits to arbitrage. In this subsection, we evaluate and refit

Table IA.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
Evaluated for 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
Evaluated for $\geq$ 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
Estimated and evaluated for size $\geq$ 0.001% of total market cap									
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

The table shows monthly Sharpe ratios (SR) of the SDF factors, explained time series variation (EV) and cross-sectional R^2 for the GAN, FFN, EN and LS models. In the first two subtables 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. In the last subtable the model is estimated and evaluated on stocks market with capitalization larger than 0.001%.

the model without small cap stocks, compare the performance and structure of the SDF for different tuning parameters and time periods and control the information used to construct the test assets.

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 IA.4 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, that is, the sample is close to the S&P 500 index.

Table IA.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 annual 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

Table IA.IX: Best Performing GAN Models on the Validation Data

Model	SMV	CSMV	HL	CHL	CHU	LR	SR (Train)	SR (Valid)	SR (Test)
GAN 1	4	32	4	0	32	0.001	2.78	1.47	0.72
GAN 2	4	32	2	0	8	0.001	3.02	1.39	0.77
GAN 3	4	32	4	0	16	0.0005	2.55	1.38	0.74
GAN 4	4	16	3	1	16	0.0005	2.44	1.38	0.77

This tables shows a re-estimation of the GAN model that is independent of the benchmark GAN model and reports the network structure. GAN 1 has 4 layers, 32 instruments and 4 hidden states. GAN 2 has 2 layers, 8 instruments and 4 hidden states and the hence the same architecture as our benchmark model. GAN 3 has 4 layers, 16 instruments and 4 hidden states, while GAN 4 has 3 layers, 16 instruments and 4 hidden states.

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 IA.VIII also estimates and evaluates the different models on stocks with market capitalization larger than 0.001% of the total market capitalization.[5] 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 full 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 full 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.

We re-estimate the GAN model independently of our benchmark fit and list the tuning parameters of the best four models on the validation data, labeled GAN 1, 2, 3 and 4 in Table IA.IX. All models have four macroeconomic states, but differ in terms of the depth of the network and the number of instruments that construct the test assets. The tuning parameters of our benchmark model would only be the second best model in this independent fit. Table IA.XIV reports the performance for the different fits and tuning parameters. The asset pricing performance is essentially identical for all models. Moreover, Table IA.XI shows that the SDFs for the various models all have a correlation higher than 80%. Section IA.C.5 collects the variable importance results and functional form of the SDF weights ω for the alternative fits. In summary, we conclude that not only the pricing performance is extremely robust to the tuning parameters, but we are actually discovering the same economic model for different tuning parameters and our results are replicable.

As another robustness test we estimate the GAN model on a rolling window of 240 months. In more detail, every year we move the training data window by one year to re-estimate the SDF weights

[5] 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.

Table IA.X: Performance of Alternative GAN Models

Model	SR Train	SR Valid	SR Test	EV Train	EV Valid	EV Test	Cross-Sectional R^2 Train	Cross-Sectional R^2 Valid	Cross-Sectional R^2 Test
GAN 1	2.78	1.47	0.72	0.18	0.08	0.07	0.12	0.01	0.21
GAN 2	3.02	1.39	0.77	0.18	0.08	0.07	0.12	0.00	0.22
GAN 3	2.55	1.38	0.74	0.22	0.11	0.09	0.17	0.04	0.25
GAN 4	2.44	1.38	0.77	0.19	0.08	0.07	0.11	0.01	0.22
GAN Rolling	N/A	N/A	0.88	N/A	N/A	0.08	N/A	N/A	0.24
GAN No Frict	2.94	1.37	0.77	0.20	0.10	0.08	0.14	0.01	0.23

This table shows the monthly Sharpe ratio (SR) of the SDF, explained time series variation (EV) and cross-sectional R^2 for alternative GAN models. GAN 1, 2, 3 and 4 are the four best GAN models on the validation data from an independent re-estimation of the model. GAN Rolling is re-estimated every year on a rolling window of 240 months. GAN No Frict is estimated without trading frictions and past returns for the conditioning function g.

Table IA.XI: Correlation of Benchmark GAN SDF with SDF of Alternative GAN Estimations

	GAN	GAN 1	GAN 2	GAN 3	GAN 4	GAN Rolling	GAN No Frict
GAN	1	0.84	0.87	0.84	0.80	0.70	0.78
GAN 1	0.84	1	0.88	0.92	0.89	0.79	0.89
GAN 2	0.87	0.88	1	0.87	0.88	0.73	0.83
GAN 3	0.84	0.92	0.87	1	0.89	0.74	0.86
GAN 4	0.80	0.89	0.88	0.89	1	0.78	0.84
GAN Rolling	0.70	0.79	0.73	0.74	0.78	1	0.78
GAN No Frict	0.78	0.89	0.83	0.86	0.84	0.78	1

This table reports the factor correlations for alternative estimations of GAN. GAN is the benchmark model, GAN 1, 2, 3 and 4 are the top four performing models from a re-estimation. GAN Rolling is estimated on a rolling window of 240 months. GAN No Frict is an estimation of the GAN model without trading friction variables and past returns for the conditioning instruments g.

and loadings.[6] Not surprisingly, we obtain slightly better asset pricing results, in particular for the Sharpe ratio, as reported in Table IA.XIV. Overall, the results are very close and the rolling window GAN SDF has a correlation of 70% with our benchmark SDF. Figures IA.1 and IA.2 show the variable importance and functional form of the rolling window SDF which share the same general patterns. We conclude that a time-varying estimate of GAN does not lead to major improvements and fits a similar economic structure. In particular, this confirms that our results are robust to the time window chosen for estimation.

Our GAN is not simply capturing pricing information that is subject to limits to arbitrage. A potential concern is that GAN constructs test assets that do not represent a risk-premium but anomalies of illiquid stocks that cannot be exploited. In that case the GAN SDF would not represent the economic risk that we want to capture for asset pricing. We want to avoid that GAN explicitly targets stocks that have high trading frictions. We exclude characteristics from the trading frictions and past return category from the conditional network and re-estimate the model labeled as GAN No Frict. The resulting variable importance for g is shown in Figure IA.3. This SDF has the same pricing

[6] The estimation of the GAN models is computationally very expensive and for this reason we are not re-estimating it every month.

performance and a correlation of 78% with our benchmark SDF suggesting that our results are robust to this change.

Figure IA.1. Characteristic Importance for Rolling Window GAN

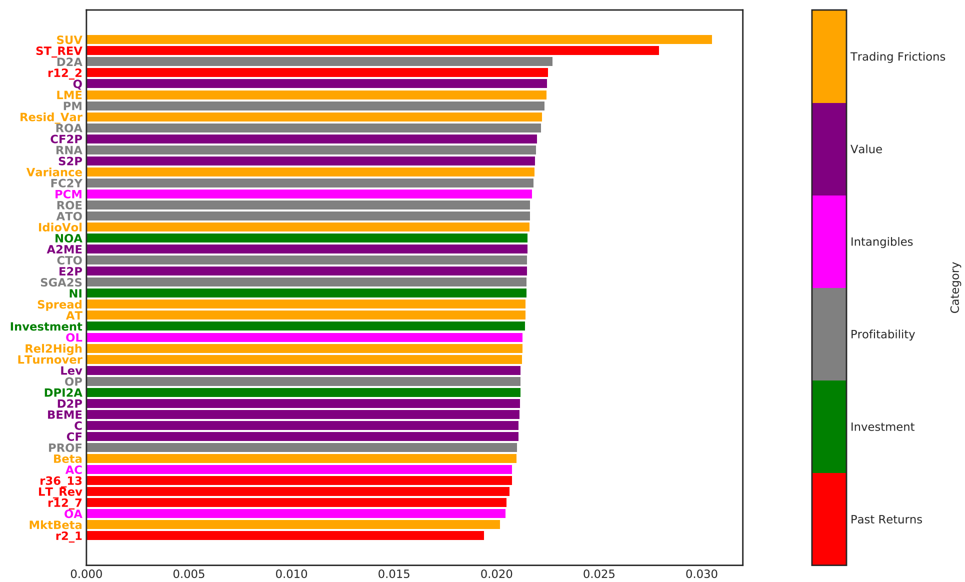
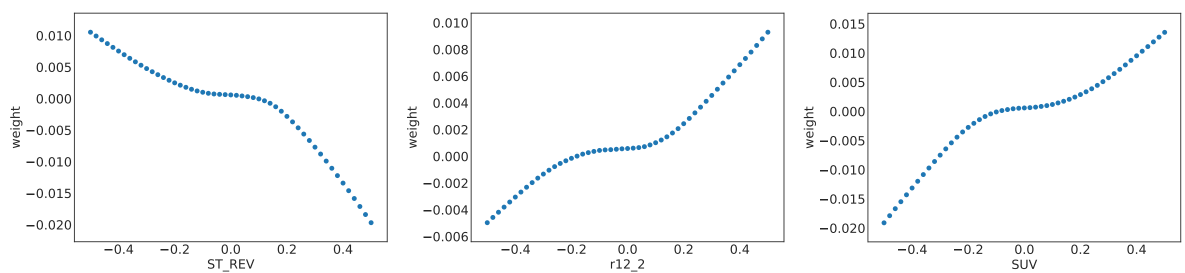
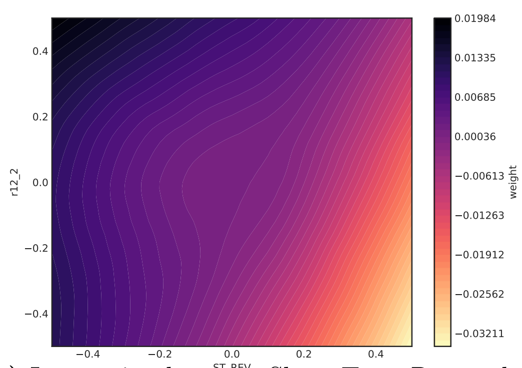


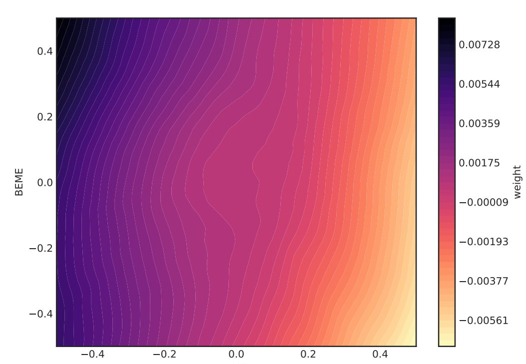
Figure IA.2: SDF weight ω as a Function of Covariates for Rolling Window Fit



(a) SDF weight ω as a function of one characteristic



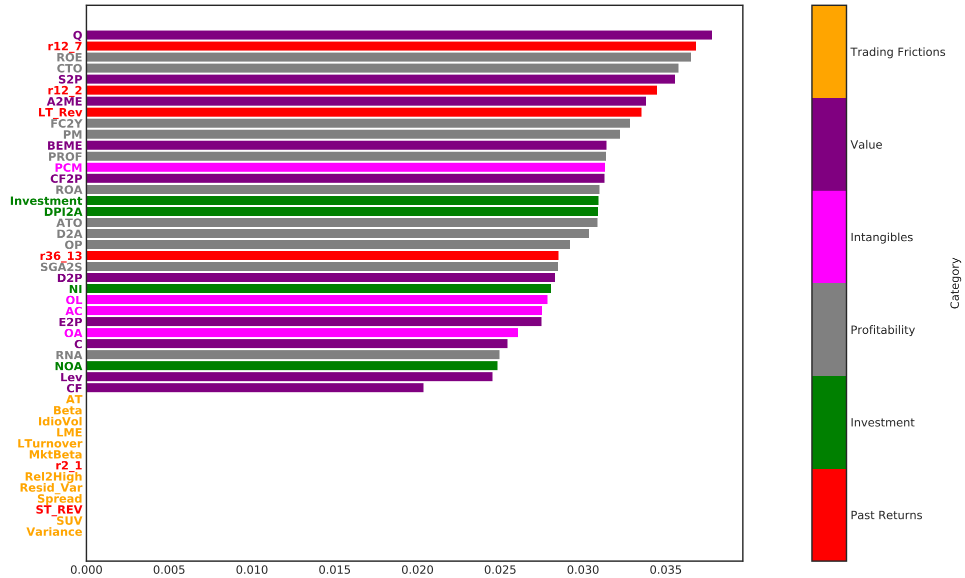
(b) Interaction between Short-Term Reversal (ST_REV) and Momentum (r12_2)



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

These figures show the variable importance and functional form of the SDF estimated on a rolling window of 240 months. The sensitivities and SDF weights ω are the average over those rolling window estimates.

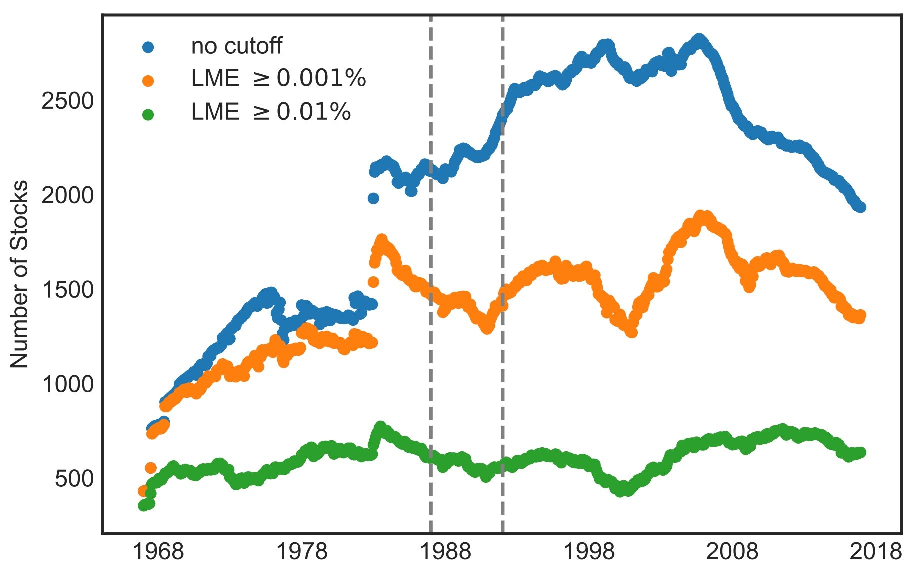
Figure IA.3: Characteristic Importance for Conditioning Function g for GAN No Frict



The figure shows the GAN variable importance ranking for the conditioning function g of the 46 firm-specific characteristics in terms of average absolute gradient on the test data. The ranking is the average of the absolute gradients for the nine ensemble fits. This GAN model excludes trading frictions and past returns in the construction of the conditioning function g. The values are normalized to sum up to one.

IA.C.2. Market Capitalization

Figure IA.4: Number of Stocks per Month



This plots shows the 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.

IA.C.3. *Missing Data*

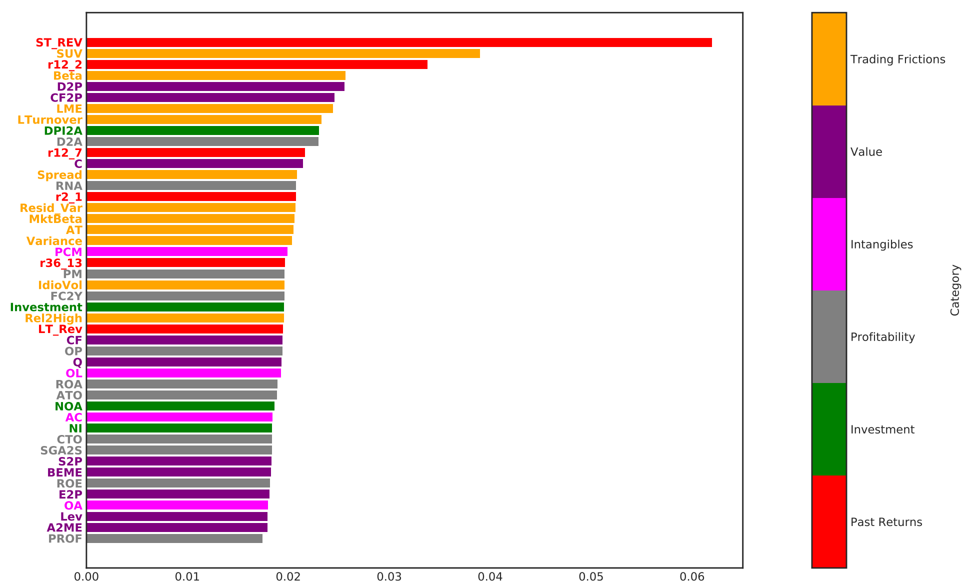
Table IA.XII: Performance of Different SDF Models on Imputed Data

	SR			EV			XS-R^2		
Model	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
LS	0.57	0.26	0.29	0.07	0.03	0.02	0.15	-0.02	0.07
EN	0.68	0.56	0.23	0.09	0.04	0.04	0.17	-0.02	0.11
FFN	0.44	0.43	0.46	0.08	0.03	0.02	0.13	-0.04	0.08
GAN	1.85	1.32	0.55	0.14	0.06	0.06	0.16	-0.04	0.13

This table shows the monthly Sharpe ratio (SR) of the SDF, explained time series variation (EV) and cross-sectional mean R^2 for the GAN, FFN, EN and LS model using a larger data set with imputed values. We imputed missing characteristics with their cross-sectional median. The asset pricing models are estimated on the data with fully observed data and evaluated on the stocks that include imputed missing characteristics. The fully observed data used in the main text has 336,113/132,167/750,275 asset-month observations and 3,686/3,347/7,141 different stocks on the training/validation/test data. The imputed data has 791,415/294,770/1,642,043 asset-month observations and 8,167/7,287/16,983 different assets. We exclude stocks with a market capitalization below 0.0001% of the total market capitalization.

