Macroeconomic Content of Characteristics-Based Asset Pricing Models: A Machine Learning Analysis*

Oleg Rytchkov[†] Xun Zhong[‡]

December 2020

Abstract

We consider seven characteristics-based asset pricing models and explore whether the non-market components of their stochastic discount factors (SDFs) are associated with macroeconomic shocks. Our analysis involves a comprehensive set of 120 macroeconomic variables and uses machine learning techniques to mitigate the overfitting problem caused by a large number of explanatory variables. We find that macroeconomic shocks are totally unrelated to the non-market SDF components. This conclusion extends to several theory-motivated macroeconomic shocks. Our results suggest that the empirical success of characteristics-based asset pricing models is produced by their ability to identify behavioral factors in stock returns rather than macroeconomic risks.

*We thank Gurdip Bakshi, Federico Bandi, Sudipta Basu, Konstantin Bauman, Victor Chernozhukov, Ming Fang, Ulas Misirli, Alberto Rossi, Guofu Zhou, participants of the 2020 FMA Virtual Conference, and seminar participants at Fordham University, Rutgers University – Camden, Temple University, and Washington State University for valuable comments and suggestions. All remaining errors are our own.

[†]Fox School of Business, Temple University, 1801 Liacouras Walk, Philadelphia, PA 19122. E-mail: rytchkov@temple.edu.

[‡]Gabelli School of Business, Fordham University, 45 Columbus Avenue, Room 620, New York, NY 10023. E-mail: xzhong21@fordham.edu.

1. Introduction

One of the most important and challenging objectives of asset pricing is the construction of models that can explain the cross section of expected stock returns. The most successful among such models are characteristics-based models whose pricing factors include the market returns and returns on trading strategies based on firm characteristics. The prime example is the Fama-French three-factor model (Fama and French, 1993); more recent examples include the Fama-French five-factor model (Fama and French, 2015), q-factor model (Hou, Xue, and Zhang, 2015), and behavioral three-factor model (Daniel, Hirshleifer, and Sun, 2020). The factors in the latest generation of characteristics-based asset pricing models combine information from multiple firm characteristics (e.g., Stambaugh and Yuan, 2017; Kozak, Nagel, and Santosh, 2018, 2020; Kelly, Pruitt, and Su, 2019; Gu, Kelly, and Xiu, 2020a). Although many characteristics-based asset pricing models can concisely describe the main patterns in the cross section of expected stock returns, they all share a serious limitation: their non-market factors lack a clear economic interpretation. In particular, it is still an open question whether expected returns are determined by exposures of stock returns to macroeconomic risks or investor sentiment (e.g., Nagel, 2013).

The objective of our paper is to shed new light on the characteristics-based asset pricing models by exploring whether their non-market factors are associated with macroeconomic shocks. Although a number of studies attempt to link characteristics-based factors to macroeconomic risks (e.g., Liew and Vassalou, 2000; Vassalou, 2003; Petkova, 2006; Hahn and Lee, 2006), our paper differs from them in several crucial respects. Most importantly, instead of focusing on a few prominent shocks, our analysis involves shocks to 120 macroeconomic variables that represent a wide range of macroeconomic activities. Although individual shocks are only weakly related to stock returns, taken together they are more likely to track macroeconomic risks priced by investors. However, it is impossible to use the OLS regression to measure the joint explanatory power of so many shocks because the results would be severely distorted by the overfitting problem.

To overcome this problem, we use the machine learning framework. We run the elastic net regression to identify a combination of macroeconomic shocks that is correlated with asset pricing factors not only in sample but also out of sample. Out-of-sample R^2 (R^2_{os}) produced by the elastic net regression fitted values allows us to quantify the explanatory power of macroeconomic shocks. To the best of our knowledge, our paper is the first finance study that uses machine learning to measure the explainability of one time series by a large number of others, not to predict returns or construct a parsimonious asset pricing model.

Another distinguishing feature of our paper is the focus on the models' stochastic discount factors (SDFs) and their non-market components rather than on individual pricing factors. Because all factors in the characteristics-based models are returns on tradable portfolios, the SDFs constructed from them are also tradable and uniquely characterize the models (e.g., Cochrane, 2005; Back, 2010). Moreover, because each model contains the market returns as one of the factors, the SDFs can be uniquely decomposed into a component proportional to the market and an orthogonal non-market component that includes all other factors. In contrast, the decomposition of the SDF into individual factors such as SMB and HML is not unique, and it is always possible to find another set of factors with even a different number of factors that would represent exactly the same SDF. Therefore, there might exist situations in which individual factors are associated with macroeconomic shocks, but their linear combination in the SDF is not. By examining the SDF's non-market component as a whole, we avoid the potential ambiguity.

We start our analysis with constructing the SDFs and their non-market components for six characteristics-based asset pricing models. The considered models include the Fama-French three- and five-factor models (Fama and French, 1993, 2015), the q-factor model (Hou, Xue, and Zhang, 2015), the Barillas-Shanken six-factor model (Barillas and Shanken, 2018), the mispricing four-factor model (Stambaugh and Yuan, 2017), and the behavioral three-factor model (Daniel, Hirshleifer, and Sun, 2020). For each model, we estimate its SDF by GMM. All considered models include the market returns as one of the factors, which allows us to decompose the SDFs into market and non-market components. Because the non-market factors are typically correlated with market returns, the non-zero correlation is inherited by the non-market SDF components and complicates their interpretation. Therefore, we also construct the SDFs' non-market components that are orthogonal to the market. We augment the set of constructed SDFs with the market-orthogonalized SDF of Kozak, Nagel, and Santosh (2020).

Next, we build a comprehensive set of macroeconomic shocks. For this purpose, we use 120 macroeconomic variables from FRED-MD, which is a publicly available macroeconomic data set maintained by the Federal Reserve Bank of St. Louis. FRED-MD contains information on a broad spectrum of economic activities, and it is designed to be a standard data set for the "big data" macroeconomic research. We construct macroeconomic shocks as residuals of the AR(1) processes fitted for each macroeconomic variable.

The main objective of our analysis is to explore whether the non-market factors of characteristics-based asset pricing models have any macroeconomic content. Conceptually, the answer to this question can be obtained by regressing the SDFs' non-market components on macroeconomic shocks and measuring the macroeconomic interpretability of a model by the regression R^2. This procedure is reliable and easily implementable when the number of shocks is small. However, the OLS regression works poorly when the number of explanatory variables is large because in this case it overfits the data, and the OLS regression R^2 is a severely upward biased estimator of the population R^2. To confront overfitting, we use the elastic net regression. Elastic net imposes a regularization on the linear regression parameters that constrains them in sample thereby mitigating overfitting and improving the model performance out of sample. To find the regularization parameters, we use the cross-validation, which is another standard machine learning technique.

We implement the elastic net-based estimation of the population R^2 in several steps. First, we split the whole sample into the training and testing samples. Then, using only the training sample, we find the optimal regularization parameters of the elastic net regression through cross-validation and estimate the regression parameters. Finally, we predict the non-market component of the SDF in the testing sample by the fitted values from the model with the estimated parameters and compute R^2_{os}. Using simulations, we demonstrate that in contrast to the R^2_{os} produced by the OLS regression, the R^2_{os} of the elastic net regression is a reliable estimator of the population R^2.

By itself, R^2_{os} is a statistic that depends on the particular realizations of the training and testing samples. First, because of the finite number of observations in the testing period, the computed R^2_{os} may be higher or lower than it would be if we had an infinite testing sample for the same model. Second, the observations in the training sample determine the parameter estimates, which, in turn, determine the quality of the prediction. To assess both types of errors, we implement a bootstrap procedure. To formally test the null hypothesis $R^2 = 0$, we use the van de Wiel, Berkhof, and van Wieringen (2009) test, which takes into account the randomness in splitting the observations into the training and testing samples.