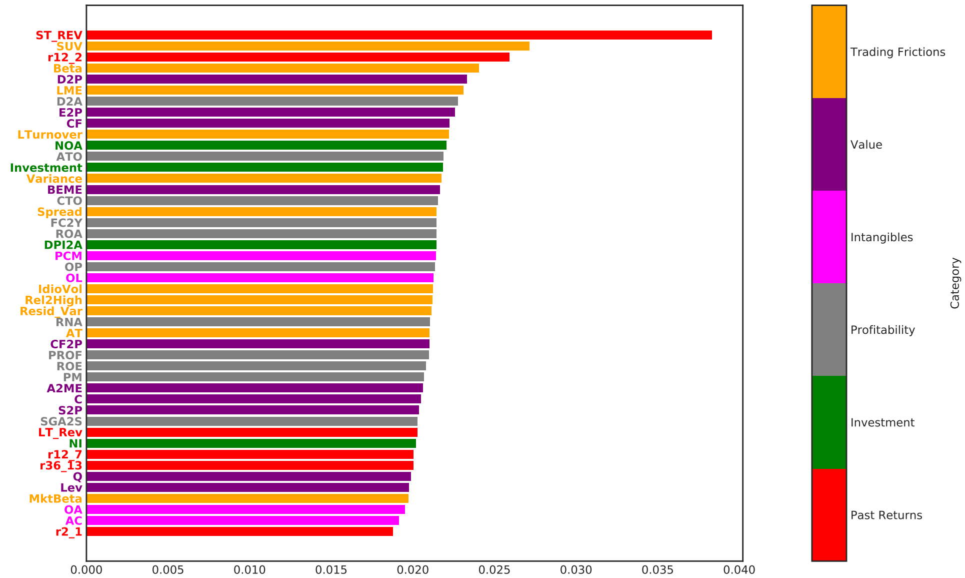
IA.C.4. *Characteristic Importance for Alternative GAN Models*

Figure IA.5: Characteristic Importance of SDF for GAN 1



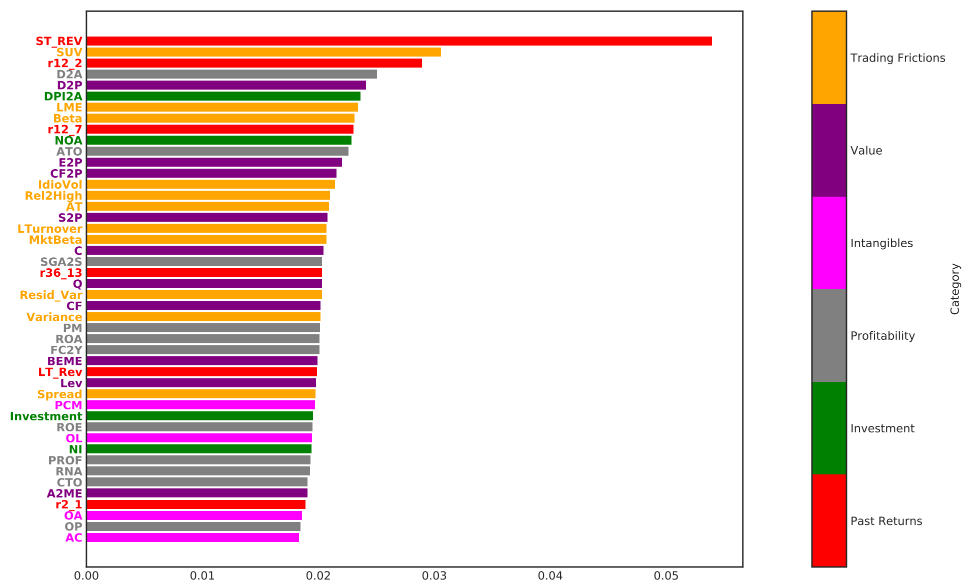
The figure shows the GAN 1 SDF variable importance ranking of the 46 firm-specific characteristics in terms of average absolute gradient (VI) on the test data. The values are normalized to sum up to one.

Figure IA.6: Characteristic Importance of SDF for GAN 2



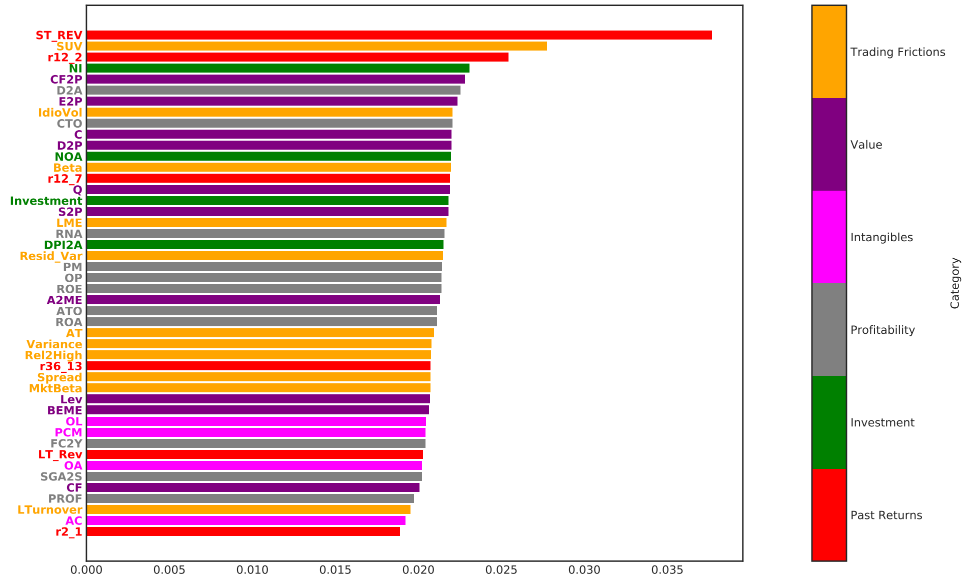
The figure shows the GAN 2 SDF variable importance ranking of the 46 firm-specific characteristics in terms of average absolute gradient (VI) on the test data. The values are normalized to sum up to one.

Figure IA.7: Characteristic Importance of SDF for GAN 3



The figure shows the GAN 3 SDF variable importance ranking of the 46 firm-specific characteristics in terms of average absolute gradient (VI) on the test data. The values are normalized to sum up to one.

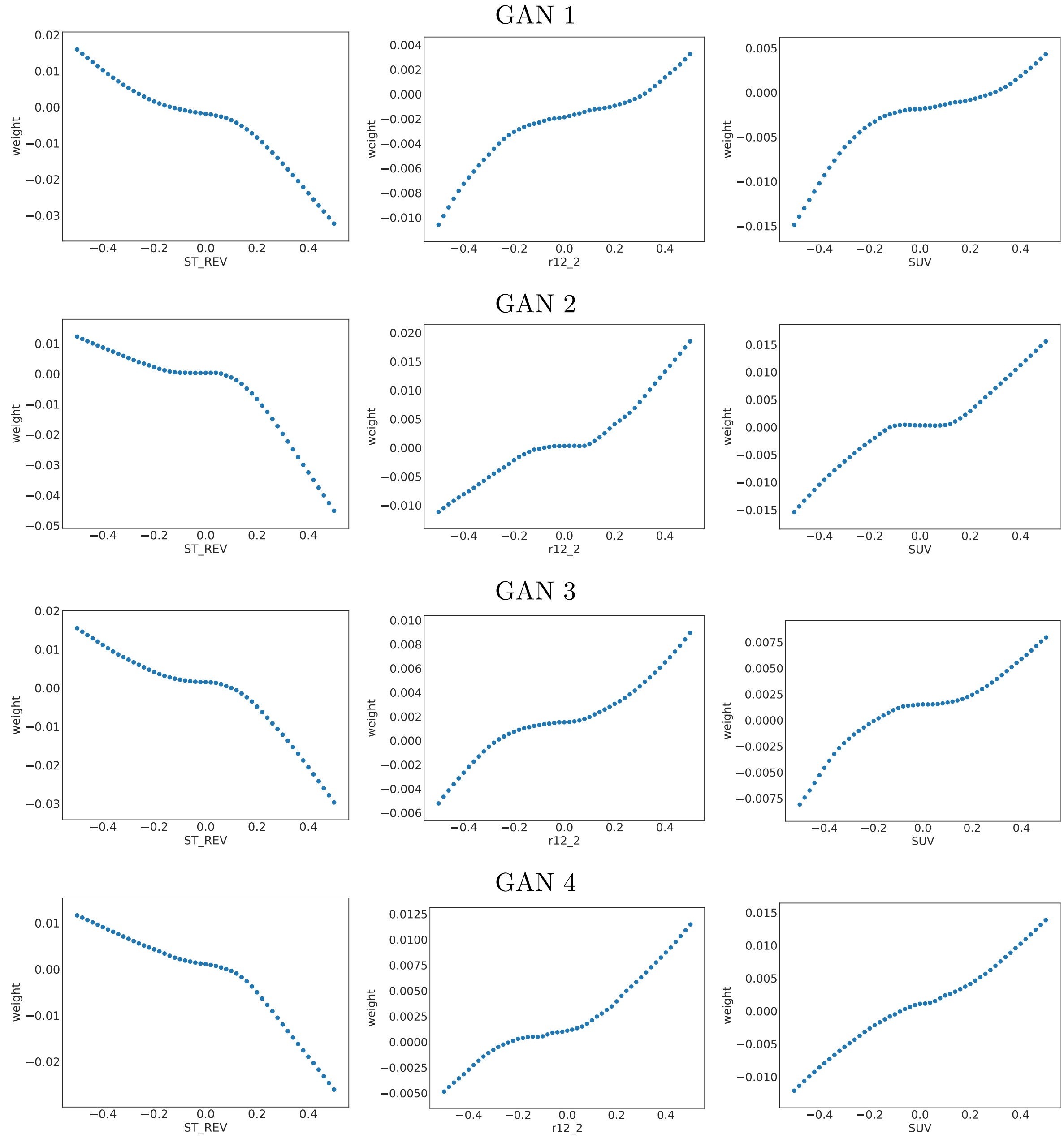
Figure IA.8: Characteristic Importance of SDF for GAN 4



The figure shows the GAN 4 SDF variable importance ranking of the 46 firm-specific characteristics in terms of average absolute gradient (VI) on the test data. The values are normalized to sum up to one.

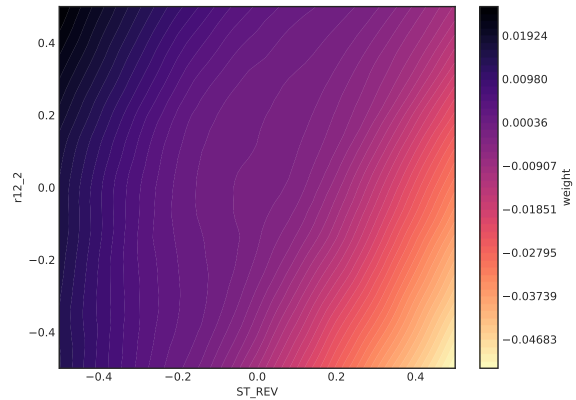
IA.C.5. SDF Structure for Alternative GAN Models

Figure IA.9: SDF weight ω as a Function of Characteristics for GAN 1-4

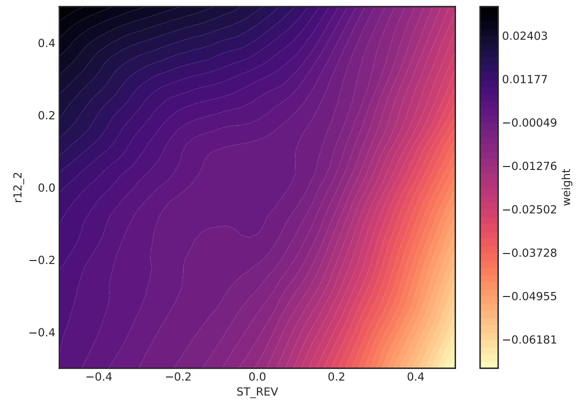


This figure shows the SDF weight ω as a one-dimensional function of characteristics keeping the other covariates at their mean level for alternative GAN models. The covariates are Short-Term Reversal (`ST_REV`), Momentum (`r12_2`) and Standard Unexplained Volume (`SUV`). The re-estimation of the GAN model is independent of the benchmark GAN model. GAN 1 has 4 layers, 32 instruments and 4 hidden states. GAN 2 has 2 layers, 8 instruments and 4 hidden states and the hence the same architecture as our benchmark model. GAN 3 has 4 layers, 16 instruments and 4 hidden states, while GAN 4 has 3 layers, 16 instruments and 4 hidden states.

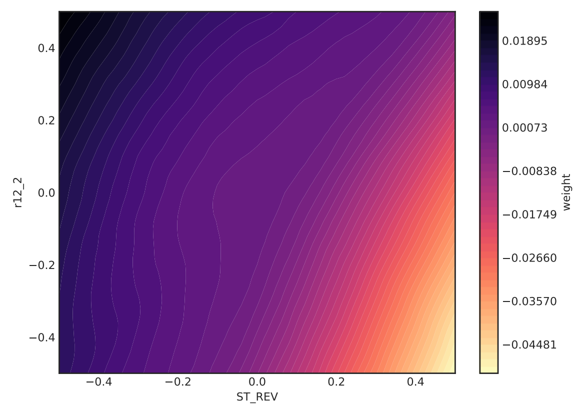
Figure IA.10: SDF weight ω as a Function of Characteristics for GAN



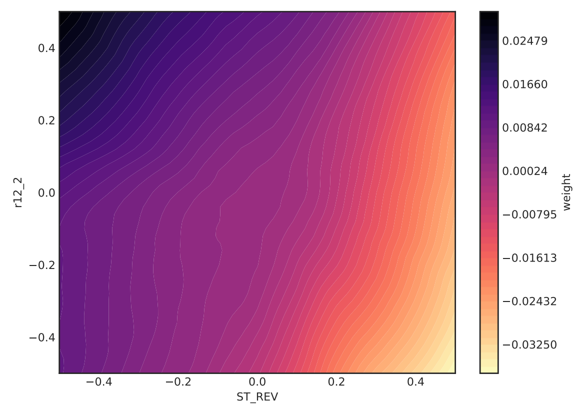
(a) Interaction between Short-Term Reversal (ST_REV) and Momentum (r12_2) for GAN 1



(b) Interaction between Short-Term Reversal (ST_REV) and Momentum (r12_2) for GAN 2



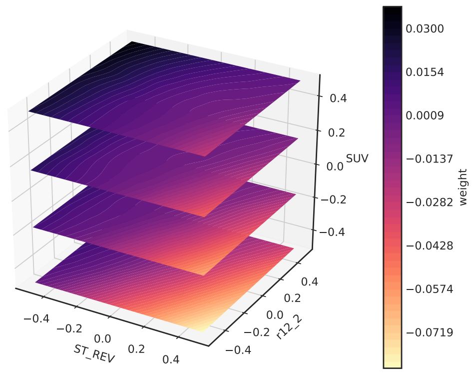
(c) Interaction between Short-Term Reversal (ST_REV) and Momentum (r12_2) for GAN 3



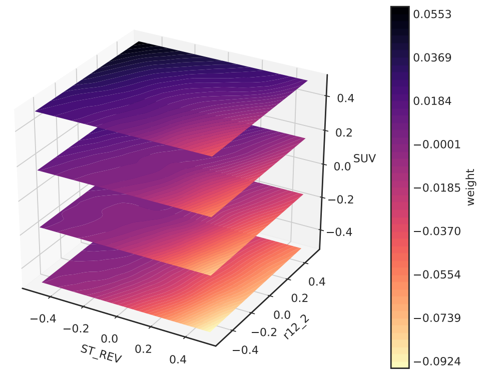
(d) Interaction between Short-Term Reversal (ST_REV) and Momentum (r12_2) for GAN 4

These figures show the SDF weight ω for alternative GAN models as two-dimensional function of characteristics keeping the remaining variables at their mean level. The re-estimation of the GAN model is independent of the benchmark GAN model. GAN 1 has 4 layers, 32 instruments and 4 hidden states. GAN 2 has 2 layers, 8 instruments and 4 hidden states and the hence the same architecture as our benchmark model. GAN 3 has 4 layers, 16 instruments and 4 hidden states, while GAN 4 has 3 layers, 16 instruments and 4 hidden states.

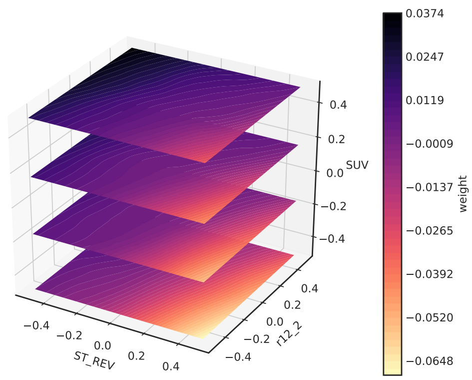
Figure IA.11: SDF weight ω as a Function of Characteristics for GAN



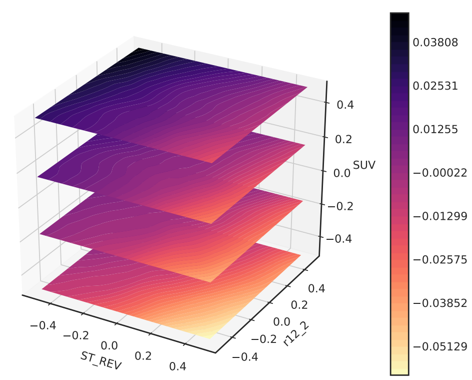
(a) Interaction between Short-Term Reversal (ST_REV), Momentum (r12_2) and Standard Unexplained Volume (SUV) for GAN 1



(b) IInteraction between Short-Term Reversal (ST_REV), Momentum (r12_2) and Standard Unexplained Volume (SUV) for GAN 2



(c) Interaction between Short-Term Reversal (ST_REV), Momentum (r12_2) and Standard Unexplained Volume (SUV) for GAN 3



(d) Interaction between Short-Term Reversal (ST_REV), Momentum (r12_2) and Standard Unexplained Volume (SUV) for GAN 4

These figures show the SDF weight ω for alternative GAN models as three-dimensional function of characteristics keeping the remaining variables at their mean level. The re-estimation of the GAN model is independent of the benchmark GAN model. GAN 1 has 4 layers, 32 instruments and 4 hidden states. GAN 2 has 2 layers, 8 instruments and 4 hidden states and the hence the same architecture as our benchmark model. GAN 3 has 4 layers, 16 instruments and 4 hidden states, while GAN 4 has 3 layers, 16 instruments and 4 hidden states.

IA.D. 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 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.[7]

Excess returns follow a no-arbitrage model $R^e_{t+1,i} = \beta_{t,i}F_{t+1} + \epsilon_{t+1,i}$. In our simple model the SDF 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 cyclical 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 with trend $Z_t = \mu_M t + h_t$. 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 and summarized in Table IA.XIII. 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 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 IA.12 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 IA.XIII 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 reaches the same value as the population SDF 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 neither the SDF nor the correct loading structure. Note, that the simple forecasting approach can generate a high Sharpe Ratio but fails in explaining the systematic component.

Figure IA.12 explains why we observe the above performance results. Note, that the SDF 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

[7] 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.

Table IA.XIII: 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

This table reports the Sharpe Ratio (SR) of the SDF, 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 are generated with an SDF with Sharpe Ratio $SR = 1$ and $\sigma_F^2 = 0.1$ and the idiosyncratic noise has $\sigma_e^2 = 1$. The macroeconomic time-series has the trend $\mu_M = 0.05$. The number of observations is $N = 500, T = 600$, $T_{train} = 250$, $T_{valid} = 100$ and $T_{test} = 250$.

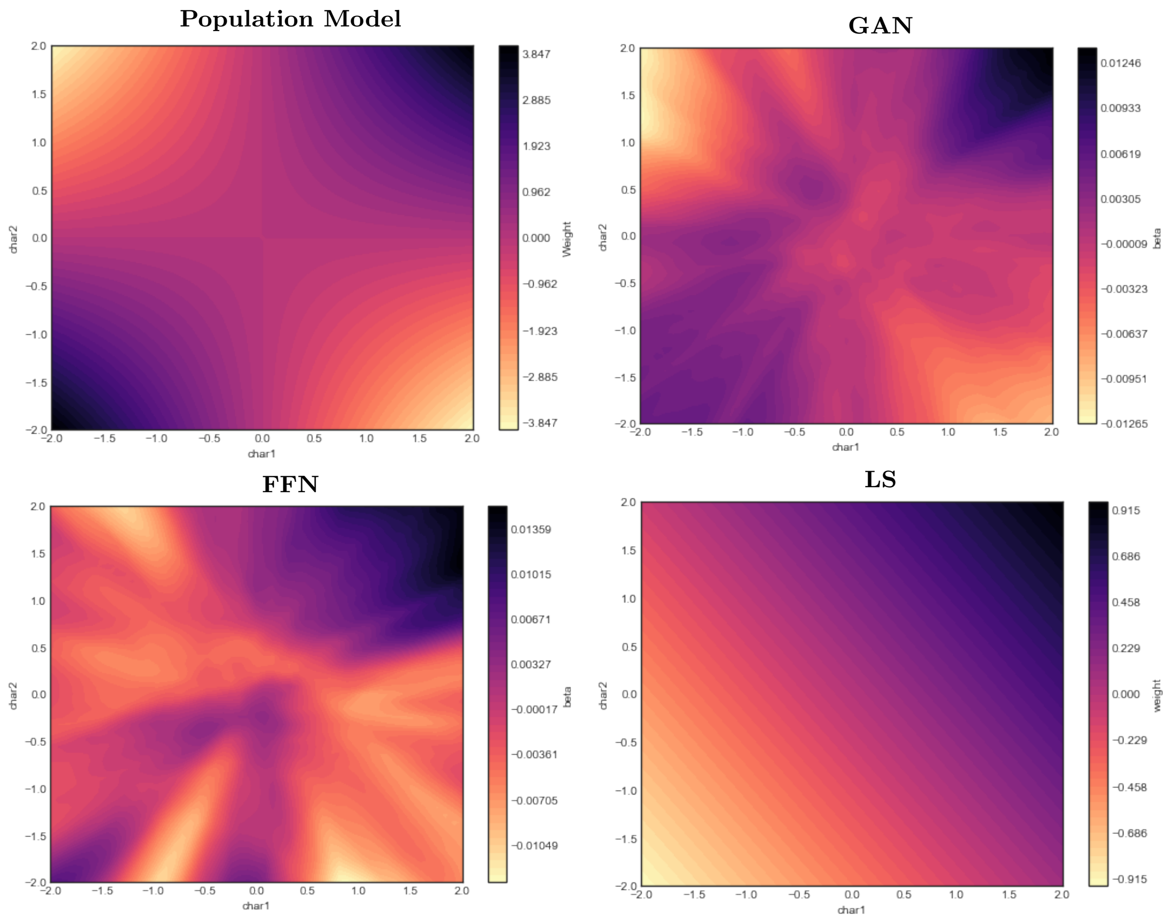
for low/low combinations. The FFN SDF still loads mainly on the extreme portfolios which results in the high Sharpe Ratio. However, the FFN fails to capture the loadings correctly which leads to high unexplained variation and pricing errors. The linear model can obviously not capture the non-linear interaction.

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 affects the assets differently during a boom and recession cycle. Note that in our example the 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 only depends on firm-specific information.

Figure IA.13 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 IA.13 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 only the most recent first differences can by construction not infer any dynamics in the macroeconomic variables.

Table IA.XIII 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

Figure IA.12: SDF Weights ω for the First Model with 2 Characteristics

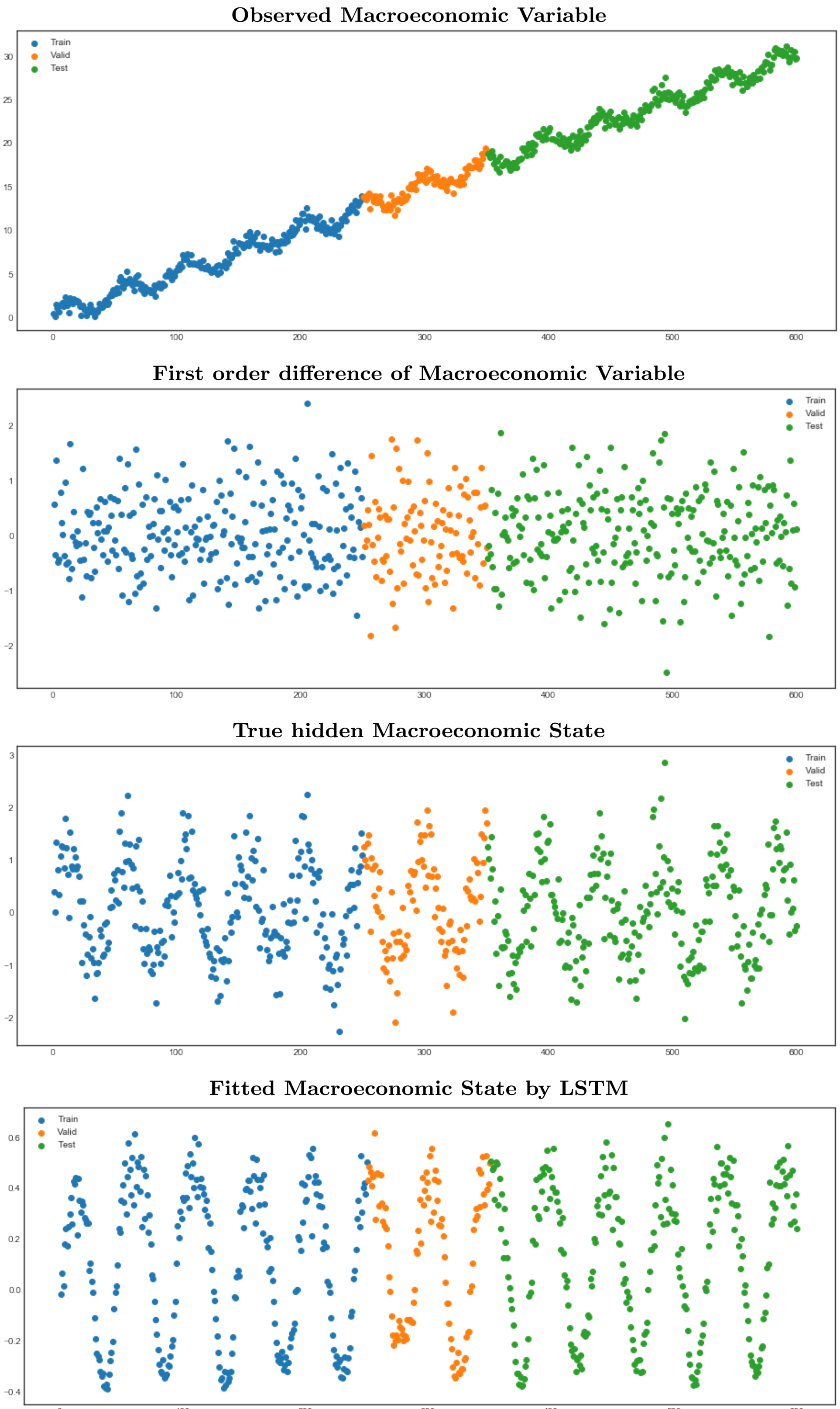


This figures shows the SDF weights ω as the function of the two characteristics estimated by different methods. Note that in our simple simulation the SDF weights ω coincide with the SDF loadings β.

roughly half of the times it uses the wrong sign for the SDF weights.

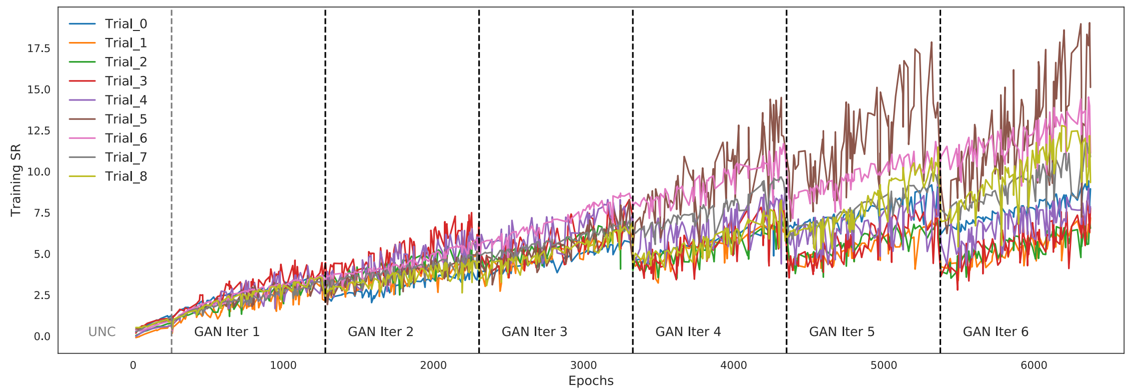
The simulation section illustrates three findings: (1) All three evaluation metrics (SR, EV and XS-R2) are necessary to assess the quality of the SDF factor. (2) By conditioning only on the most recent macroeconomic observations, general macroeconomic dynamics are ruled out. (3) The no-arbitrage condition in the GAN model helps to deal with a low signal-to-noise ratio.

Figure IA.13. Dynamics of the Macroeconomic State Variable

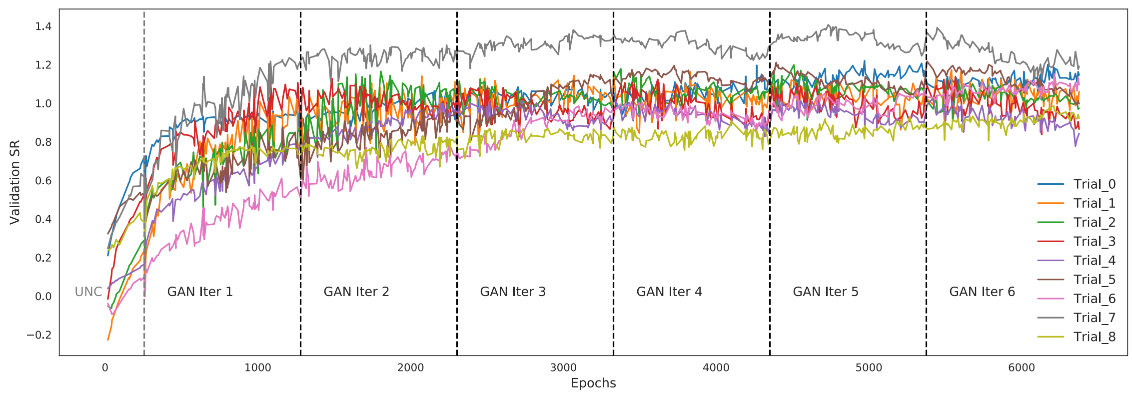


IA.E. Empirical Implementation

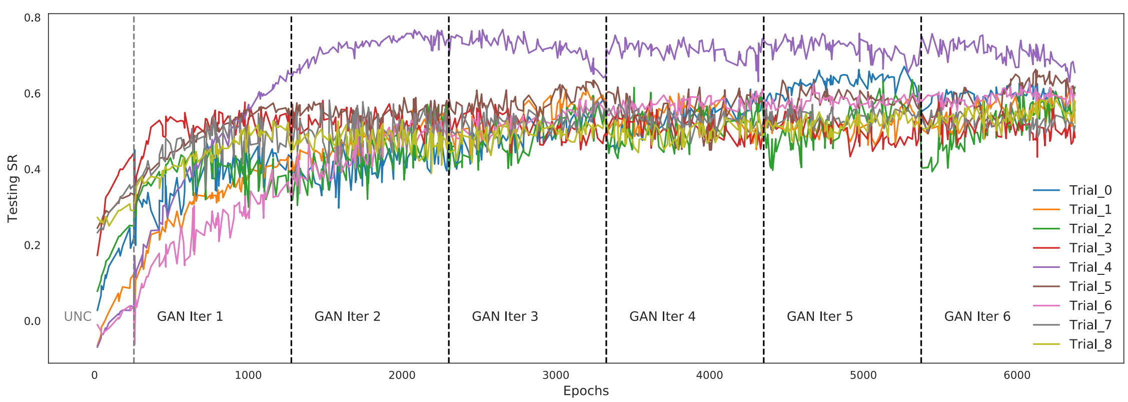
Figure IA.14: Evaluation of GAN for Different Number of GAN Iterations



(a) Training



(b) Validation



(c) Test

This figure shows the Sharpe ratio on the training, validation and test data for different number of iterations of the GAN fit for the benchmark model in our empirical analysis. The number of epochs is the number of complete passes through the training data set. The first 256 epochs are for the first stage of the unconditional GAN fit, that is, this represents the UNC fit. This is followed by 64 epochs for constructing the conditioning function g which are not counted in these plots. The next 1024 epochs fit the GAN model for the first set of moment conditions given by g. Hence, after the first 256 epochs, every 1024 epochs represent one iteration of the GAN fit. Our ensemble fit is based on 9 independent estimations of the model which are all included in the plots. Subfigure (b) shows the Sharpe ratio on the validation data. The benchmark model in the paper uses one GAN iteration, that is, we use the model represented by the second vertical line. The third vertical line corresponds to two GAN iterations which does not improve the Sharpe ratio. In contrast, the unconditional model UNC represented by the first vertical line is strictly inferior to the GAN fit.

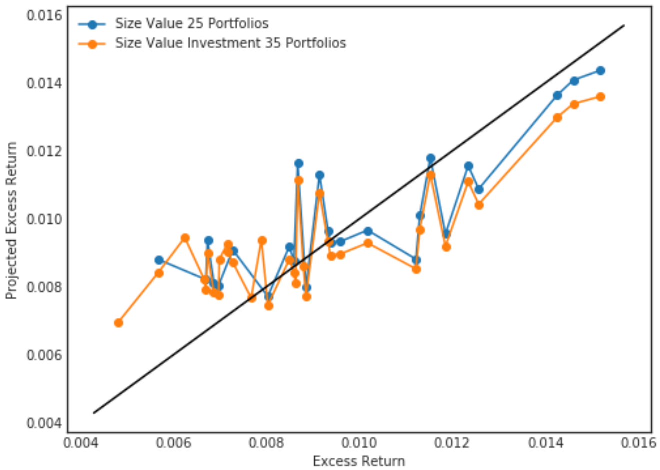
IA.F. Illustrative GAN Example

Table IA.XIV: Performance of Different SDF Models for Portfolio Sorts based on Size, Value and Investment

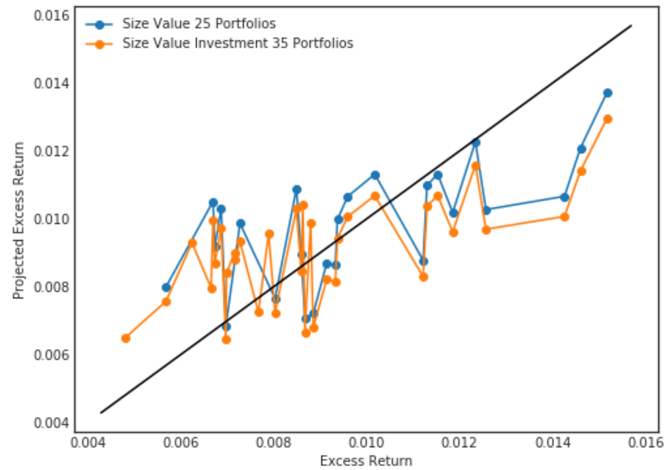
	SR			EV			Cross-Sectional R^2		
Model	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
	25 Double-Sorted Size and Value Portfolios								
UNC (SV)	0.22	0.42	0.27	0.50	0.60	0.48	0.50	0.83	0.60
UNC (SVI)	0.30	0.21	0.28	0.64	0.71	0.60	0.77	0.83	0.82
GAN (SV-SV)	0.32	0.40	0.38	0.61	0.64	0.55	0.88	0.76	0.84
GAN (SVI-SV)	0.50	0.58	0.43	0.78	0.80	0.72	0.91	0.73	0.96
GAN (SVI-SVI)	0.50	0.57	0.51	0.81	0.82	0.75	0.90	0.69	0.98
GAN	2.68	1.43	0.75	0.80	0.80	0.76	0.92	0.74	0.98
	10 Decile Investment Portfolios + 25 Double-Sorted Size and Value Portfolios								
UNC (SV)	0.22	0.42	0.27	0.52	0.64	0.48	0.45	0.88	0.60
UNC (SVI)	0.30	0.21	0.28	0.65	0.74	0.60	0.72	0.88	0.82
GAN (SV-SV)	0.32	0.40	0.38	0.63	0.68	0.57	0.85	0.80	0.86
GAN (SVI-SV)	0.50	0.58	0.43	0.76	0.80	0.70	0.87	0.77	0.96
GAN (SVI-SVI)	0.50	0.57	0.51	0.79	0.81	0.72	0.88	0.73	0.98
GAN	2.68	1.43	0.75	0.79	0.80	0.72	0.91	0.77	0.97

This table shows the monthly Sharpe ratio (SR), explained time-series variation (EV) and cross-sectional R^2 for different SDF models in the illustrative GAN example. UNC (SV) and UNC (SVI) are unconditional models with respect to the test assets, that is, they use only size and value respectively size, value and investment for the SDF weights, but set g to a constant. The GAN models use a non-trivial g. GAN (SVI-SV) allows the SDF weight ω to depend on size, value and investment, but the test asset function g to depend only on size and investment. The model labeled as GAN is our benchmark model estimated with all characteristics and macroeconomic information. We evaluate the model on 25 double-sorted size and book-to-market portfolios (SV 25) and we add another 10 decile portfolios sorted on investment (SVI 35). The portfolios are value-weighted.