The paper contains several empirical results. Most importantly, we find that the non-market SDF components of all considered models are almost unrelated to macroeconomic shocks: the obtained R^2_{os} are low and statistically insignificant. The orthogonalization of the non-market components to the market further weakens the results: R^2_{os} become close to zero and even negative.

A similar conclusion holds for individual non-market factors: none of them produces R^2_{os} larger than 5%. Given that individual macroeconomic variables are known to be only weakly related to stock returns, it is not surprising that the obtained R^2_{os} are substantially below 1. However, it is remarkable that *all* considered macroeconomic shocks taken together can explain *none* of the variation in the non-market SDF components. This finding suggests that the empirical success of the characteristics-based models is produced by their ability to identify behavioral factors in stock returns rather than to describe macroeconomic risks.

To demonstrate that our results are not driven by the specifics of our analysis, we conduct a battery of robustness tests. We find that our conclusions hold for alternatively estimated SDFs, for alternatively defined macroeconomic shocks, for alternative compositions and sizes of training and testing samples, and for an alternative specification of the elastic net regression. In all cases, we obtain consistently low R^2_{os} and high p-values of the van de Wiel, Berkhof, and van Wieringen (2009) test. Given that the standard errors of R^2_{os} do not exceed a few percentage points, the results cannot be explained by a low statistical power of our analysis.

Although our list of 120 macroeconomic shocks is comprehensive, several theory-motivated shocks that presumably determine asset prices are missing from it. In an additional set of tests, we explore how the characteristics-based SDFs are related to the consumption growth shock and the shock to the intermediary capital ratio; the latter together with the market return constitutes the SDF of the intermediary asset pricing model (e.g., Adrian, Etula, and Muir, 2014; He, Kelly, and Manela, 2017). We find that consumption growth is unrelated to the SDFs and their components, and this result is consistent with the inability of the standard consumption-based asset pricing model to explain the cross section of stock returns. In contrast, the intermediary capital ratio is weakly related to the non-market SDF components of several models, and its explanatory power can be slightly amplified by adding all other macroeconomic shocks. The latter observation suggests that the intermediary capital ratio does not capture all pricing information contained in the macroeconomic factors. However, the intermediary capital ratio appears to be unrelated to the orthogonalized non-market SDF component, and the result does not change when all other macroeconomic factors are also included.

Two comments are in order. First, our results do not imply that macroeconomic risks are completely irrelevant for the cross section of stock returns. We only argue that the non-market components of characteristics-based SDFs are unrelated to macroeconomic shocks. It is possible that the characteristics-based models and macrofinance models describe different determinants of asset prices and complement each other's abilities to explain expected stock returns. Second, we do not take a stand on whether the ICAPM holds and the characteristics-based factors represent shocks to investment opportunities. On the one hand, we include all available macroeconomic shocks, not only those that are associated with state variables that track investment opportunities, and allow data to speak. On the other hand, we exclude the variables like the price-dividend ratio that are known as predictors of stock returns but that are not directly associated with interpretable macroeconomic risks. Those variables reflect all changes in expected returns and expected cash flows and can be driven by behavioral factors rather than macroeconomic shocks.

Our paper is closely related to the studies that aim to provide an economic interpretation to the characteristics-based asset pricing factors. For example, Liew and Vassalou (2000) find that HML and SMB predict the GDP growth. Vassalou (2003) shows that news about the future GDP growth subsume the pricing power of the HML and SMB factors. Petkova (2006) relate HML and SMB to innovations in several variables that predict investment opportunities. The conclusions regarding the momentum factor are more controversial. Griffin, Ji, and Martin (2003) demonstrate that momentum returns are unrelated to macroeconomic factors in 17 markets. In contrast, Liu and Zhang (2008) find that the growth rate of industrial production explains more than half of momentum profits. Aretz, Bartram, and Pope (2010) extended that line of research by examining simultaneously HML, SMB, and momentum factors and considering a larger set of six macroeconomic variables. Maio and Santa-Clara (2012) take a broader view on the interpretability of asset pricing models and argue that many multifactor models do not satisfy the restrictions imposed by the ICAPM. That conclusion is challenged by Boons (2016). Our paper differs from those studies in three ways. First, our objective is to interpret the whole non-market component of each SDF, not the individual factors. Second, our analysis involves a much larger set of macroeconomic shocks. Third, we simultaneously consider several models including the recent ones whose relations to macroeconomic factors have not been studied at all.

Our paper also belongs to a rapidly growing literature that applies machine learning techniques to various asset pricing problems such as extracting information about future returns from a large number of variables and constructing optimal portfolios. For example, LASSO and its modifications have been used to predict country-level stock returns (e.g., Rapach, Strauss, and Zhou, 2013), industry returns (e.g., Rapach, Strauss, Tu, and Zhou, 2019), high-frequency returns (e.g., Chinco, Clark-Joseph, and Ye, 2019), and the cross section of stock returns (e.g., Han, He, Rapach, and Zhou, 2019). Freyberger, Neuhierl, and Weber (2020) use grouped LASSO to select firm characteristics that provide incremental information about the cross section of stock returns and combine them nonparametrically. DeMiguel, Martin-Utrera, Nogales, and Uppal (2020) exploit the ability of LASSO to identify the characteristics that are incrementally important for constructing optimal portfolios in the presence of transaction costs. More sophisticated machine learning tools such as regression trees and neural networks have also been used for measuring conditional risk premia on stocks (e.g., Moritz and Zimmermann, 2016; Messmer, 2017; Rossi, 2018; Gu, Kelly, and Xiu, 2020b) and bonds (e.g., Huang and Shi, 2019; Bianchi, Büchner, and Tamoni, 2020; Feng, Fulop, and Li, 2020). Borochin and Zhao (2020) apply machine learning to forecasting implied volatilities of individual stocks.

Another set of papers apply machine learning methods to constructing asset pricing models. Bryzgalova (2016) develops shrinkage-based estimators of linear factor models that are robust to the presence of spurious factors. Kozak, Nagel, and Santosh (2020) and Kozak (2019) propose an estimation of SDF that incorporates information from a large number of firm characteristics and features the regularization that is similar to the elastic net regularization. Feng, Giglio, and Xiu (2020) demonstrate how double-selection LASSO can be used for evaluating the contribution of a new asset pricing factor relative to a high-dimensional set of existing factors. Gu, Kelly, and Xiu (2020a) and Feng, Polson, and Xu (2020) use neural networks to build asset pricing models that nonlinearly incorporate information from a large number of covariates. Chen, Pelger, and Zhu (2020) construct the SDF by applying deep neural networks to the conditional information from both firm characteristics and macroeconomic state variables. In contrast to those studies, we use machine learning to interpret existing asset pricing models, not to propose new models or predict returns.

The rest of the paper is organized as follows. Section 2 introduces the main idea of the paper and our empirical methodology. In particular, it describes our estimation procedure and how it helps to assess the ability of macroeconomic shocks to explain SDFs. Section 3 introduces the characteristics-based asset pricing models and their SDFs, as well as the macroeconomic variables used in the empirical analysis. Section 4 presents the main empirical results. Section 5 contains robustness tests. Section 6 concludes the paper.

2. Empirical framework

In this section, we describe the main idea of our study and introduce machine learning tools used in the paper. We also conduct simulations that illustrate the ability of our approach to produce an estimator of population R^2 that is much less biased than the OLS regression R^2.

2.1. SDFs and their components

According to the standard textbook argument (e.g., Cochrane, 2005), the law of one price implies that there exists an SDF m_t such that for excess returns on any asset R_t^e,

$$E(m_t R_t^e) = 0. \tag{1}$$

Any asset pricing model is characterized by its SDF. The most popular class of asset pricing models consists of linear factor models whose SDFs are linear functions of several factors:

$$m_t = 1 - b'(f_t - E(f_t)). \tag{2}$$

Here f_t is a vector of factor realizations at time t, and b is a vector of risk prices. Equation (1) is invariant to rescaling of the SDF by a constant nonzero factor, so without loss of generality the mean of the SDF in equation (2) is normalized to 1.