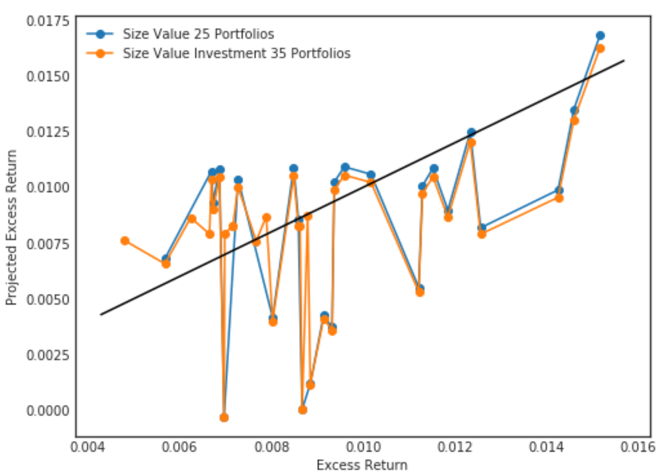
Figure IA.15: GAN Portfolio Pricing



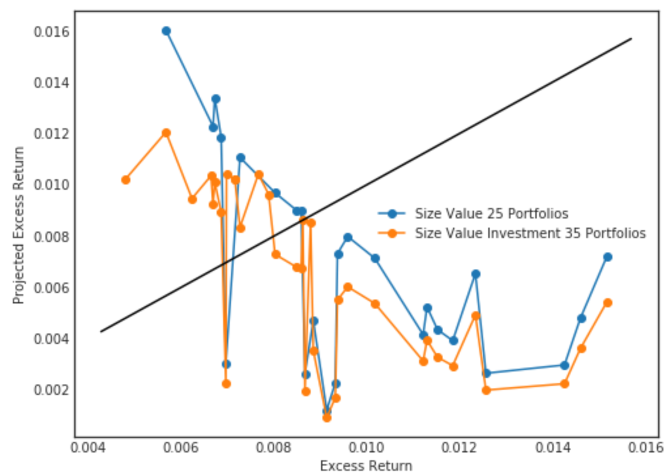
(a) GAN



(b) GAN (SVI-SV)



(c) GAN (SV-SV)



(d) UNC (SV)

This figure shows the predicted and average excess returns for the 25 and 35 sorted portfolios. The tuning parameters are chosen optimally on the validation data set and are different from the general benchmark GAN. UNC (SV) and UNC (SVI) are unconditional models with respect to the test assets, that is, they use only size and value respectively size, value and investment for the SDF weights, but set g to a constant. The GAN models use a non-trivial g. GAN (SVI-SV) allows the SDF weight ω to depend on size, value and investment, but the test asset function g to depend only on size and investment. The model labeled as GAN is our benchmark model estimated with all characteristics and macroeconomic information. We evaluate the model on 25 double-sorted size and book-to-market portfolios (SV 25) and we add another 10 decile portfolios sorted on investment (SVI 35). The portfolios are value-weighted.

Table IA.XV: Explained Variation and Pricing Errors for Double-Sorted Portfolios based on Size and Value

		UNC		GAN			GAN Full	UNC		GAN			GAN Full
		SV	SVI	SV-SV	SVI-SV	SVI-SVI	GAN	SV	SVI	SV-SV	SVI-SV	SVI-SVI	GAN
LME	BEME	Explained Variation						Alpha					
1	1	0.19	0.15	-0.00	0.45	0.52	0.63	0.12	0.13	0.17	0.03	0.00	-0.06
1	2	0.12	0.32	0.34	0.62	0.67	0.74	0.16	0.11	0.10	0.01	-0.01	-0.04
1	3	0.32	0.55	0.68	0.73	0.75	0.76	0.22	0.16	0.09	0.07	0.05	0.01
1	4	0.53	0.73	0.74	0.78	0.78	0.75	0.19	0.12	0.02	0.05	0.04	0.01
1	5	0.68	0.76	0.59	0.73	0.71	0.76	0.16	0.09	-0.03	0.03	0.02	0.02
2	1	0.31	0.33	-0.03	0.60	0.70	0.68	0.08	0.07	0.14	0.00	-0.03	-0.02
2	2	0.30	0.54	0.42	0.80	0.84	0.81	0.14	0.09	0.11	0.01	-0.01	-0.01
2	3	0.37	0.76	0.79	0.88	0.89	0.86	0.20	0.10	0.09	0.05	0.04	0.03
2	4	0.56	0.83	0.87	0.88	0.88	0.85	0.14	0.05	0.01	0.00	0.00	-0.01
2	5	0.72	0.86	0.88	0.85	0.85	0.85	0.11	0.03	-0.00	0.00	0.00	0.02
3	1	0.51	0.54	0.14	0.69	0.77	0.73	0.08	0.07	0.15	0.03	0.00	0.02
3	2	0.55	0.76	0.62	0.85	0.88	0.85	0.14	0.09	0.11	0.05	0.03	0.05
3	3	0.57	0.88	0.88	0.91	0.91	0.90	0.16	0.07	0.06	0.03	0.03	0.05
3	4	0.68	0.90	0.90	0.89	0.89	0.88	0.12	0.03	0.02	0.01	0.01	0.02
3	5	0.78	0.86	0.86	0.83	0.82	0.81	0.06	-0.02	-0.01	-0.02	-0.02	0.01
4	1	0.85	0.77	0.49	0.76	0.80	0.79	-0.03	0.00	0.08	0.01	-0.01	0.01
4	2	0.90	0.88	0.84	0.85	0.85	0.85	-0.01	-0.01	0.00	-0.01	-0.01	-0.00
4	3	0.86	0.88	0.87	0.83	0.82	0.83	0.04	-0.01	-0.02	-0.01	-0.00	0.00
4	4	0.87	0.89	0.89	0.85	0.85	0.85	0.03	-0.02	-0.03	-0.02	-0.01	0.01
4	5	0.84	0.77	0.83	0.78	0.78	0.79	-0.01	-0.07	-0.05	-0.05	-0.03	-0.01
5	1	-0.94	0.33	0.57	0.50	0.50	0.44	-0.20	-0.09	-0.02	-0.05	-0.04	-0.06
5	2	0.06	0.47	0.67	0.60	0.64	0.58	-0.13	-0.09	-0.05	-0.05	-0.03	-0.05
5	3	0.53	0.45	0.62	0.56	0.63	0.57	-0.07	-0.09	-0.06	-0.05	-0.03	-0.04
5	4	0.61	0.56	0.69	0.64	0.66	0.67	-0.10	-0.11	-0.08	-0.07	-0.04	-0.02
5	5	0.47	0.29	0.58	0.52	0.56	0.56	-0.11	-0.14	-0.08	-0.08	-0.05	-0.03
		Explained Variation						Cross-Sectional R^2					
All		0.48	0.60	0.55	0.72	0.75	0.76	0.60	0.82	0.84	0.96	0.98	0.98

This table shows the out-of-sample explained variation and pricing errors for double sorted portfolios based on size (`LME`) and book-to-market ratio (`BEME`). UNC (SV) and UNC (SVI) are unconditional models with respect to the test assets, that is, they use only size and value respectively size, value and investment for the SDF weights, but set g to a constant. GAN (SVI-SV) allows the SDF weight ω to depend on size, value and investment, but the test asset function g to depend only on size and investment. The model labeled as GAN is our benchmark model estimated with all characteristics and macroeconomic information. The portfolios are value-weighted.

Table IA.XVI: Explained Variation and Pricing Errors for Double-Sorted Portfolios based on Size and Value and Decile Sorted Portfolios based on Investment

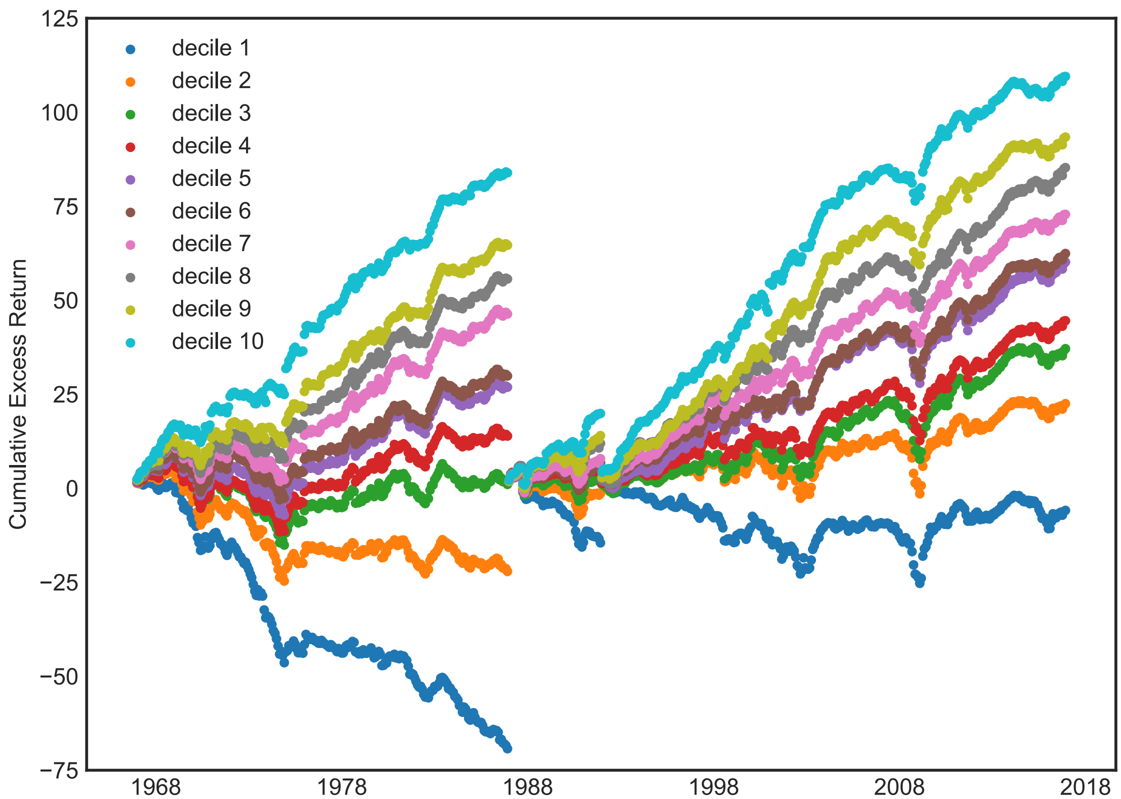
		UNC		GAN			GAN Full	UNC		GAN			GAN Full
		SV	SVI	SV-SV	SVI-SV	SVI-SVI	GAN	SV	SVI	SV-SV	SVI-SV	SVI-SVI	GAN
Investment		Explained Variation						Alpha					
1		0.66	0.69	0.69	0.65	0.67	0.61	-0.00	-0.02	0.01	-0.03	-0.02	0.01
2		0.75	0.75	0.76	0.70	0.71	0.71	0.00	-0.01	0.00	-0.02	-0.01	0.00
3		0.63	0.61	0.69	0.58	0.59	0.56	-0.03	-0.04	-0.01	-0.03	-0.02	-0.03
4		0.55	0.53	0.68	0.55	0.56	0.54	-0.06	-0.07	-0.04	-0.05	-0.05	-0.06
5		0.57	0.59	0.71	0.62	0.62	0.57	-0.05	-0.05	-0.02	-0.03	-0.03	-0.04
6		0.40	0.47	0.64	0.52	0.54	0.45	-0.05	-0.05	-0.02	-0.03	-0.02	-0.03
7		0.49	0.61	0.71	0.64	0.65	0.59	-0.06	-0.05	-0.02	-0.03	-0.01	-0.03
8		0.71	0.72	0.72	0.70	0.69	0.70	-0.07	-0.05	-0.02	-0.02	-0.01	-0.03
9		0.71	0.69	0.64	0.64	0.63	0.65	-0.05	-0.02	0.00	0.01	0.01	-0.00
10		0.73	0.69	0.63	0.61	0.62	0.62	-0.10	-0.07	-0.05	-0.03	-0.03	-0.04
LME	BEME	Explained Variation						Alpha					
1	1	0.14	0.13	-0.00	0.40	0.48	0.58	0.12	0.12	0.16	0.04	0.01	-0.04
1	2	0.09	0.28	0.32	0.57	0.62	0.69	0.15	0.11	0.09	0.02	-0.01	-0.03
1	3	0.23	0.48	0.64	0.69	0.71	0.73	0.22	0.16	0.08	0.08	0.06	0.02
1	4	0.40	0.65	0.68	0.74	0.74	0.73	0.20	0.13	0.03	0.06	0.04	0.02
1	5	0.54	0.70	0.53	0.70	0.69	0.74	0.17	0.10	-0.02	0.04	0.03	0.03
2	1	0.23	0.29	-0.03	0.56	0.67	0.64	0.08	0.07	0.13	0.01	-0.02	-0.01
2	2	0.22	0.47	0.41	0.76	0.81	0.78	0.14	0.09	0.10	0.02	-0.00	0.00
2	3	0.28	0.69	0.76	0.84	0.86	0.83	0.19	0.11	0.08	0.05	0.04	0.04
2	4	0.42	0.76	0.83	0.85	0.85	0.82	0.15	0.06	0.02	0.02	0.01	0.00
2	5	0.58	0.80	0.83	0.83	0.83	0.82	0.13	0.05	0.01	0.01	0.01	0.02
3	1	0.39	0.48	0.13	0.65	0.74	0.70	0.10	0.08	0.14	0.04	0.01	0.02
3	2	0.43	0.69	0.61	0.83	0.86	0.83	0.15	0.09	0.11	0.05	0.04	0.05
3	3	0.44	0.82	0.86	0.90	0.90	0.89	0.16	0.08	0.06	0.04	0.04	0.05
3	4	0.53	0.84	0.87	0.88	0.88	0.86	0.13	0.05	0.03	0.02	0.02	0.03
3	5	0.64	0.82	0.84	0.83	0.82	0.80	0.09	0.01	-0.00	-0.01	-0.01	0.02
4	1	0.76	0.73	0.49	0.75	0.80	0.79	0.01	0.02	0.07	0.01	-0.00	0.01
4	2	0.82	0.86	0.86	0.87	0.87	0.87	0.03	0.01	0.01	0.00	-0.00	0.00
4	3	0.75	0.88	0.88	0.86	0.85	0.85	0.07	0.01	-0.01	-0.00	0.00	0.01
4	4	0.75	0.88	0.89	0.87	0.86	0.85	0.06	0.00	-0.02	-0.01	0.00	0.01
4	5	0.76	0.81	0.83	0.80	0.80	0.79	0.03	-0.04	-0.04	-0.03	-0.02	-0.01
5	1	0.28	0.60	0.66	0.63	0.62	0.59	-0.11	-0.06	-0.02	-0.03	-0.03	-0.05
5	2	0.71	0.72	0.77	0.74	0.75	0.71	-0.06	-0.06	-0.04	-0.03	-0.02	-0.04
5	3	0.83	0.69	0.72	0.70	0.73	0.70	-0.02	-0.05	-0.05	-0.04	-0.02	-0.03
5	4	0.80	0.72	0.75	0.73	0.73	0.73	-0.04	-0.07	-0.06	-0.05	-0.03	-0.02
5	5	0.72	0.54	0.65	0.62	0.64	0.63	-0.05	-0.10	-0.07	-0.06	-0.04	-0.02
		Explained Variation						Cross-Sectional R^2					
All		0.48	0.60	0.57	0.70	0.72	0.72	0.60	0.82	0.86	0.96	0.98	0.97

This table shows the out-of-sample explained variation and pricing errors for the combined double sorted portfolios based on size (`LME`) and book-to-market ratio (`BEME`) and decile sorted portfolios based on investment (`Investment`). UNC (SV) and UNC (SVI) are unconditional models with respect to the test assets, that is, they use only size and value respectively size, value and investment for the SDF weights, but set g to a constant. The GAN models use a non-trivial g. GAN (SVI-SV) allows the SDF weight ω to depend on size, value and investment, but the test asset function g to depend only on size and investment. The model labeled as GAN is our benchmark model estimated with all characteristics and macroeconomic information. The portfolios are value-weighted.

IA.G. Predictive Portfolios

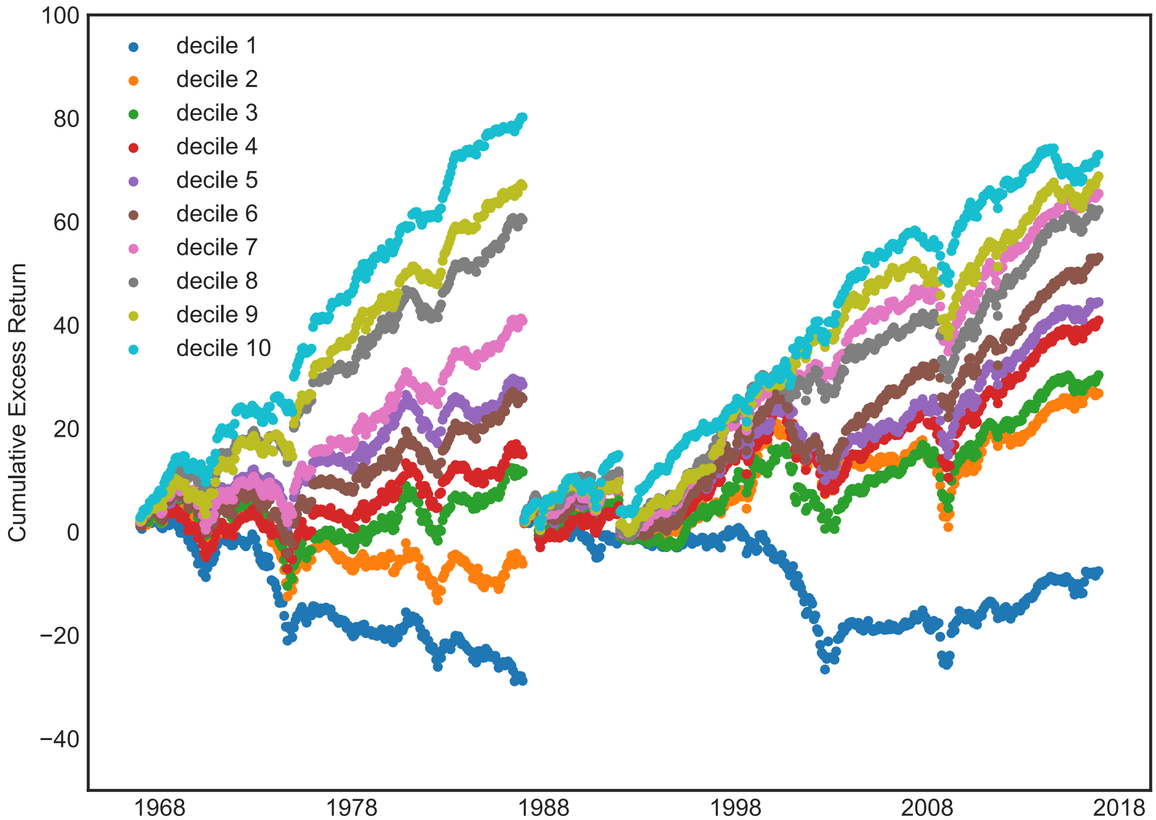
IA.G.1. Decile Sorted Portfolios for GAN

Figure IA.16: Cumulative Excess Return of Equally Weighted Decile Sorted Portfolios with GAN



This figure shows the 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.