A popular way to construct asset pricing models is to use the market return and returns on long-short strategies based on various firm characteristics as the pricing factors. Such models are known as the characteristics-based factor models, and the examples include the models of Fama and French (1993), Fama and French (2015), Hou, Xue, and Zhang (2015), Stambaugh and Yuan (2017), Barillas and Shanken (2018), and Daniel, Hirshleifer, and Sun (2020). By construction, all factors in the characteristics-based models are returns on tradable assets, so the SDFs are uniquely defined.

Isolating the market factor, the SDF from equation (2) can be decomposed into the market component m_t^M and the component that combines all other factors m_t^N:

$$m_t = 1 - \underbrace{b_m(\mathrm{MKT}_t - E(\mathrm{MKT}_t))}_{m_t^M} - \underbrace{b'_{\tilde{f}}(\tilde{f}_t - E(\tilde{f}))}_{m_t^N}, \tag{3}$$

where MKT_t and $\tilde{f}_t$ are the market factor and the vector of non-market factors, respectively, and b_m and b_f are the corresponding prices of risk. Note that even though asset pricing models are typically described in terms of the individual factors $\tilde{f}_t$, only the combination of the factors m_t^N is relevant for pricing the assets. Thus, any two sets of factors that produce the same m_t^N should be interpreted as representing the same asset pricing model. Below we mostly consider m_t and m_t^N, which determine the pricing ability of the model, not individual factors. The latter are often inspired by asset pricing anomalies and may be hard to interpret.

There is an alternative way to decompose the SDF into the market and non-market components. Many individual characteristics-based factors are correlated with the market factor, and the correlation is likely to be inherited by m_t^N. Noting that m_t^N can always be represented as $m_t^N = b_N(\mathrm{MKT}_t - E(\mathrm{MKT}_t)) + m_t^\perp$, where $m_t^\perp$ is orthogonal to MKT_t, we can rewrite the SDF as

$$m_t = 1 - (b_m + b_N)(\mathrm{MKT}_t - E(\mathrm{MKT}_t)) - m_t^\perp. \tag{4}$$

In this representation, $m_t^\perp$ represents the factors whose pricing ability is unrelated to the market.

2.2. Main idea

The main objective of our paper is to identify the economic content of characteristics-based asset pricing models by examining their SDFs and the SDF components m_t^N and $m_t^\perp$. In particular, we ask to which extent the variation in m_t, m_t^N, and $m_t^\perp$ can be explained by macroeconomic shocks. Because the ability of macroeconomic factors to explain m_t and m_t^N can result from the correlation between the factors and market returns, we base our main conclusions on the evidence for $m_t^\perp$.

To quantify the ability of macroeconomic shocks to explain m_t, m_t^N, and $m_t^\perp$, it is natural to consider a linear regression model with one of those variables as a dependent variable and shocks as independent variables. By definition, the fraction of the variation in the dependent variable explained by macro shocks is measured by the population R^2, which is our main object of interest. However, the number of potentially relevant shocks is large, and the OLS regression would suffer from the overfitting problem. As a result, the standard OLS R^2 would be a severely upward biased estimator of the population R^2: it can be large even when the shocks are unrelated to the SDF and its components.[1]

[1] The adjusted R^2 is not much better: the adjustment corrects for finite sample biases in the estimates of total variation and unexplained variation that determine R^2 but not for model overfitting.

To construct a better estimator of the population R^2 in the presence of many explanatory variables, we propose an alternative procedure, which is motivated by the machine learning framework. The procedure is implemented through the following steps: (i) split the whole sample into two subsamples referred to as training and testing samples, (ii) find the best explanatory model for the dependent variable using only the training sample and penalizing the model complexity, (iii) construct the estimator of the population R^2 as the out-of-sample R^2 of the obtained model in the testing sample. The construction of the best explanatory model in the second step ensures that all relevant information from the shocks is used and R^2 is not underestimated. The penalty imposed on the model complexity reduces overfitting thereby decreasing the model's explanatory power in the training sample but increasing it in the testing sample.[2] The use of an independent testing sample to assess the explanatory power of the model further minimizes the concerns about overfitting and resulting upward bias in the estimator for R^2. The simulations conducted in Section 2.7 demonstrate good performance of the obtained estimator in a sample commensurate with our empirical sample of SDFs and macroeconomic shocks.

2.3. Training and testing samples

In our empirical analysis, we assign two thirds of the time periods to the training sample and the rest of them to the testing sample. The assigning process is deterministic: every two subsequent periods of the training sample are followed by a period assigned to the testing sample. This procedure ensures that all years are equally represented in both samples, and the results are not driven by special periods such as recessions, crisis years, etc. Note that because our objective is to explain the target variable by contemporaneous shocks, not to predict it, the alternation of training and testing periods over time does not create a look-ahead bias. As robustness tests, we also implement the versions of our estimation procedure with alternative sizes of the training and testing samples and alternative allocation of observations to them.

2.4. Elastic net

As the explanatory model, we use the elastic net regression (Zou and Hastie, 2005). It is a linear regression with a regularization that penalizes a large number of non-zero regression slopes and

[2] Mitigation of the overfitting problem by using various regularization schemes that restrict the complexity of the fitted model is a standard machine learning method. See Hastie, Tibshirani, and Friedman (2016) for a textbook introduction to machine learning.

substantial heterogeneity among them. Denote the realization of the target variable at time t as y_t (it is m_t, m_t^N, or $m_t^\perp$ in our analysis) and the realizations of Q explanatory variables as x_{it}, $i = 1, \ldots, Q$ (they are macroeconomic shocks in our analysis). Each x_{it} is assumed to be demeaned and standardized using its mean and standard deviation in the training sample. The elastic net-based predictor $\hat{y}_t$ is constructed as a linear combination $\hat{y}_t = \hat{\beta}_0 + \sum_{i=1}^{Q} x_{it}\hat{\beta}_i$, where the intercept $\hat{\beta}_0$ and slopes $\hat{\beta}_i$, $i = 1, \ldots, Q$, minimize the mean-squared error (MSE) augmented with a regularization term:

$$(\hat{\beta}_0, \ldots, \hat{\beta}_Q) = \underset{\beta_0,\ldots,\beta_Q}{\operatorname{argmin}} \left[\underbrace{\frac{1}{T} \sum_{s=1}^{T} \left(y_s - \beta_0 - \sum_{i=1}^{Q} \beta_i x_{is} \right)^2}_{\text{mean-squared error}} + \underbrace{\lambda\alpha \sum_{i=1}^{Q} |\beta_i|}_{L^1 \text{ penalty}} + \underbrace{\lambda(1-\alpha) \sum_{i=1}^{Q} \beta_i^2}_{L^2 \text{ penalty}} \right]. \tag{5}$$

The objective function in equation (5) has two additional parameters (tuning parameters). The parameter $\lambda \geq 0$ controls the tightness of the regularization: when $\lambda = 0$, the regularization disappears, the objective function in (5) reduces to the MSE, and the elastic net reduces to the OLS regression. The parameter $\alpha \in [0, 1]$ determines the relative importance of the two components of the regularization term, which are associated with LASSO (L^1 penalty) and ridge regression (L^2 penalty), respectively.

When $\alpha = 1$, elastic net reduces to LASSO (Tibshirani, 1996). Because LASSO minimizes the MSE on a polytope, corner solutions are very typical: from multiple predictors LASSO usually selects only few, and the slopes of the others are zero. In particular, if there is a group of several highly correlated predictors, LASSO tends to pick only one predictor from the group, and the information contained in the others is completely ignored. As a result, LASSO may demonstrate suboptimal prediction performance when predictors are highly correlated.

When $\alpha = 0$, elastic net reduces to the ridge regression (Hoerl and Kennard, 1970). Because of the L^2 penalty, the ridge regression shrinks all parameters toward zero and each other without dropping any predictor. On the one hand, the use of all predictors allows the ridge regression to better exploit the information from highly correlated predictors than LASSO does. On the other hand, the ridge regression suffers from useless predictors that would be discarded by LASSO. By incorporating both the L^1 and L^2 penalties, elastic net combines the advantages of LASSO and ridge regression: it can simultaneously produce a parsimonious model like LASSO and efficiently handle correlated predictors like the ridge regression.

There are several reasons why elastic net is an appropriate tool for our analysis of discount factors. First, in contrast to more flexible machine learning techniques such as regression trees, random forests, and neural networks, it predicts the target variable by linear combinations of individual predictors. Therefore, it preserves the structure of linear factor models. Second, the macroeconomic shocks are correlated within several groups, and the ability of elastic net to efficiently aggregate information from correlated predictors reduces the chance that we underestimate the ability of macroeconomic shocks to explain discount factors. Third, many macroeconomic shocks are likely to be irrelevant for asset prices, and the ability of elastic net to discard weak predictors likely diminishes the amount of noise in the estimated model parameters.

To choose the elastic net regularization parameters λ and α, we use the cross-validation, which is another standard technique in machine learning. First, we randomly split the training sample into ten equal parts (folds).[3] Second, we choose a grid in the space of the parameters (λ, α) with 100 values of λ and 10 values of α. Third, for each combination of the parameters we compute the cross-validated MSE. To do that, we sequentially exclude the fold k, $k = 1, 2, ..., 10$ and solve the optimization problem (5) on the remaining nine folds. For each observation s from the training sample, denote by $\hat{\beta}_0^{-s}(\lambda, \alpha)$ and $\hat{\beta}_i^{-s}(\lambda, \alpha)$ the results of the minimization on all folds excluding the one containing s. The cross-validated MSE is defined as

$$MSE^{CV}(\lambda, \alpha) = \frac{1}{T} \sum_{s=1}^{T} \left(y_s - \hat{\beta}_0^{-s}(\lambda, \alpha) - \sum_{i=1}^{Q} \hat{\beta}_i^{-s}(\lambda, \alpha) x_{is} \right)^2 . \tag{6}$$

By construction, $MSE^{CV}(\lambda, \alpha)$ uses all observations from the training sample, but for each observation the prediction is out of sample. Finally, we choose λ and α that minimize the cross-validated MSE and use them for estimating the parameters of the model in the whole training sample.

2.5. Out-of-sample R^2 and its standard errors

To estimate the population R^2 that measures the explanatory power of macroeconomic shocks, we use R^2 of the constructed predictor computed for the testing period as

$$R_{os}^2 = 1 - \frac{\sum_{t \in \{\text{testing period}\}} (y_t - \hat{y}_t)^2}{\sum_{t \in \{\text{testing period}\}} (y_t - \bar{y})^2},$$

[3] The decision about the number of folds involves the trade-off between the variance and bias of the constructed predictor. The five- or ten-fold cross-validation is recommended as a good compromise (Tibshirani, 1996).

where y_t is either m_t, or m_t^N, or $m_t^\perp$, $\bar{y}$ is its training sample average, and $\hat{y}_t$ is the constructed predictor. We will refer to such R^2 as out-of-sample R^2. By construction, the observations from the testing period are not involved in the estimation of the model coefficients, so R_{os}^2 does not suffer from the upward bias produced by overfitting, and in small samples it is likely to be a much better estimator of the population R^2 than the standard in-sample R^2. Note that in contrast to the in-sample R^2, R_{os}^2 can be negative.

The obtained R_{os}^2 is only an estimate of the population R^2, and it is affected by two types of errors. First, the training sample is finite, so the model coefficients are estimated with an error, which depends on the realization of the training sample. As a result, even having an infinite testing sample, it is impossible to uncover the population R^2; we can get only a very precise estimate of it produced by the estimated model. Second, the testing sample is finite, so even for a model with known coefficients we can estimate the population R^2 only with an error, which depends on the realization of the testing sample. To find the dispersion of the estimated R_{os}^2 produced by the randomness of training and testing samples, we implement the following bootstrap-type procedure.

First, we construct a bootstrapped training sample by randomly drawing with replacement the periods from the original training sample, and similarly draw a bootstrapped testing sample using only the observations from the original testing sample. By construction, the bootstrapped training and testing samples never overlap. Then, using the bootstrapped training sample, we estimate the linear model coefficients by elastic net with the regularization parameters obtained by ten-fold cross-validation on the empirical sample. Note that the bootstrapped training sample is likely to contain repeating observations, and using the fixed regularization parameters allows us to avoid conducting cross-validation with the same observations in the estimation and validation folds. The one-time tuning of elastic net also substantially reduces the computational burden of the procedure. Finally, we compute R_{os}^2 of the estimated model on the bootstrapped testing sample. Repeating this procedure 1,000 times, we obtain a sample of simulated R_{os}^2. As usual, the standard errors of the empirical R_{os}^2 are estimated as sample standard deviations of simulated R_{os}^2.

We also implement two modifications of this procedure. In the first modification, we bootstrap only the training sample and use the empirical testing sample to find R_{os}^2. In the second modification, we bootstrap only the testing sample and use the empirical training sample to estimate the model parameters. Those modifications allow us to separately estimate how the randomness of the training sample and randomness of the testing sample contribute to the distribution of R_{os}^2.

2.6. Testing

Along with the computation of standard errors for R^2_{os}, we formally test the hypothesis that the macroeconomic shocks are totally unrelated to the SDFs, that is, the population R^2 is zero. The latter implies that the explanatory power of the best model with macroeconomic shocks is equal to the explanatory power of the SDF average. To construct the test, we employ the approach proposed by van de Wiel, Berkhof, and van Wieringen (2009). Its main idea is to consider the distributions of the squared errors $(y_t - \hat{y}_t)^2$ and $(y_t - \bar{y})^2$ on the testing sample and state the null hypothesis as H_0: $(y_t - \bar{y})^2 - (y_t - \hat{y}_t)^2$ are symmetrically distributed around zero.[4] The hypothesis is tested by the one-sided Wilcoxon signed-rank test: H_0 is rejected when the test statistic exceeds the critical value obtained from the asymptotic distribution of the Wilcoxon signed-rank statistic under the null.

Although this procedure provides a valid p-value for the given split of the whole sample into the training and testing samples, it might be sensitive to how the split is conducted, and this fact invalidates the inference.[5] To deal with this problem, van de Wiel, Berkhof, and van Wieringen (2009) propose to consider multiple random splits of the data into training and testing samples and use the median of the obtained p-values for the ultimate inference. In our empirical analysis, we implement this approach with 100 random splits: for each split, we train elastic net on the training sample, compute $(y_t - \hat{y}_t)^2$ and $(y_t - \bar{y})^2$ on the testing sample, and obtain the p-value of the Wilcoxon signed-rank test. The null hypothesis is rejected if the median of the p-values is below the nominal size of the test.

2.7. Simulations

To illustrate the ability of the proposed procedure to estimate the population R^2 of linear regression models with different numbers of regressors, we conduct a simulation analysis. We fix the population R^2 at the selected level R^2_p and produce N samples with T time periods generated by a linear regression model with Q regressors. Then, in each sample, we find R^2_{os} as described above. The mean and dispersion of the simulated R^2_{os} characterize the ability of our procedure to estimate R^2_p. For consistency with our empirical analysis in Section 4, we set $T = 700$ (this is approximately the length of our empirical sample) and consider Q in the range from 1 to 120 (the latter is the number

[4]Obviously, this condition implies that $E(y_t - \bar{y})^2 = E(y_t - \hat{y}_t)^2$.