Figure IA.17: Cumulative Excess Return of Value Weighted Decile β Portfolios with GAN



This figure shows the 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.

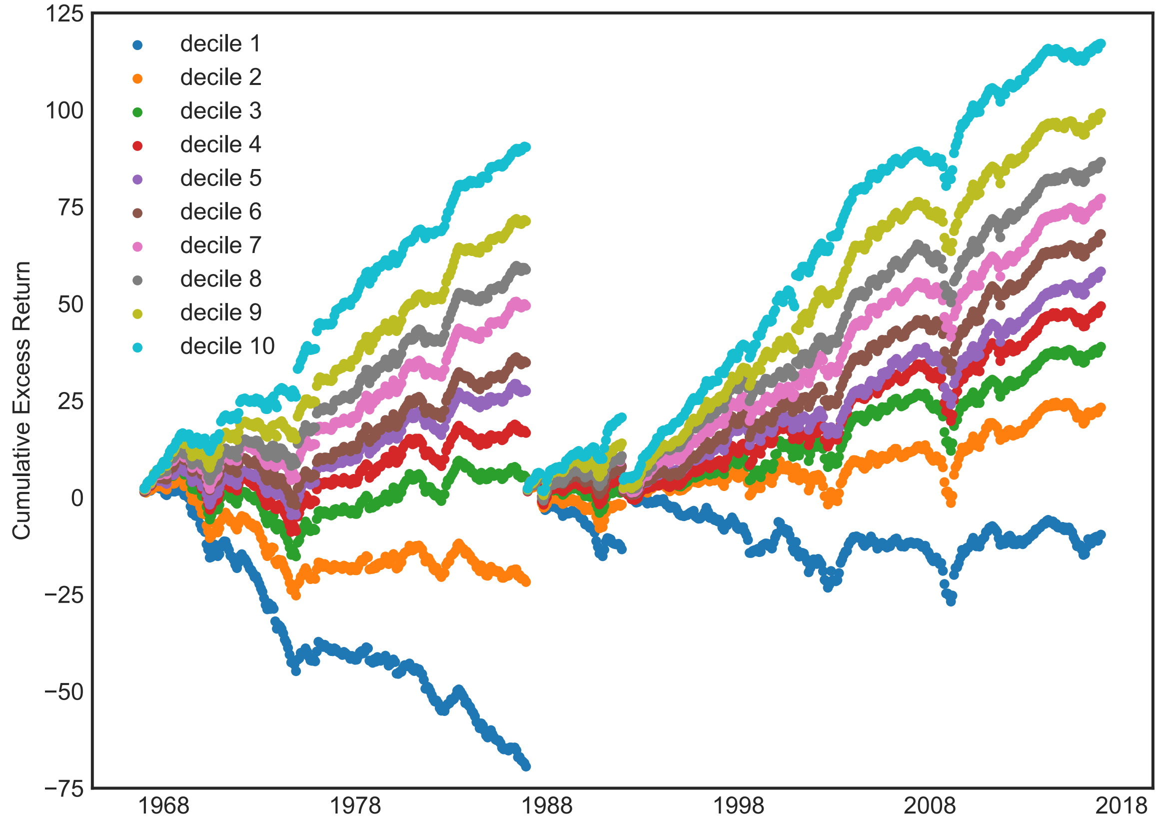
Table IA.XVII: Time Series Pricing Errors for Value Weighted β-Sorted Portfolios

	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

This table shows the 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 full time period and the test period. Within each decile the stocks are value weighted.

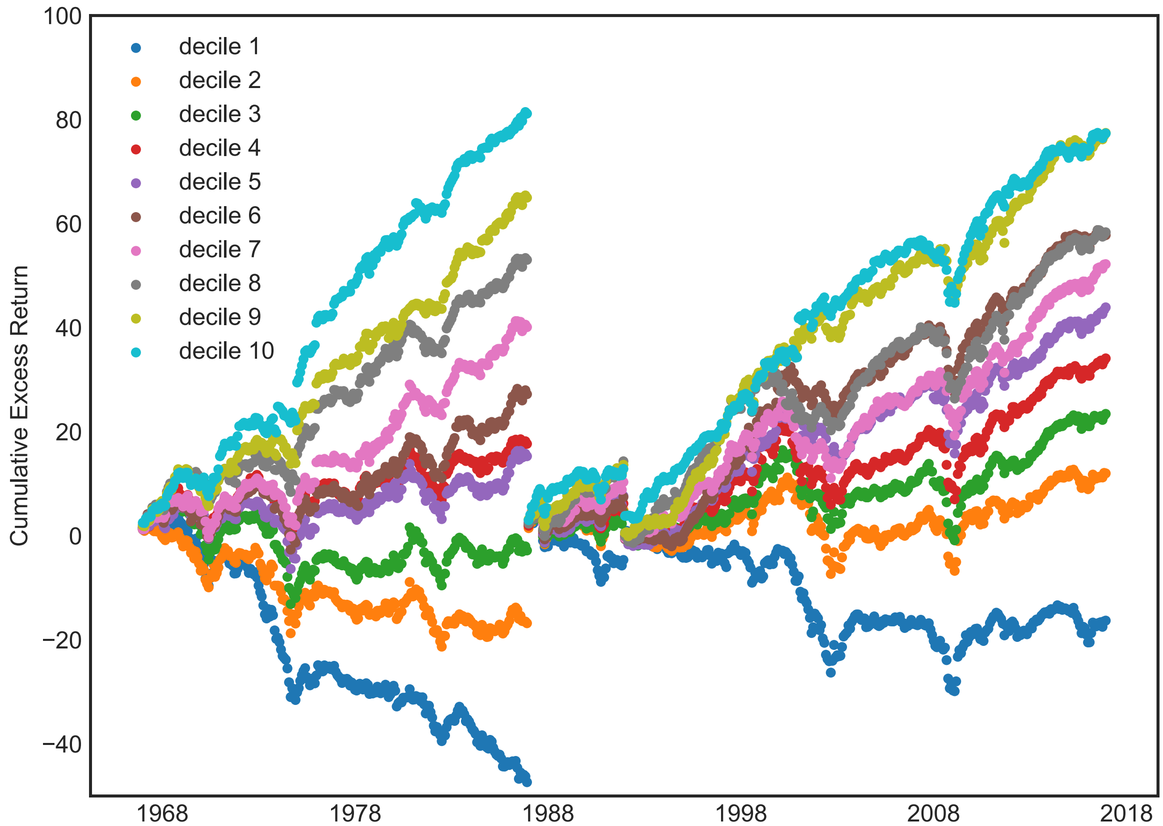
IA.G.2. Decile Sorted Portfolios for FFN

Figure IA.18: Cumulative Excess Return of Equally Weighted Decile β Portfolios with FFN



This figure shows the 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.

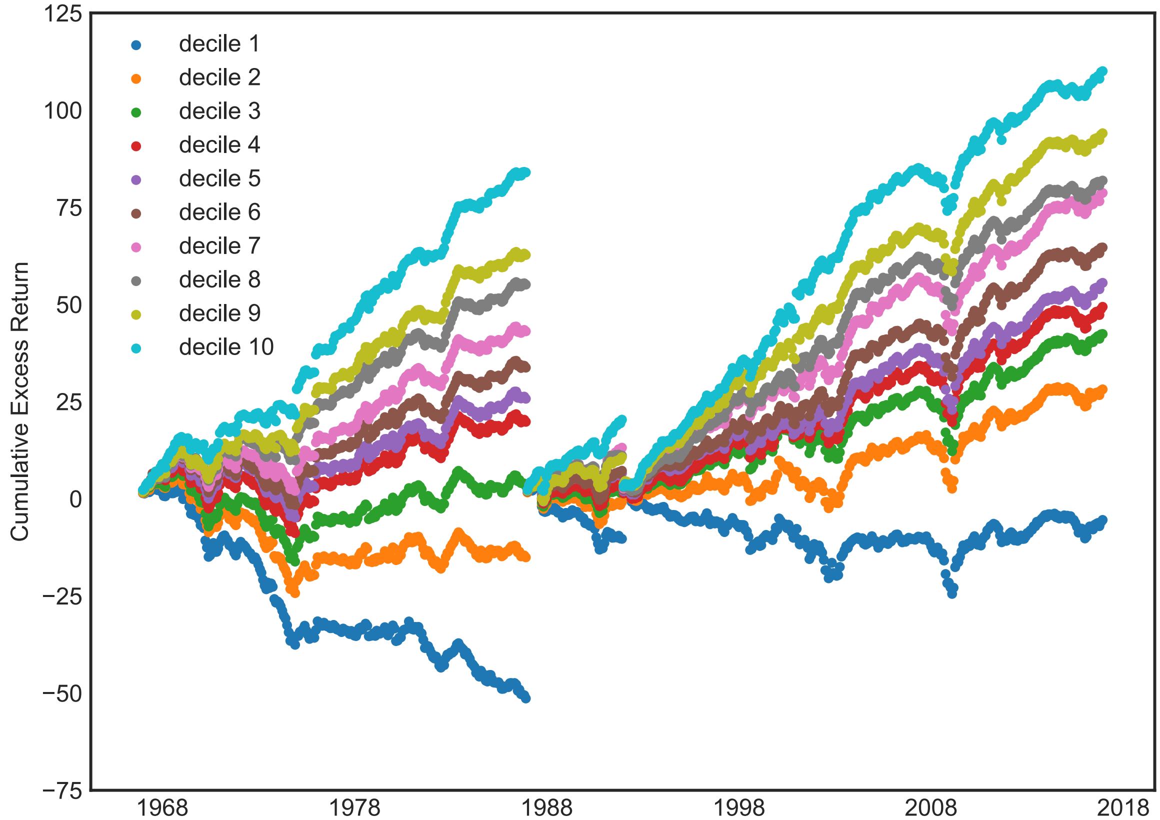
Figure IA.19: Cumulative Excess Return of Value Weighted Decile β Portfolios with FFN



This figure shows the 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.

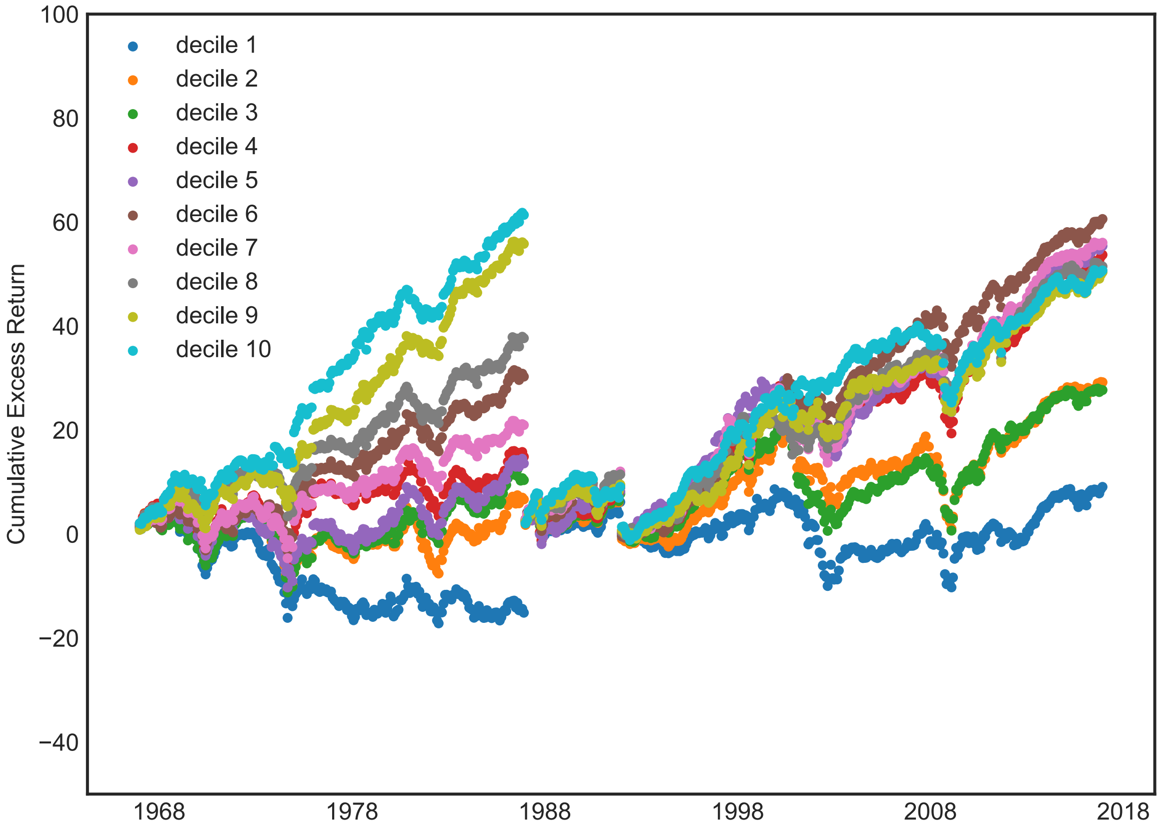
IA.G.3. Decile Sorted Portfolios for EN

Figure IA.20: Cumulative Excess Return of Equally Weighted Decile β Portfolios with EN



This figure shows the 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.

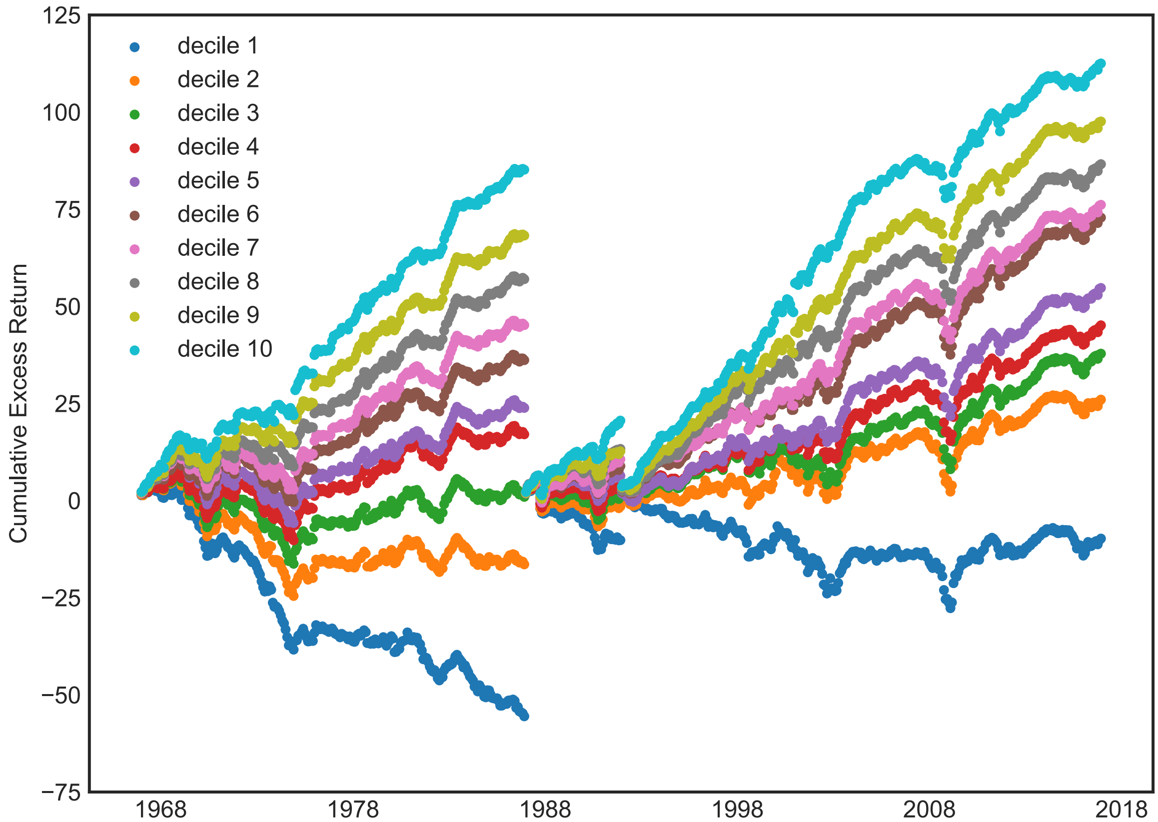
Figure IA.21: Cumulative Excess Return of Value Weighted Decile β Portfolios with EN



This figure shows the 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.