[5]The general testing procedures that eliminate the sensitivity of inference to data splitting are discussed by Romano and DiCiccio (2019).

of macro shocks). We separately explore how our estimation works for $R_p^2 = 0$ and $R_p^2 = 0.2$. To maintain a reasonable balance between the precision of simulated results and their computational feasibility, we set $N = 200$.

To generate each sample, we (i) simulate Q predictors x_{qt}, $q = 1, \ldots, Q$, with the observations at $t = 1, \ldots, T$ by randomly drawing each realization of x_{qt} from the standard normal distribution, (ii) simulate b_q, $q = 1, \ldots, Q$, as random draws from the standard normal distribution, (iii) define the population slopes of the linear regression model as

$$\beta_q = \frac{b_q \sqrt{R_p^2}}{\sqrt{\sum_{q=1}^{Q} b_q^2}}, \tag{7}$$

and (iv) simulate the target variable y_t as

$$y_t = \sum_{q=1}^{Q} x_{qt} \beta_q + u_t, \tag{8}$$

where u_t are randomly drawn from the normal distribution with the mean 0 and standard deviation $\sqrt{1 - R_p^2}$. The procedure implies that each simulated sample is produced by a different model with different coefficients β_q. However, equations (7) and (8) ensure that the population R^2 of each model is R_p^2:

$$R^2 = \frac{Var\left(\sum_{q=1}^{Q} x_{qt}\beta_q\right)}{Var(y_t)} = \frac{\sum_{q=1}^{Q} \beta_q^2}{\sum_{q=1}^{Q} \beta_q^2 + Var(u_t)} = \frac{R_p^2}{R_p^2 + 1 - R_p^2} = R_p^2.$$

Moreover, because $\sum_{q=1}^{Q} \beta_q^2 = R_p^2$ and b_q are independent standard normal random variables, β_q are uniformly distributed on the $(Q-1)$-dimensional sphere with the radius $\sqrt{R_p^2}$ (e.g., Muller, 1959). Therefore, all models with the given R^2 appear with equal probabilities, and it is unlikely that the results are driven by particular types of them (for example, by models with sparsity).

For each simulated sample, we apply our estimation procedure and compute R_{os}^2. To compare the in- and out-of-sample performance of the estimated model, we also compute the model's in-sample R^2 (R_{is}^2) on the training sample. In addition, we consider the estimates of R^2 (both R_{is}^2 and R_{os}^2) when the elastic net regression is replaced with the standard OLS regression.

[FIGURE 1 IS HERE]

Figure 1 shows the averages of the simulated R^2_{os} and R^2_{is} as functions of the number of the regressors Q. We also plot the corresponding 95% confidence bands. Consider first the case with $R^2_p = 0$ presented in Panel A. When there are only few regressors in the model, the OLS regression produces accurate results: both R^2_{os} and R^2_{is} are close to zero. However, as Q increases, R^2_{os} and R^2_{is} diverge: average R^2_{is} increases and for $Q = 120$ becomes close to 30%, whereas average R^2_{os} decreases and approaches -40%.[6] This pattern clearly demonstrates the inability of the OLS R^2 to estimate the population R^2 when the number of regressors is large compared to the number of observations. The increasing R^2_{is} shows that the OLS regression overfits the data as the flexibility of the model (the number of the parameters) increases. However, the parameter estimation adds only noise to the fitted values out of sample. As a result, they explain the target variable even worse than the average $\bar{y}$ (which is very close to zero in our case), and the average R^2_{os} becomes negative.[7] Moreover, the amount of noise grows with Q, and starting from approximately 60 regressors, the probability to get a positive R^2_{os} is less than 2.5%.

The results are drastically different for the elastic net regression. In this case, both R^2_{os} and R^2_{is} are very close to zero for all numbers of regressors, and zero always lies within the 95% confidence band of R^2_{os}. Intuitively, elastic net suppresses the estimation noise by forcing many coefficients to be zero and shrinking others toward zero. As a result, elastic net provides much more reliable estimates of the population R^2 than the OLS regression when $R^2_p = 0$, and this fact justifies its use in our empirical analysis.

There might be a concern that the case $R^2_p = 0$ is special, and the results are different when the regressors can at least partially explain the variation in the target variable. To alleviate it, we also consider the case with $R^2_p = 0.2$ and present the results in Panel B of Figure 1. As before, the average R^2_{is} and R^2_{os} produced by the OLS regression are close to R^2_p for small Q but move away from R^2_p as Q increases: R^2_{is} increases, and for $Q = 120$ it is slightly above 40%. In contrast, R^2_{os} decreases, and for $Q = 120$ it drops below -10%. The results for the elastic net regression are starkly different: it produces R^2_{os} and R^2_{is} that are close to 0.2 for all Q, although both estimators are slightly biased and underestimate R^2_p when Q is large. Note that due to the noise in the estimated model coefficients, the dispersions of R^2_{os} and R^2_{is} are larger than in the case with $R^2_p = 0$. However, they are still moderate and smaller than that produced by R^2_{os} of the OLS

[6] We have also computed the adjusted in-sample R^2 and obtained the results that are very similar to those produced by R^2_{is}.

[7] This intuition is justified more formally by Hansen and Timmermann (2015).

regression. More importantly, the confidence bands are tight enough to exclude zero unless Q is really large, but even in those cases the probability to get low R^2_{os} is small. Thus, the inference based on R^2_{os} is likely to be powerful enough to detect even weak ability of macroeconomic shocks to explain SDFs and their components.

3. Asset pricing models and macroeconomic shocks

3.1. Characteristics-based asset pricing models

In our empirical analysis, we explore whether macroeconomic shocks can explain the SDFs of seven characteristics-based asset pricing models. The selected models are listed below.

- FF3: Fama-French three-factor model (Fama and French, 1993) with the factors MKT, SMB, and HML;
- FF5: Fama-French five-factor model (Fama and French, 2015) with the factors MKT, SMB, HML, RMW, and CMA;
- Q4: q-factor model (Hou, Xue, and Zhang, 2015) with the factors MKT, ME, IA, and ROE;
- BS6: Barillas-Shanken six-factor model (Barillas and Shanken, 2018) with the factors MKT, SMB, $\mathrm{HML}^{\mathrm{m}}$, IA, ROE and UMD;
- M4: mispricing four-factor model (Stambaugh and Yuan, 2017) with the factors MKT, SMB, MGMT, and PERM;
- BF3: behavioral three-factor model (Daniel, Hirshleifer, and Sun, 2020) with the factors MKT, FIN, and PEAD;
- KNS: multi-characteristic model of Kozak, Nagel, and Santosh (2020).

Although the list of the considered models is not exhaustive, it covers a broad spectrum of the models proposed in the literature. Along with the classic FF3 model, it includes its recent popular modifications such as FF5 and Q4, which highlight the importance of the profitability and investment factors. In the BS6 model, the factors are selected from a pool of prominent factors using the Bayesian approach. The BF3 model features behavior factors. In the M4 model, the factors are constructed from many characteristics. The KNS model makes a step further and directly specifies the SDF rather than individual factors.

The construction of all factors is described in Appendix A. The data on the market factor, SMB, HML, RMW, CMA, and UMD have been downloaded from Ken French's data library at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. The factors of the q-factor model are available at http://global-q.org/factors.html. The $\mathrm{HML}^{\mathrm{m}}$ factor is provided by Andrea Frazzini at http://people.stern.nyu.edu/afrazzin/data_library.htm. Robert Stambaugh's website http://finance.wharton.upenn.edu/∼stambaug/ is the source of the factors MGMT and PERM. The factors FIN and PEAD have been downloaded from Lin Sun's website https://sites.google.com/view/linsunhome. The non-market SDF of the Kozak, Nagel, and Santosh (2020) model has been computed using the code provided by Serhiy Kozak on his website https://sites.google.com/site/serhiykozak/. All variables are measured at the monthly frequency.

3.2. Estimation of the SDFs and their non-market components

The starting point of our analysis is the SDFs of the selected characteristics-based asset pricing models. Because the characteristics-based factors are returns on tradable portfolios, linear combinations of them are also tradable, and the SDFs constructed from them uniquely characterize the models. However, for all models except KNS the SDFs should be estimated.[8] To do that, we first estimate the risk prices b_m and $b_{\tilde{f}}$ from equation (3) by GMM using the factors themselves as test assets whose returns must be explained by the model. This is the minimal and most conservative choice of test assets that produces a sufficient number of moment conditions. Although the inclusion of other assets would increase the precision of the GMM estimates, it would also increase the chance that some moment conditions are misspecified. Indeed, none of the six asset pricing models for which we construct SDFs can perfectly price all tradable assets, and including anomalous portfolios in the estimation may distort the results. Moreover, by having a just-identified set of moment conditions, we avoid the problem of choosing the GMM weighting matrix.

[TABLE 1 IS HERE]

The GMM estimation results are reported in Table 1. As expected, almost all estimated risk

[8] Kozak, Nagel, and Santosh (2020) do not explicitly include the market factor in the SDF but orthogonalize all factors with respect to the market return instead and directly obtain $m_t^{\perp}$, which is the main focus of our analysis.

prices $\hat{b}$ are positive and statistically different from zero. The notable exception is the HML factor in the FF5 model, whose risk price is negative and statistically insignificant. This result is consistent with Fama and French (2015), who note the redundancy of the HML factor in the presence of the RMW and CMA factors. Another observation is that the risk prices associated with all factors included in multiple models (MKT, SMB, HML) vary across the models. This is a consequence of non-trivial correlations between the factors.

Having the estimated risk prices, we construct the SDFs m_t and their non-market components m_t^N according to equations (2) and (3), in which the population values of b are replaced with the estimates $\hat{b}$, and the factor expectations are replaced with the factor averages. The non-market components $m_t^{\perp}$ are obtained as residuals from the regressions of m_t^N on the market returns. The exception is the KNS model, for which $m_t^{\perp}$ is directly available. The summary statistics for m_t, m_t^N, and $m_t^{\perp}$ are reported in Table 2.

[TABLE 2 IS HERE]

Panel A of Table 2 shows that the SDFs of the considered asset pricing models are substantially correlated. The positive correlations are not surprising because all SDFs contain the market factor, and many of them share identical or highly correlated other factors. Also, the ordering of the correlations looks reasonable: the highest correlation is observed between FF5 and Q4, which both include the size, profitability, and investment factors, whereas the lowest correlation is between FF3 and BF3, which arguably contain rational and behavioral factors, respectively.

A more interesting observation is that the correlation matrices in Panels A and B of Table 2 are almost identical, that is, the exclusion of the market component from the SDFs leaves the correlations between the SDFs almost unchanged. Thus, the similarity of the considered models goes far beyond the fact that they all include the market factor.

Because the characteristics-based factors are typically correlated with the market, the non-market components m_t^N are likely to be correlated with the market as well. Indeed, as reported in Panel B of Table 2, the market and non-market components tend to be negatively correlated, and the correlation coefficient can reach -0.5. Interestingly, m_t^N of the behavioral models M4 and BF3 tend to be stronger negatively correlated with the market than the corresponding components of the more conventional FF3, FF5, Q4, and BS6 models.

The correlations between the orthogonalized non-market components $m_t^{\perp}$ reported in Panel C of Table 2 are typically lower than their non-orthogonalized counterparts, but they are still quite high. Thus, even after removing their exposure to market shocks, the SDFs still appear to contain common information. The result is not completely surprising given that the non-market components m_t^N and $m_t^{\perp}$ are highly correlated.

Finally, Table 2 reports the volatilities of the SDFs and their components. Because the non-market component m_t^N is negatively correlated with the market, the removal of the latter has only a weak effect on the volatility of the SDFs, and the sign of the effect varies across the models. Our results also indicate that the volatilities of m_t, m_t^N, and $m_t^{\perp}$ tend to be lower for rational models (FF3, FF5, and Q4) and higher for behavioral models (M4, BF3). Because $E(m_t) = 1$, the standard deviation of m_t is the upper bound on the Sharpe ratios (Hansen and Jagannathan, 1991). According to our estimates, those bounds are not tight: even for the FF3 model the annualized $\sigma(m_t)$ exceeds 0.7, and it is close to 1.8 for the BF3 model. Those results are consistent with the general observation that the portfolios produced by factor investing are more mean-variance efficient than the market portfolio.

3.3. Macroeconomic shocks

Next, we construct a comprehensive set of macroeconomic shocks using FRED-MD, which is a large publicly available macroeconomic data set. FRED-MD is maintained by the Federal Reserve Bank of St. Louis and described in McCracken and Ng (2016). The variables contained in FRED-MD measure a broad spectrum of economic activities and can be classified in eight categories: (1) output and income, (2) labor market, (3) housing, (4) consumption, orders, and inventories, (5) money and credit, (6) interest rates and exchange rates, (7) prices, and (8) stock market. FRED-MD is updated monthly and can be downloaded from https://research.stlouisfed.org/econ/mccracken/fred-databases/. We use the data vintage from October 2020, which contains monthly realizations of 128 macro variables. We have dropped four variables (ACOGNO, ANDENOx, TWEXAFEGSMTHx, and UMCSENTx) whose values are missing in the early part of the sample period. To avoid discovering a mechanical relation between the SDFs and explanatory variables, we have also excluded the composite stock price index (S&P 500), stock price index of industrials (S&P: indust), dividend yield (S&P div yield), and price-earnings ratio (S&P PE ratio), which are directly determined by stock market prices. The remaining 120 variables are listed in Appendix B. Because many macroeconomic indicators are announced with one- or even two-month delay, we take lags of those variables so that each value is assigned to the month in which it becomes publicly available.

Next, we transform nonstationary variables into stationary ones. In particular, we remove stochastic trends by taking a first- or second-order difference. When a variable changes in relative rather than absolute terms, we difference its logarithm. Each variable is treated individually, and the appropriate transformation is chosen following McCracken and Ng (2016). Those transformations are indicated in Appendix B.[9] Finally, we winsorize the outliers that deviate from the sample median by more than ten interquartile ranges.

By definition, macroeconomic shocks are unexpected innovations in a macroeconomic variable. Because the majority of macroeconomic indicators are highly persistent, their own lags are strong predictors of future realizations. Therefore, to preserve simplicity, we construct macroeconomic shocks as AR(1) residuals. To avoid the look-ahead bias, in each period the AR(1) model is estimated using the realized observations available in that period.

The constructed set of macroeconomic shocks contains a small number of standard shocks that are frequently used in the empirical asset pricing literature, as well as many other shocks. The standard shocks include the innovations in the consumption expenditure, industrial production, oil price, the term spread, default spread, one-month Treasury Bill rate, and inflation (e.g., Chen, Roll, and Ross, 1986). The shocks to the last four variables are often interpreted as innovations in the economic state variables that predict future investment opportunities (e.g., Maio and Santa-Clara, 2012; Boons, 2016).

[FIGURE 2 IS HERE]

Figure 2 shows a heat map of absolute pairwise correlations between various macroeconomic shocks. The ordering of the shocks coincides with the ordering of macroeconomic variables in Appendix B, and a darker shade corresponds to a larger correlation. The figure shows that there are several clusters of non-trivial correlations that correspond to macroeconomic shocks from the same category. However, the vast majority of the correlations are surprisingly low. Therefore, the considered shocks are likely to represent relatively unrelated aspects of the economy, and the majority of them are not redundant when considered jointly. The latter fact justifies using all shocks as potential explanatory variables for the SDFs.