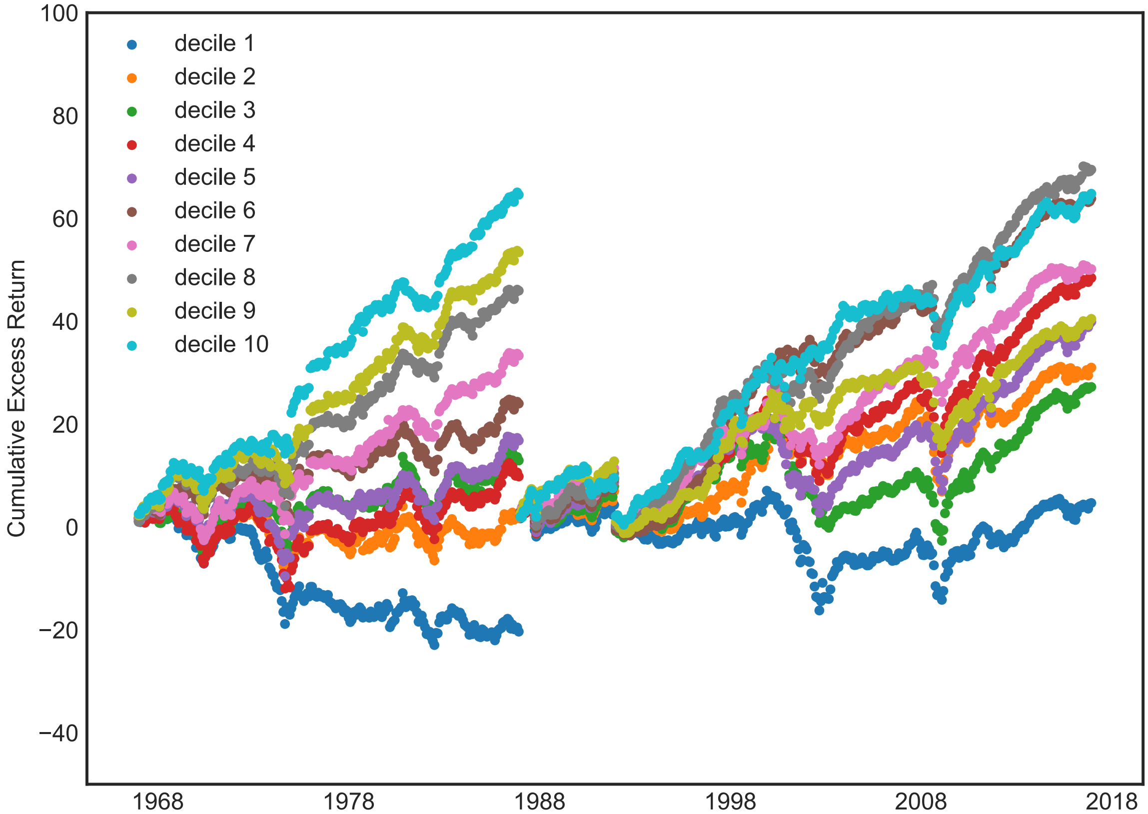
IA.G.4. Decile Sorted Portfolios for LS

Figure IA.22: Cumulative Excess Return of Equally Weighted Decile β Portfolios with LS



This figure shows the 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.23: Cumulative Excess Return of Value Weighted Decile β Portfolios with LS

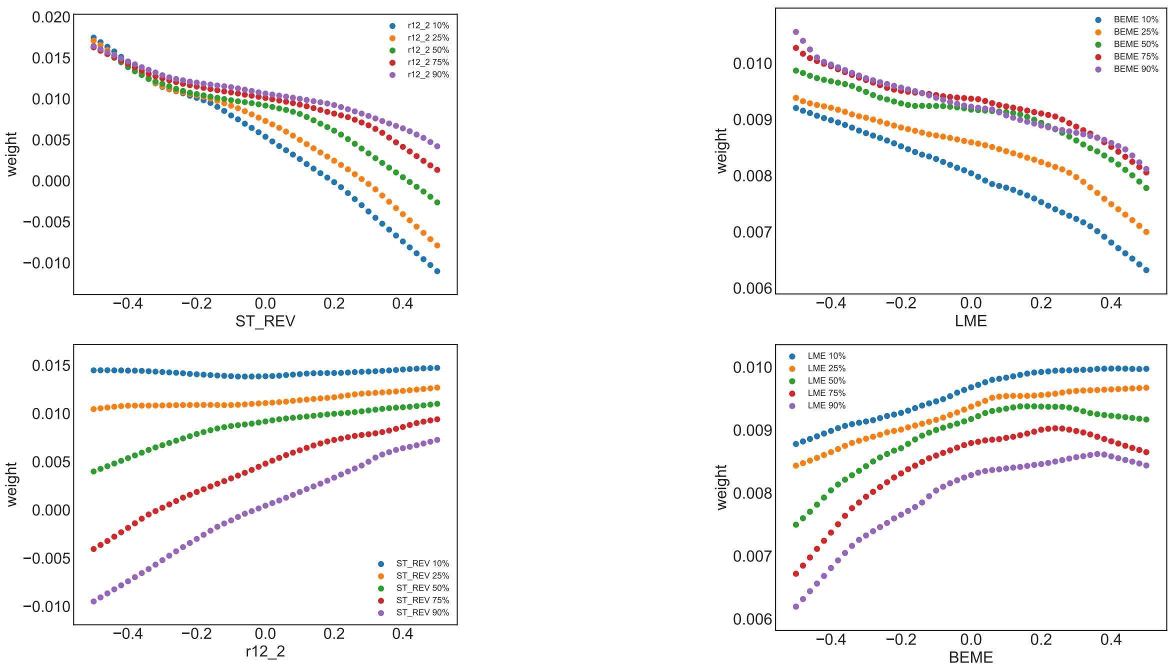


This figure shows the cumulative excess return of decile sorted portfolios based on the risk loadings β for LS. The first portfolio is based on the smallest decile of risk loadings, while the last decile portfolio is constructed with the largest loading decile. Within each decile the stocks are value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

IA.H. SDF Structure for FFN, EN and LS

IA.H.1. SDF Structure for FFN

Figure IA.24: SDF weight ω as a Function of Characteristics for FFN

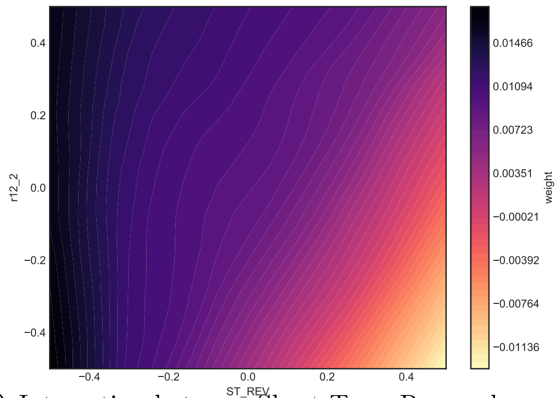


(a) Interaction between Short-Term Reversal (ST_REV) and Momentum (r12_2)

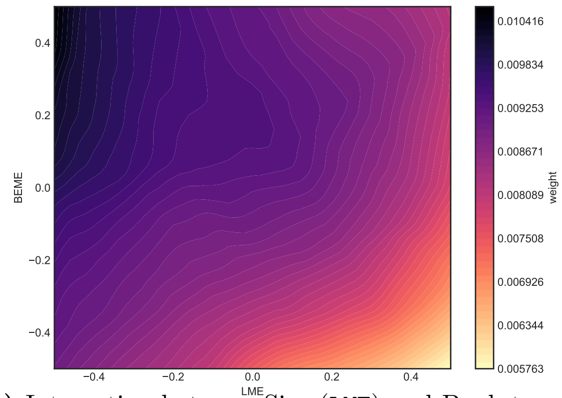
(b) Interaction between Size (LME) and Book to Market Ratio (BEME)

This figures show the SDF weight ω as function of short-term reversal, momentum, size and book-to-market ratio for different quantiles of the second variable while keeping the remaining variables at their mean level.

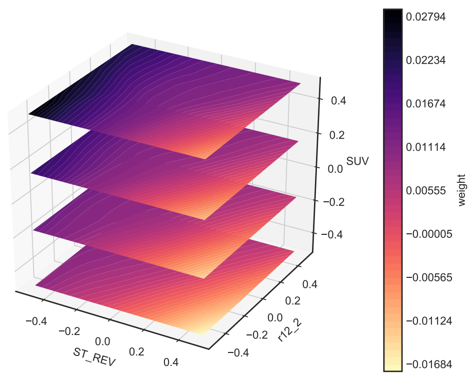
Figure IA.25: SDF weight ω as a Function of Characteristics for FFN



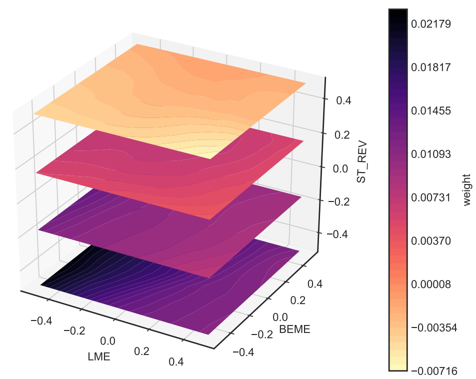
(a) Interaction between Short-Term Reversal (ST_REV) and Momentum (r12_2)



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



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

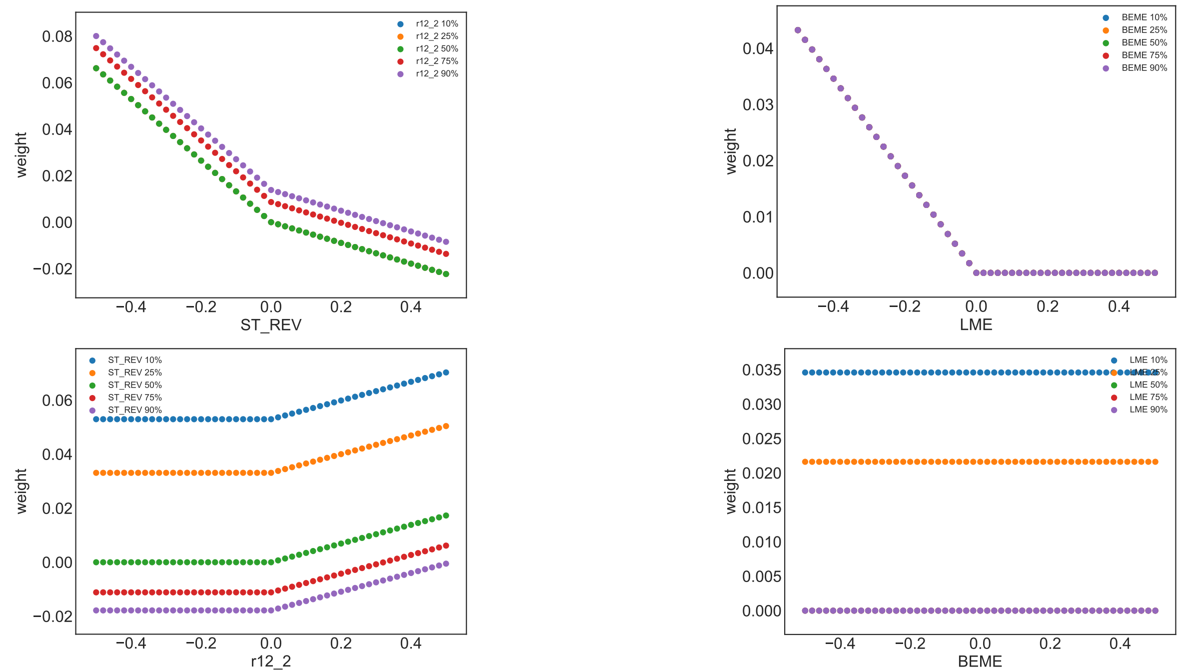


(d) Interaction between Size (LME), Book to Market Ratio (BEME) and Short-Term Reversal (ST_REV)

These figures show the SDF weight ω as two- and three-dimensional functions of characteristics keeping the remaining variables at their mean level.

IA.H.2. SDF Structure for EN

Figure IA.26: SDF weight ω as a Function of Characteristics for EN

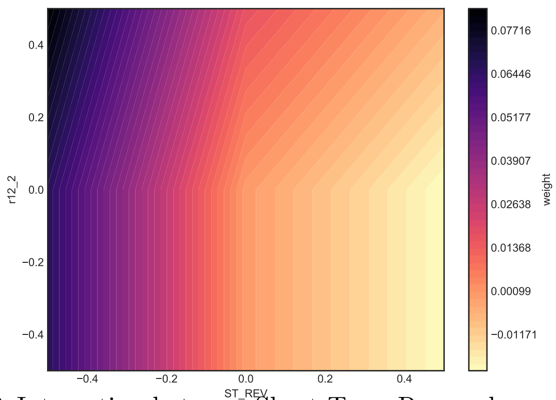


(a) Interaction between Short-Term Reversal (`ST_REV`) and Momentum (`r12_2`)

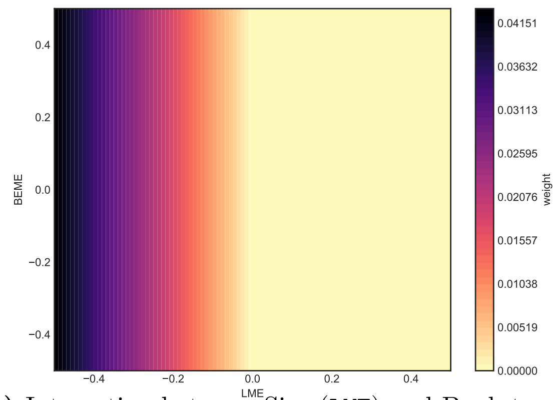
(b) Interaction between Size (`LME`) and Book to Market Ratio (`BEME`)

This figures show the SDF weight ω as function of short-term reversal, momentum, size and book-to-market ratio for different quantiles of the second variable while keeping the remaining variables at their mean level.

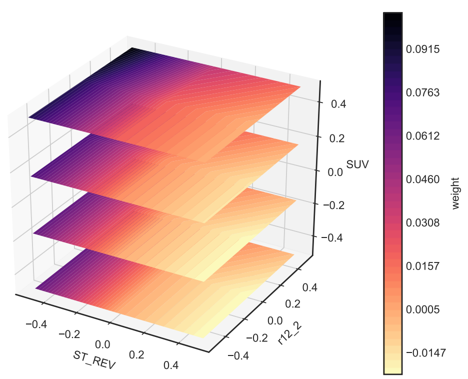
Figure IA.27: SDF weight ω as a Function of Characteristics for EN



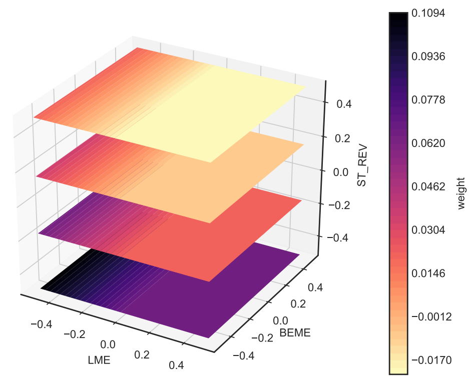
(a) Interaction between Short-Term Reversal (ST_REV) and Momentum (r12_2)



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



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

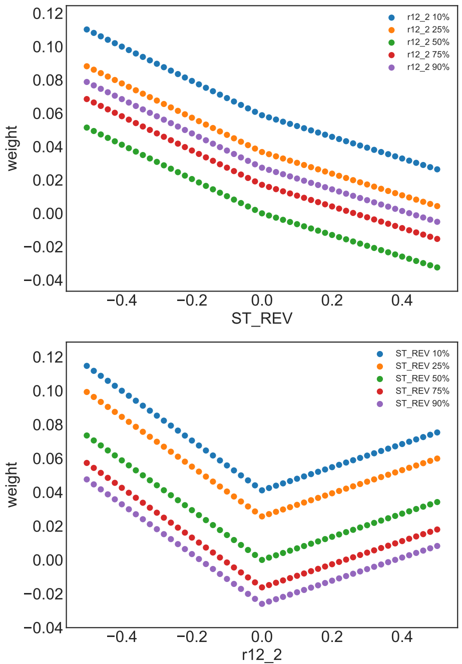


(d) Interaction between Size (LME), Book to Market Ratio (BEME) and Short-Term Reversal (ST_REV)

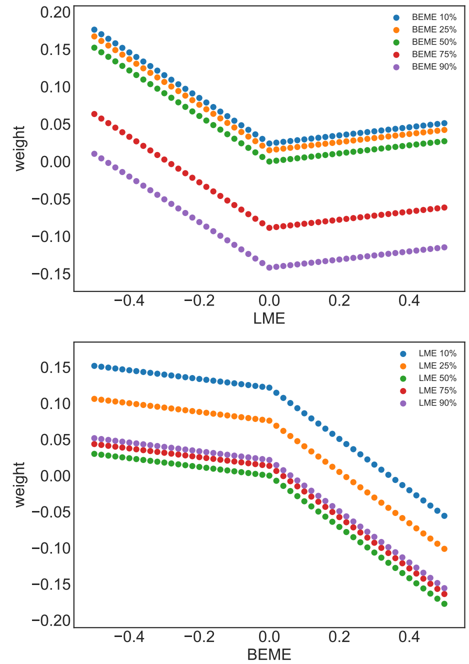
These figures show the SDF weight ω as two- and three-dimensional functions of characteristics keeping the remaining variables at their mean level.

IA.H.3. SDF Structure for LS

Figure IA.28: SDF weight ω as a Function of Characteristics for LS



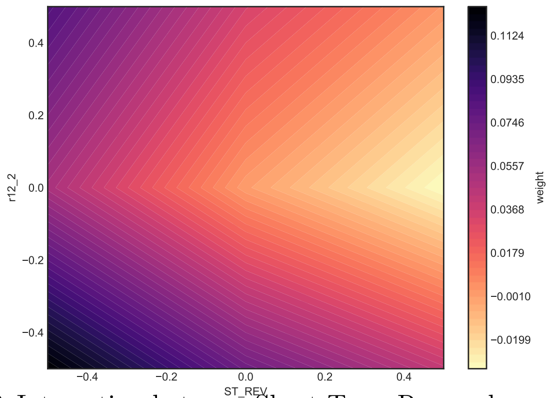
(a) Interaction between Short-Term Reversal (`ST_REV`) and Momentum (`r12_2`)



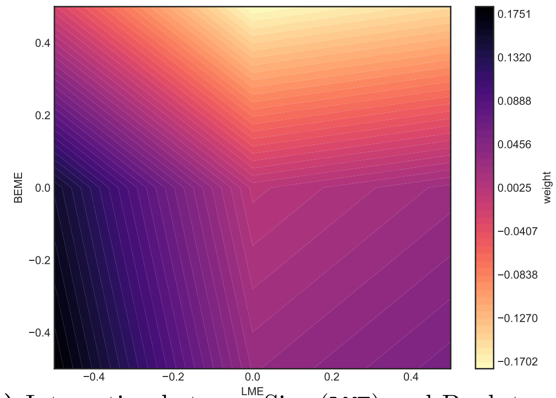
(b) Interaction between Size (`LME`) and Book to Market Ratio (`BEME`)

This figures show the SDF weight ω as function of short-term reversal, momentum, size and book-to-market ratio for different quantiles of the second variable while keeping the remaining variables at their mean level.