[9] McCracken and Ng (2016) is supplemented by the Matlab code that performs those transformations.

4. Empirical results

4.1. Explaining m_t, m_t^N, and $m_t^{\perp}$

Our main objective is to explore the joint ability of macroeconomic shocks to explain the variation in the SDFs and their non-market components m_t^N and $m_t^{\perp}$ of seven asset pricing models. Having the SDFs and their components obtained in Section 3.2 and macroeconomic shocks constructed in Section 3.3, we apply the machine learning-based estimation of the linear regression model R^2 described in Section 2.2. The results are reported in Table 3.

[TABLE 3 IS HERE]

Panel A of Table 3 shows that only the SDFs of the FF3 and FF5 models have detectable relations to the macroeconomic shocks: the obtained R^2_{os} are around 19% and 4% respectively, and the null hypothesis $R^2 = 0$ is reliably rejected in both cases. Although R^2_{os} of the M4 and BF3 models are also positive, they are not statistically significant. The R^2_{os} of the Q4 and BS6 models are indistinguishable from zero.

The conclusions are more uniform for the non-market components m_t^N and $m_t^{\perp}$: the macroeconomic shocks have no power to explain them, and Panels B and C of Table 3 show that this result holds for all considered asset pricing models. In particular, the obtained R^2_{os} for m_t^N are zero or negative in the case of the FF3, FF5, Q4, and BS6 models. Although, the R^2_{os} for the M4 and BF3 models are positive, the p-values of the van de Wiel, Berkhof, and van Wieringen (2009) test are high. The obtained R^2_{os} for $m_t^{\perp}$ are closer to zero than those for m_t^N, and none of them is statistically significant. Moreover, for the Q4, BS6, M4, and KNS models, R^2_{os} are exactly zero because the cross-validated elastic net does not include the macroeconomic shocks at all.

There might be a concern that the obtained high p-values result from a low statistical power of the van de Wiel, Berkhof, and van Wieringen (2009) test caused by a low precision of R^2_{os} as an estimator of the population R^2. However, this is an unlikely explanation for our results. The magnitudes of R^2_{os} are consistently low across the models for both m_t^N and $m_t^{\perp}$, which would not be the case if R^2_{os} were an inefficient estimator. To provide further evidence, we estimate the standard errors of R^2_{os} using bootstrap as described in Section 2.5. The randomness in R^2_{os} is produced by the randomness in the estimated model parameters and by the finiteness of the testing sample, and both of them are taken into account when both samples are bootstrapped. Table 3 demonstrates that the standard errors of R^2_{os} do not exceed several percentage points, and they are particularly small when R^2_{os} is close to zero. In the vast majority of the cases, zero is within one standard error from the empirical R^2_{os}.

We also consider the distributions of R^2_{os} when only either the training or testing sample is bootstrapped. Not surprisingly, the dispersions of those distributions are smaller than in the case when both samples are bootstrapped. More interestingly, the estimation and testing errors comparably contribute to the dispersion of R^2_{os}, although in the case of non-market components the standard errors when only the training sample is bootstrapped tend to be slightly larger than when only the testing sample is bootstrapped.

Along with the bootstrapped standard errors of R^2_{os}, we also report the means of the bootstrapped R^2_{os}, which themselves can be viewed as estimates of the population R^2. Table 3 shows that in the main case with bootstrapped training and testing samples, the mean R^2_{os} are almost always lower than the empirical R^2_{os}. This result is likely to be produced by the finiteness of the training sample and the noise that the parameter estimation introduces to the bootstrap-based estimator. Indeed, in all panels of Table 3, the means of the bootstrapped R^2_{os} are close to the empirical R^2_{os} when only the testing sample is bootstrapped, but they are substantially lower when only the training sample is bootstrapped.

Overall, our results indicate that the macroeconomic shocks are largely unrelated to the SDFs of the considered asset pricing models. This finding casts doubts on the possibility to attribute the empirical success of characteristics-based asset pricing models to their ability to identify priced macroeconomic risks. Instead, it suggests that the characteristics-based asset pricing factors are likely to have the behavioral nature.

4.2. Explaining individual asset pricing factors

The inability of macroeconomic shocks to explain the SDFs and their components might be caused by an imprecise estimation of the risk prices and the resulting noise in the SDFs. To partially alleviate this concern, we explore whether the macroeconomic shocks are related to the individual characteristics-based asset pricing factors by conducting the same analysis as in the previous section with the individual factors from Section 3.1 used as the left hand side variables.

[TABLE 4 IS HERE]

The results are reported in Panel A of Table 4, which shows that the highest R^2_{os} are obtained for SMB and ME (4.23% and 4.58%, respectively), and the null hypothesis $R^2 = 0$ is rejected for SMB, ME, MGMT, and FIN at the 10% confidence level. For the other factors, the results are much weaker: none of the R^2_{os} is statistically different from zero, and for two factors the R^2_{os} are negative. As in the case of the SDFs and their components, the bootstrap results demonstrate that the uncertainty in the R^2_{os} is produced both by the estimation errors in the model parameters and by the finiteness of the testing sample.

It is well known that many characteristics-based asset pricing factors are correlated with the market. Therefore, as in the case of the SDFs, we also consider orthogonalized factors constructed as residuals from the time series regression of factor returns on the market. Panel B of Table 4 shows that after the orthogonalization, R^2_{os} substantially drop for the vast majority of the factors and equals to zero for six of them. Accordingly, the p-values from testing $R^2 = 0$ do not fall below 39%. Thus, even assuming that the estimated standard errors are too large and some positive R^2_{os} in Panel A of Table 4 are not spurious, we still have to attribute those R^2_{os} to the relations between the macroeconomic shocks and market returns, not to the ability of the shocks to track the non-market parts of the factors.

To summarize, the lack of the relations between the macroeconomic shocks and SDFs is unlikely to result from an imprecise estimation of the prices of risk. Instead, it appears to be an intrinsic property of the characteristics-based asset pricing models.

4.3. Adding theory-motivated shocks

Although we consider a large number of macroeconomic shocks, there is still a chance that a shock that can provide additional information about the characteristics-based SDFs is missing. In this section, we explore the ability of theory-motivated shocks such as the consumption growth shock and the shock to the capital of financial intermediaries to explain the SDFs of the characteristics-based asset pricing models.

4.3.1. Consumption growth shock

The most popular macroeconomic shock in asset pricing is the consumption growth shock, which is the cornerstone of consumption-based asset pricing models (e.g., Breeden, 1979). To construct the shock, we use the consumption data from the National Income and Product Accounts (NIPA). Following Jagannathan and Wang (2007), we first obtain monthly nominal consumption expenditure on nondurables and services from NIPA Table 2.8.5. Then, we convert the two series into their per capita analogs using the population numbers from NIPA Table 2.6 and adjust for inflation using the corresponding price deflators from NIPA Table 2.8.4. Finally, we aggregate the obtained real per capita consumption and expenditure on nondurables and services to get a series of monthly consumption. The consumption growth is defined as the log growth rate of the obtained series.[10]

The first question we ask is whether the variation in the SDFs and their non-market components can be explained by the consumption growth alone. Because overfitting is not an issue in this case, we use the standard OLS regression to construct predictors of the target variables. The results are presented in Table 5. To save space, we report the mean and standard deviation of bootstrapped R^2_{os} only in the case when both training and testing samples are bootstrapped.

[TABLE 5 IS HERE]

Table 5 demonstrates that consumption growth is unrelated to the SDFs of the considered theories, and R^2_{os} are particularly small for $m_t^{\perp}$. This result is consistent with a poor ability of the consumption growth shock to explain the cross section of stock returns (e.g., Lettau and Ludvigson, 2001). Adding all other macroeconomic shocks reduces the majority of the p-values and even allows us to reject the hypothesis $R^2 = 0$ for m_t of the FF3 and FF5 models. However, the obtained results are almost identical to those reported in Table 3, and this observation holds for both SDFs and their non-market components. Thus, our results are insensitive to including the consumption growth shock as one of the macroeconomic factors.