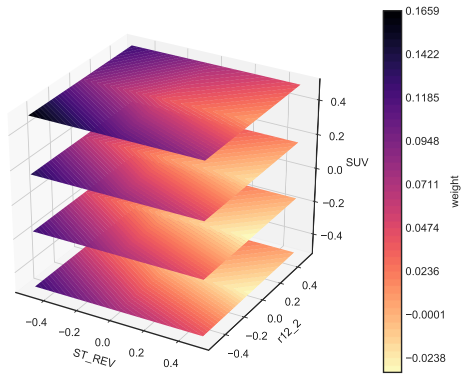
Figure IA.29: SDF weight ω as a Function of Characteristics for LS



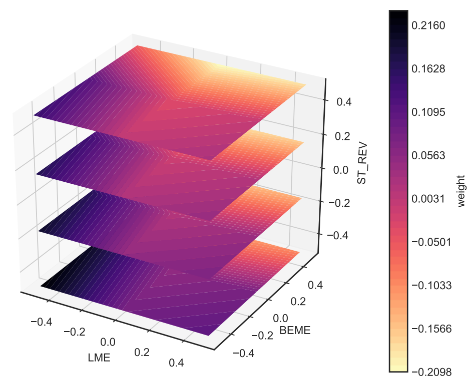
(a) Interaction between Short-Term Reversal (ST_REV) and Momentum (r12_2)



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



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



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

These figures show the SDF weight ω as two- and three-dimensional functions of characteristics keeping the remaining variables at their mean level.

IA.I. Machine-Learning Investment

Table IA.XVIII: SDF Portfolio 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

This table reports the Sharpe ratio, maximum 1-month loss and maximum drawdown of the SDF portfolios. We include the mean-variance efficient portfolio based on the 3 and 5 Fama-French factors.

Table IA.XIX: 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

This table reports the turnover for positions with positive and negative weights for the SDF portfolios. 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 $\|w_t\|_1 = 1$.

Table IA.XX: SDF Risk Measures for Large Market Cap Stocks

	SR			Max Loss			Max Drawdown		
Model	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
Size ≥ 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 ≥ 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

The table shows the monthly Sharpe Ratio, maximum 1-month loss and maximum drawdown of the SDF portfolios for the GAN, FFN, EN and LS models. The models are estimated on all stocks but evaluated on stocks with market capitalization larger than 0.01% or 0.001% of the total market capitalization.

Table IA.XXI: 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

This table shows 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 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.

IA.J. Additional Empirical Results

IA.J.1. Asset Pricing on Sorted Portfolios for Additional Characteristics

Table IA.XXII: Explained Variation and Pricing Errors for Double-Sorted Portfolios based on Size and Dividend Yield

		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

This table shows the out-of-sample explained variation and pricing errors for decile sorted portfolios based on Size (`LME`) and Dividend Yield (`D2P`).

Table IA.XXIII: Explained Variation and Pricing Errors for Decile Sorted Portfolios based on Standard Unexplained Volume

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

This table shows the out-of-sample explained variation and pricing errors for decile sorted portfolios based on Standard Unexplained Volume (SUV).

Table IA.XXIV: Explained Variation and Pricing Errors for Decile Sorted Portfolios based on Net Operating Assets

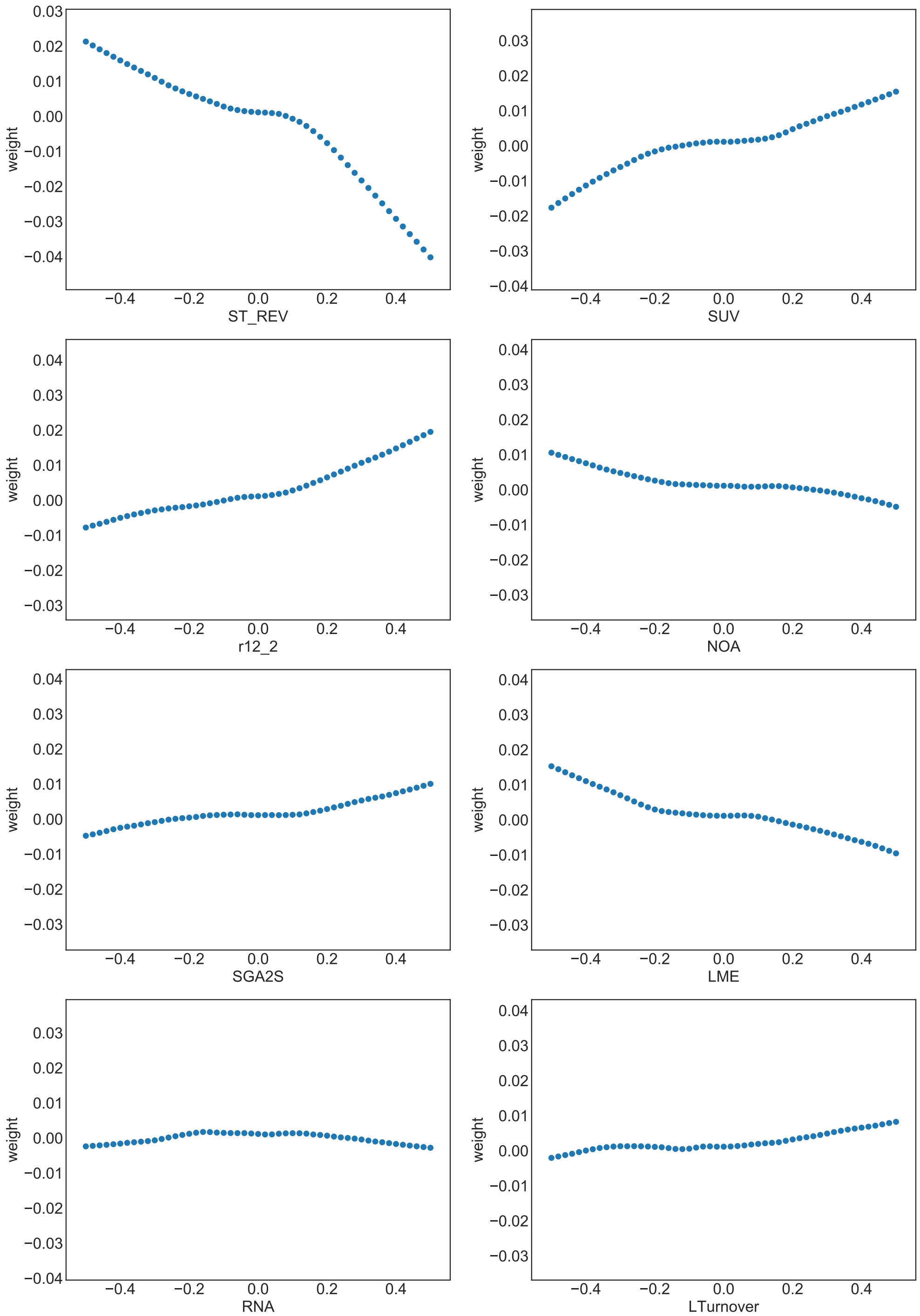
NOA	EN	FFN	GAN	EN	FFN	GAN
Decile	Explained Variation			Alpha		
1	0.41	0.55	0.66	0.17	0.10	0.09
2	0.57	0.72	0.80	0.05	-0.01	0.04
3	0.58	0.79	0.84	-0.06	-0.07	-0.03
4	0.69	0.76	0.78	0.02	0.01	0.05
5	0.73	0.75	0.77	-0.03	-0.04	0.00
6	0.64	0.75	0.75	0.06	0.03	0.05
7	0.72	0.82	0.83	0.02	-0.01	-0.00
8	0.67	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

This table shows the out-of-sample explained variation and pricing errors for decile sorted portfolios based on Net Operating Assets (NOA).

IA.J.2. SDF Structure for Additional Characteristics

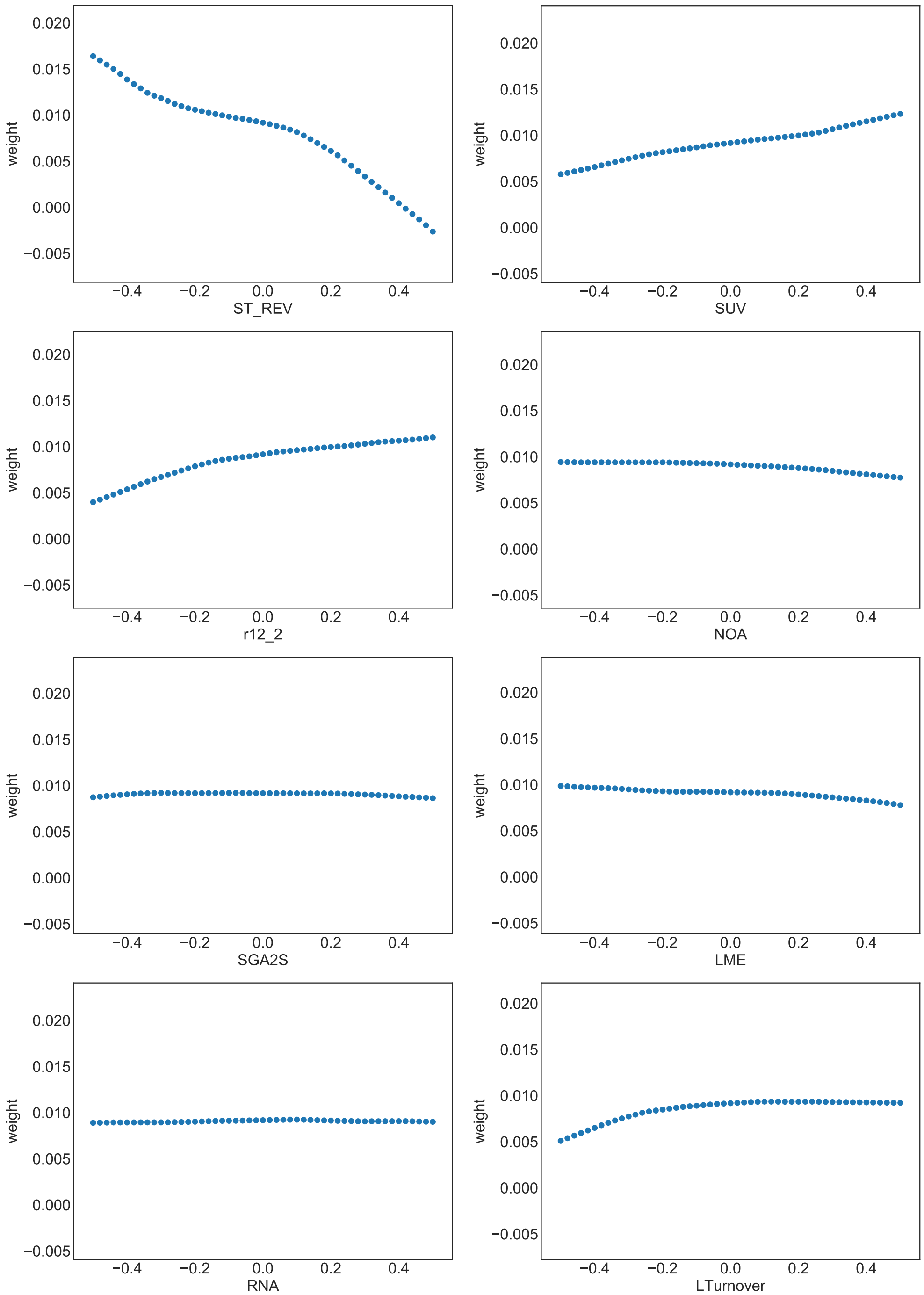
IA.J.2.1. One-Dimensional Relationship

Figure IA.30: SDF weight ω as a Function of Characteristics for GAN



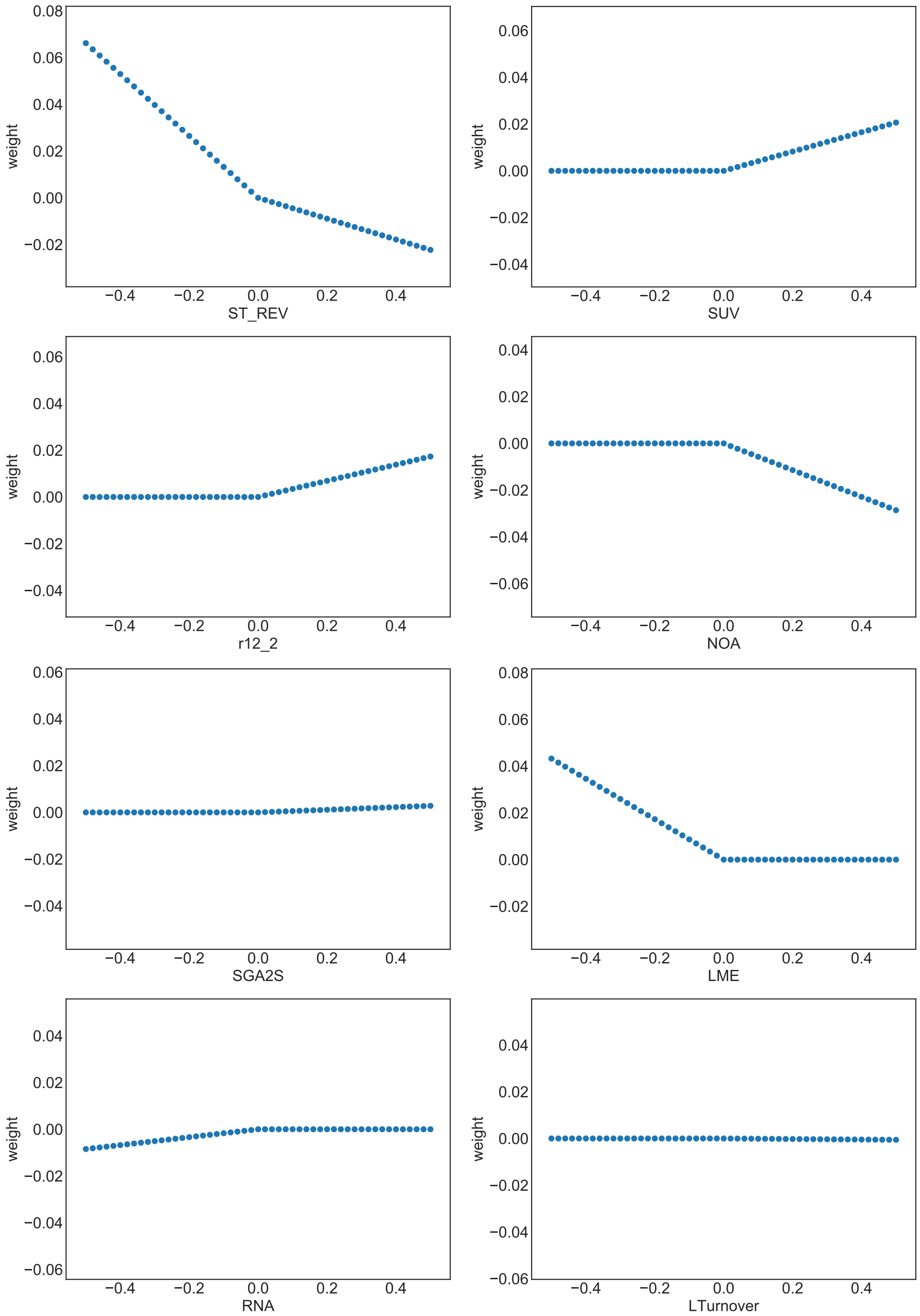
This figure shows the SDF weight ω as a one-dimensional function of characteristics keeping the other covariates at their mean level. The characteristics are Short-Term Reversal (`ST_REV`), Standard Unexplained Volume (`SUV`), Momentum (`r12_2`), Net Operating Assets (`NOA`), Selling, General and Administrative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

Figure IA.31: SDF weight ω as a Function of Characteristics for FFN



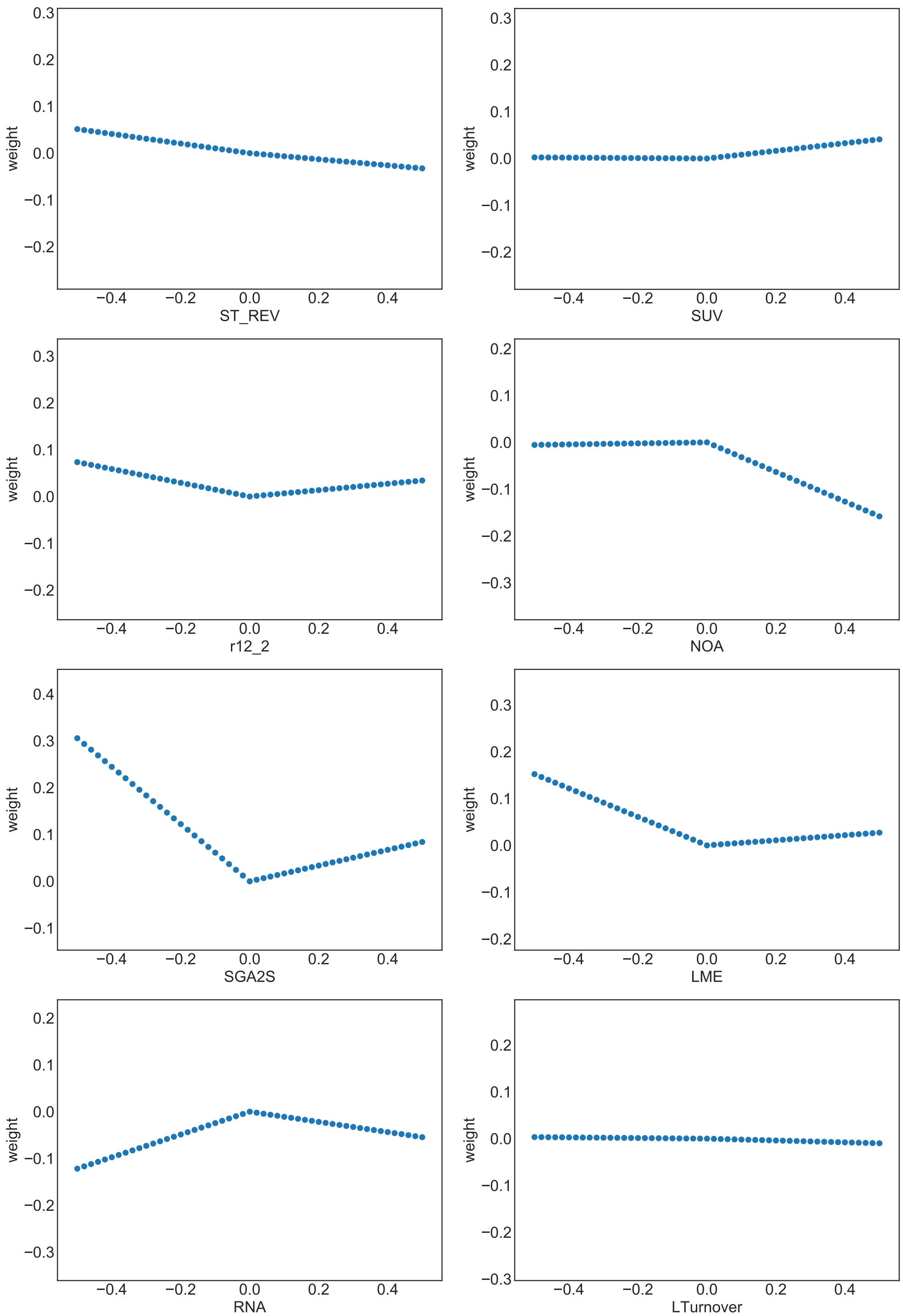
This figure shows the SDF weight ω as a one-dimensional function of characteristics keeping the other covariates at their mean level. The characteristics are Short-Term Reversal (ST_REV), Standard Unexplained Volume (SUV), Momentum (r12_2), Net Operating Assets (NOA), Selling, General and Administrative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

Figure IA.32: SDF weight ω as a Function of Characteristics for EN



This figure shows the SDF weight ω as a one-dimensional function of characteristics keeping the other covariates at their mean level. The characteristics 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`).

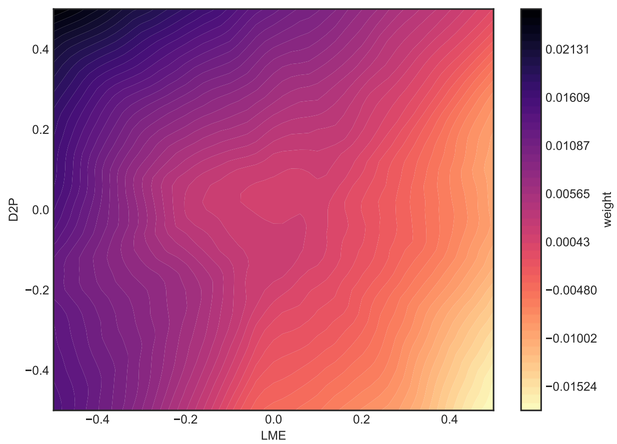
Figure IA.33: SDF weight ω as a Function of Characteristics for LS



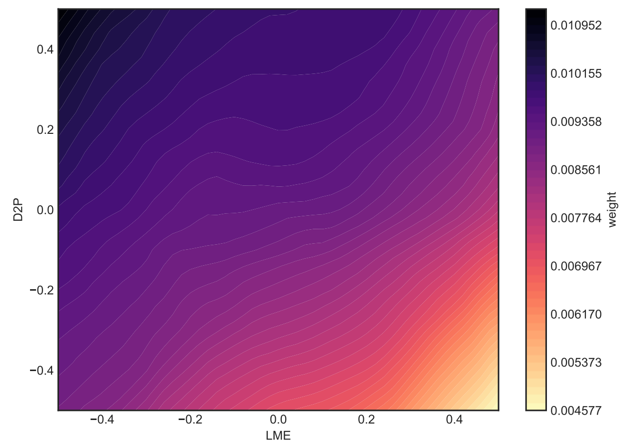
This figure shows the SDF weight ω as a one-dimensional function of characteristics keeping the other covariates at their mean level. The characteristics are Short-Term Reversal (`ST_REV`), Standard Unexplained Volume (`SUV`), Momentum (`r12_2`), Net Operating Assets (`NOA`), Selling, General and Administrative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

IA.J.2.2. Interaction between Characteristics

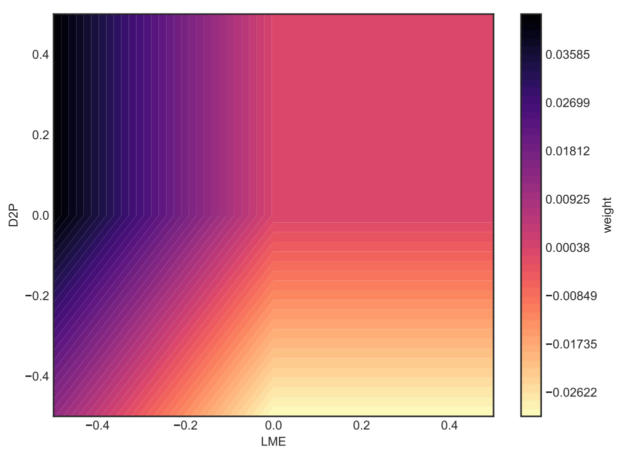
Figure IA.34: SDF weight ω as a Function of Size (LME) and Dividend Yield (D2P)



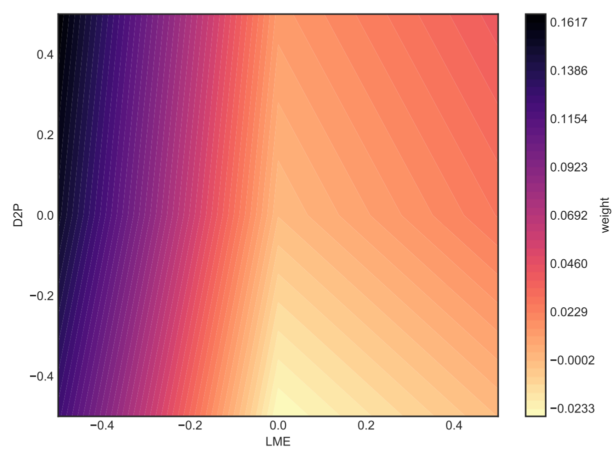
(a) GAN



(b) FFN



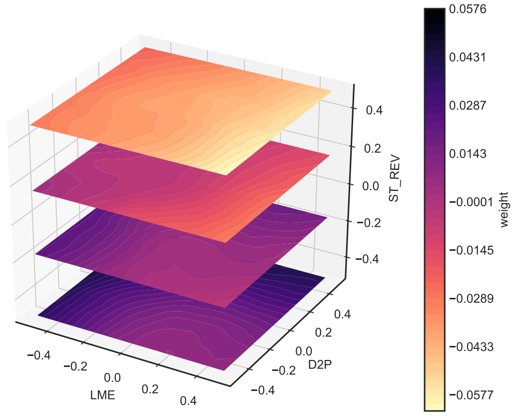
(c) EN



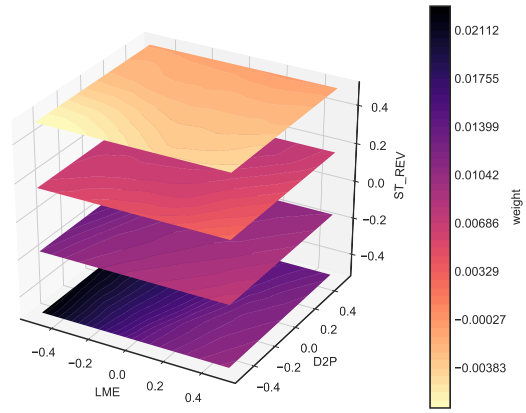
(d) LS

These figures show the SDF weight ω as two-dimensional function of characteristics keeping the remaining variables at their mean level. The two characteristics are Size (LME) and Dividend Yield (D2P).

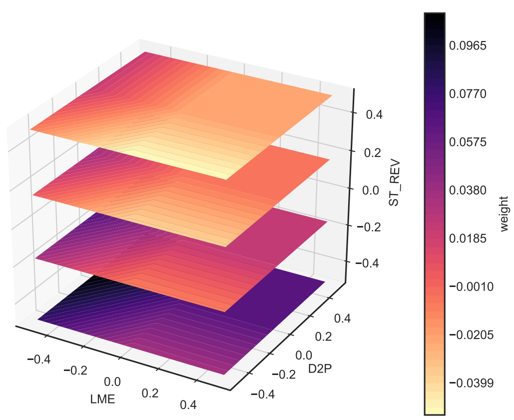
Figure IA.35: SDF weight ω as a Function of Size (LME), Dividend Yield (D2P) and Short-Term Reversal (ST_REV)



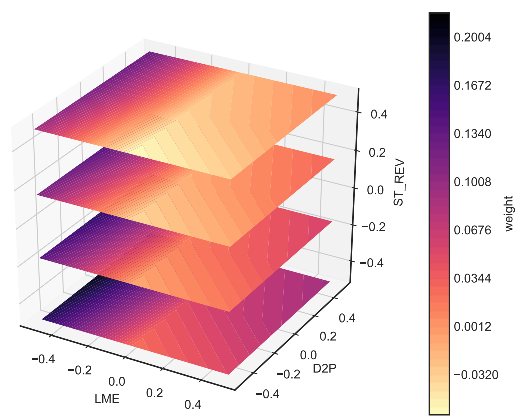
(a) GAN



(b) FFN



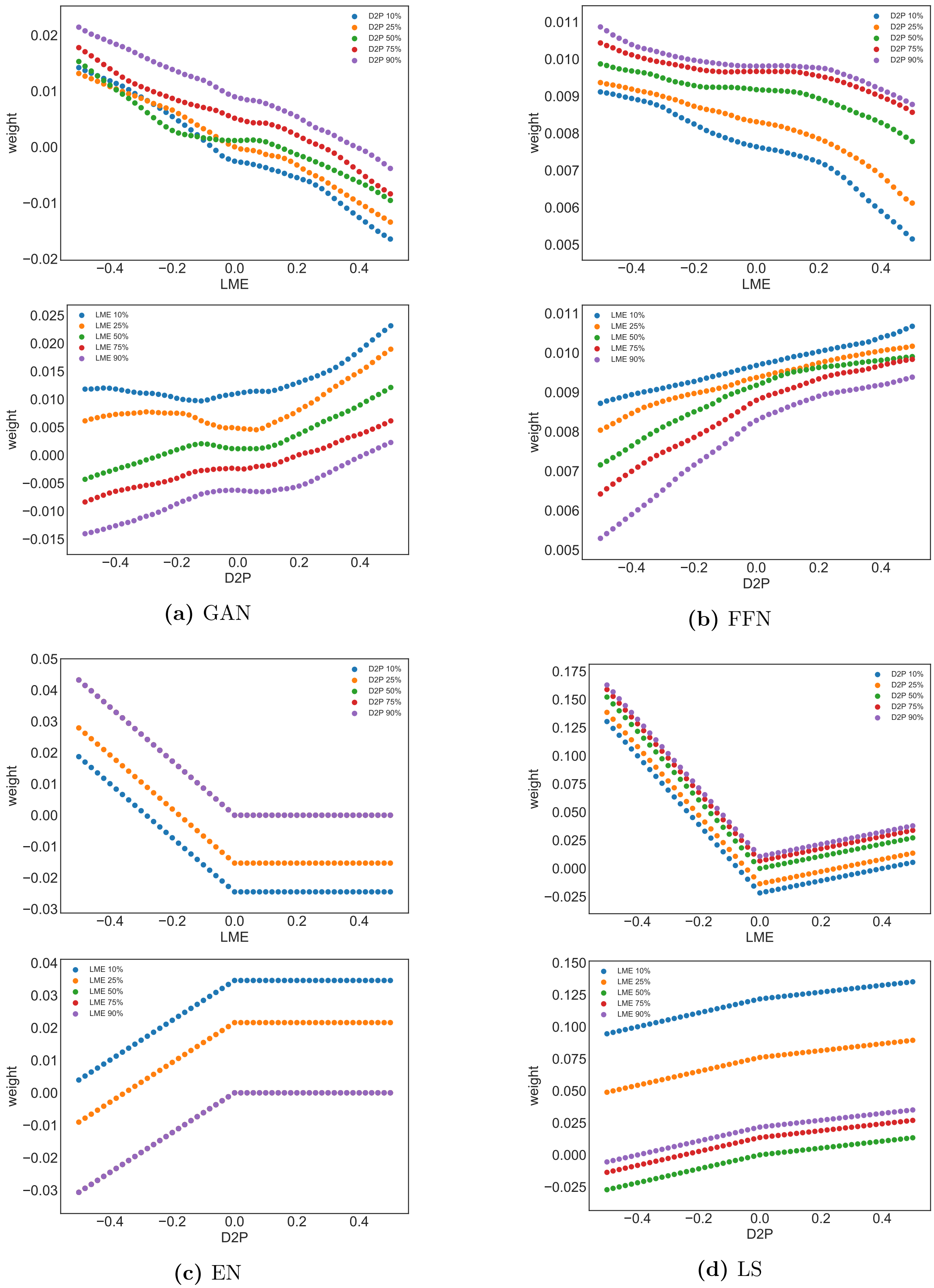
(c) EN



(d) LS

These figures show the SDF weight ω as three-dimensional function of characteristics keeping the remaining variables at their mean level. The two characteristics are Size (LME), Dividend Yield (D2P) and Short-Term Reversal (ST_REV).

Figure IA.36: SDF weight ω as a Function of Size (LME) and Dividend Yield (D2P)



(a) GAN

(b) FFN

(c) EN

(d) LS

This figures show the SDF weight ω as function of Size (LME) and Dividend Yield (D2P) for different quantiles of the second variable while keeping the remaining variables at their mean level.

IA.K. Data

IA.K.1. List of Macroeconomic Variables

Table IA.XXV: 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

Variable Name	Description	Source	tCode
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
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

Variable Name	Description	Source	tCode
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
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

This table reports the macroeconomic variables, their description, source and stationary transformation. The transformations (tCode) are (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) $\Delta(x_t/x_{t-1} - 1.0)$.

Table IA.XXVI: 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

This table describes the NBER Recessions.

IA.K.2. List of Firm-Specific Characteristics

Table IA.XXVII: Firm-Specific Characteristics

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, Karolyi, and Kho (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, Rajgopal, and Venkatachalam (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, Sun, and Zhang (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, Liu, Pflueger, and Weber (2018)
IdioVol	Idiosyncratic volatility	Standard deviation of the residuals from aregression of excess returns on the Fama and French three-factor model	Ang, Hodrick, Xing, and Zhang (2006)
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, Gulen, and Schill (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)

Acronym	Name	Definition	Reference
LT_Rev	Long-term reversal	Cumulative return from 60 months before the return prediction to 13 months before	Jegadeesh and Titman (2001)
Lturnover	Turnover	Turnover is last month's volume (VOL) over shares outstanding (SHROUT)	Datar, Naik, and Radcliffe (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, Hou, Teoh, and Zhang (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, Gerakos, Linnainmaa, and Nikolaev (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, Hodrick, Xing, and Zhang (2006)
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, Bartov, and Faurel (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, Neuhierl, and Weber (2020)

Acronym	Name	Definition	Reference
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)
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, Hodrick, Xing, and Zhang (2006)

This table summarize the firm-specific characteristics. It includes their acronym, name, definition and reference. We use the same characteristics as in the 2017 paper version of Freyberger, Neuhierl, and Weber (2020) and augment them with the characteristics listed on the Kenneth French Data Library.

REFERENCES

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006): "The Cross-Section of Volatility and Expected Returns," *Journal of Finance*, 61(1), 259–299.

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

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

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.

Bhandari, L. C. (1988): "Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence," *The Journal of Finance*, 43(2), 507–528.

Bondt, W. F. M. D., and R. Thaler (1985): "Does the Stock Market Overreact?," *The Journal of Finance*, 40(3), 793–805.

Bryzgalova, S., M. Pelger, and J. Zhu (2019): "Forest through the Trees: Building Cross-Sections of Stock Returns," *Working paper*.

Bustamante, M. C., and A. Donangelo (2017): "Product Market Competition and Industry Returns," *The Review of Financial Studies*, 30(12), 4216–4266.

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

Cooper, M. J., H. Gulen, and M. J. Schill (2008): "Asset Growth and the Cross-Section of Stock Returns," *Journal of Finance*, 63(4), 1609–1651.

D'Acunto, F., R. Liu, C. Pflueger, and M. Weber (2018): "Flexible Prices and Leverage," *Journal of Financial Economics*, 129(1), 46–68.

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

Desai, H., S. Rajgopal, and M. Venkatachalam (2004): "Value-Glamour and Accruals Mispricing: One Anomaly or Two?," *The Accounting Review*, 79(2), 355–385.

Fama, E. F., and K. R. French (1992): "The Cross-Section of Expected Stock Returns," *The Journal of Finance*, 47(2), 427–465.

Fama, E. F., and K. R. FrencH (1996): "Multifactor Explanations of Asset Pricing Anomalies," *The Journal of Finance*, 51(1), 55–84.

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

Fama, E. F., and J. D. MacBeth (1973): "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81(3), 607–636.

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

Freyberger, J., A. Neuhierl, and M. Weber (2020): "Dissecting characteristics nonparametrically," *Review of Financial Studies, forthcoming*, 33(5), 2326–2377.

Gandhi, P., and H. Lustig (2015): "Size anomalies in US bank stock returns," *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., AND C.-Y. HWANG (2004): "The 52-Week High and Momentum Investing," *The Journal of Finance*, 59(5), 2145–2176.

GORODNICHENKO, Y., AND M. WEBER (2016): "Are Sticky Prices Costly? Evidence from the Stock Market," *American Economic Review*, 106(1), 165–99.

HANSEN, L. P., AND R. JAGANNATHAN (1997): "Assessing Specification Errors in Stochastic Discount Factor Models," *The Journal of Finance*, 52(2), 557–590.

HAUGEN, R. A., AND N. L. BAKER (1996): "Commonality in the determinants of expected stock returns," *Journal of Financial Economics*, 41(3), 401 – 439.

HIRSHLEIFER, D., K. HOU, S. H. TEOH, AND Y. ZHANG (2004): "Do investors overvalue firms with bloated balance sheets?," *Journal of Accounting and Economics*, 38, 297 – 331.

HOU, K., G. A. KAROLYI, AND B.-C. KHO (2011): "What Factors Drive Global Stock Returns?," *The Review of Financial Studies*, 24(8), 2527–2574.

JEGADEESH, N., AND S. TITMAN (1993): "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *The Journal of Finance*, 48(1), 65–91.

——— (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., S. PRUITT, AND Y. SU (2019): "Characteristics Are Covariances: A Unified Model of Risk and Return," *Journal of Financial Economics*, 134(3), 501–524.

KOZAK, S., S. NAGEL, AND S. SANTOSH (2020): "Shrinking the Cross Section," *Journal of Financial Economics*, 135(2), 271–292.

LETTAU, M., AND M. PELGER (2020): "Factors that Fit the Time-Series and Cross-Section of Stock Returns," *Review of Financial Studies*, 33(5), 2274–2325.

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

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

LYANDRES, E., L. SUN, AND L. ZHANG (2008): "The New Issues Puzzle: Testing the Investment-Based Explanation," *The Review of Financial Studies*, 21(6), 2825–2855.

NOVY-MARX, R. (2011): "Operating Leverage," *Review of Finance*, 15(1), 103–134.

——— (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.

PONTIFF, J., AND A. WOODGATE (2008): "Share issuance and cross-sectional returns," *The Journal of Finance*, 63(2), 921–945.

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.