4.3.2. Intermediary capital shock

Another theory-motivated macroeconomic shock is a shock to the capital of financial intermediaries. In contrast to the consumption growth shock, the exposure of returns to this shock has been

[10] There is also a consumption growth variable in FRED-MD, but it is constructed from monthly real personal consumption expenditures, which include all types of consumption and which are reported in NIPA Table 2.8.3.

shown to have explanatory power for the cross section of expected returns on various assets (e.g., Adrian, Etula, and Muir, 2014; He, Kelly, and Manela, 2017). In particular, He, Kelly, and Manela (2017) propose a two-factor model whose factors are the excess market return and the shock to the financial intermediaries' equity capital ratio. The latter is defined as the ratio of the primary dealers' aggregate market equity to the sum of their aggregate market equity and aggregate book debt. The shocks constructed as AR(1) innovations in the equity capital ratio divided by the lagged ratio constitute the intermediary asset pricing factor.[11] Because the intermediary asset pricing model and the characteristics-based asset pricing models have the excess market return as one of the factors, we only explore the ability of the intermediary asset pricing factor to explain the non-market SDF components.

As in the case of consumption growth, we first construct contemporaneous predictors for m_t^N and $m_t^{\perp}$ as fitted values from the OLS regression of those variables on the intermediary asset pricing factor alone. The results are presented in Table 6.

[TABLE 6 IS HERE]

Panel A of Table 6 demonstrates that although for all models the intermediary asset pricing factor can explain m_t^N with a positive R^2_{os}, the standard errors are large, and only for M4 the hypothesis $R^2 = 0$ is reliably rejected. As in the case of other macroeconomic shocks, the results are much weaker for $m_t^{\perp}$. As reported in Panel B of Table 6, R^2_{os} is negative for FF5, BF3, and KNS, and even when it is positive, it is not statistically different from zero. Therefore, the weak but still detectable relation between the intermediary asset pricing factor and m_t^N should be attributed to the fact that both of them are correlated with the market. After removing the exposure to the market, the SDFs of the considered models appear to be unrelated to the intermediary asset pricing factor.

Table 6 also reports the results of explaining the characteristics-based SDFs by a combination of the intermediary asset pricing factor and all other 120 macroeconomic shocks. In the case of m_t^N, the results are qualitatively and quantitatively similar to those with a single intermediary factor. In the case of $m_t^{\perp}$, they are noticeably weaker in the presence of all macroeconomic shocks, and

[11] We are grateful to Asaf Manela for making the intermediary asset pricing factors available on his website http://apps.olin.wustl.edu/faculty/manela/data.html. The sample starts in January 1970, and we adjust our sample of the SDFs and macroeconomic shocks accordingly.

for five models R^2_{os} becomes zero. This result is likely to be explained by noise created by useless factors. Indeed, the standard errors of R^2_{os} are typically higher when all macroeconomic shocks are included as explanatory variables.

To summarize, the intermediary asset pricing model and the considered characteristics-based asset pricing models describe different aspects of the cross section of expected stock returns, and their SDFs cannot be reduced to each other. Therefore, the intermediary asset pricing factor cannot be viewed as a missing macroeconomic factor capable of explaining the SDFs of the characteristics-based models.

5. Robustness tests

This section contains several additional tests that demonstrate the robustness of our conclusions to several modifications of the estimation and testing procedures used in the main analysis.

5.1. Alternative estimation of the SDFs

In the first robustness test, we consider an alternative way to construct the SDFs. In the main analysis, we estimate the prices of risk only assuming that the SDF correctly prices the returns on the factor portfolios. However, adding more assets that are presumably priced by the given model would increase the precision of the estimates. Therefore, we expand the set of the test assets by augmenting the factor portfolios with the Fama-French 25 size and book-to-market portfolios, 10 price-to-earnings portfolios, and 10 industry portfolios. The returns on the additional portfolios have been downloaded from Ken French's data library.

Having excess returns on each test portfolio R^e_t, we estimate the SDFs of all models except KNS by the two-stage GMM with the standard moment conditions

$$E[(1 - b'(f_t - E(f_t)))R^e_t] = 0.$$

As in the main analysis, we use the estimated prices of risk to construct the SDFs and their components m^N_t and $m^{\perp}_t$. Then, we apply the estimation procedure from Section 2.2 and find R^2_{os} produced by the macroeconomic shocks. The results are reported in Panel A of Table 7. To save space, we only present the results for m^N_t and $m^{\perp}_t$ but not for m_t. Also, we do not bootstrap the training and testing samples individually.

[TABLE 7 IS HERE]

Consistent with the main results from Table 3, the macroeconomic shocks are only weakly related to the non-market component m_t^N for two out of six models, and even in those cases the p-values are higher than 10%. However, the relation is almost undetectable for $m_t^{\perp}$: for one model (BF3) the empirical R^2_{os} is negative, and for all others it is identical to zero. Thus, we obtain another evidence that the inability of macroeconomic shocks to explain the characteristics-based SDFs should not be attributed to poorly estimated prices of risk.

5.2. Alternative construction of macroeconomic shocks from real-time data

Macroeconomic data that are publicly revealed by various government agencies are typically revised in the future, and FRED-MD contains the most recent information. Because of that, the macroeconomic shocks in our analysis may not adequately represent the real-time innovations in the macroeconomic variables as they are observed by investors, which potentially can explain the disconnect between macroeconomic shocks and asset prices. This concern is exacerbated by several recent studies that demonstrate the importance of using the real-time macroeconomic data for forecasting macroeconomic variables and bond returns (e.g., Faust and Wright, 2009; Ghysels, Horan, and Moench, 2018; Feng, Fulop, and Li, 2020; Huang, Jiang, Tong, and Zhou, 2020). To show that our results are insensitive to the revisions in the macroeconomic data, we repeat the analysis using real-time macroeconomic shocks.

To construct those shocks, we use the ALFRED database, which is maintained by the Federal Reserve Bank of St. Louis and downloadable from https://alfred.stlouisfed.org/. ALFRED provides time-stamped vintages of macroeconomic data and allows researchers to retrieve real-time realizations of the variables. Unfortunately, many data vintages contain different sets of the variables. To get a balanced data set of 120 macroeconomic variables in the considered time period, we use the revised data from FRED-MD to estimate the AR(1) models and compute expectations but the real-time data from ALFRED to find shocks. All macroeconomic variables are transformed as indicated in Appendix B period by period. Also, we winsorize the transformed variables in each period so that the outliers are detected and removed in real time.

We present the results in Panel B of Table 7, which shows that for both m_t^N and $m^{\perp}$ the R^2_{os} tend to be higher than their counterparts from Table 3. This observation indicates that the real-time macroeconomic shocks indeed have a better chance to explain the SDFs than the shocks based on more accurate but unavailable to investors information. Nevertheless, the magnitudes of R_{os}^2 are still very low, and only for the BS6 and KNS models the p-values drop below 5%. Even in those two cases we exercise caution and do not interpret the results as evidence of statistical significance because we simultaneously test R^2 of multiple models, and multiple testing distorts the sizes of the tests with the standard critical values.

5.3. Alternative construction of macroeconomic shocks from a FAVAR model

Because macroeconomic shocks are unexpected innovations in macroeconomic variables, their realizations depend on the chosen model for the variable expectations. In particular, the AR(1) model used in our main analysis can be too restrictive since the lagged variable represents only a small subset of information available to market participants. As a result, the macroeconomic shocks may be poorly measured and erroneously appear as unrelated to the discount factors. To alleviate this concern, we repeat the analysis with the macroeconomic shocks constructed as residuals of a factor-augmented vector autoregression (FAVAR). The FAVAR model, which has been developed by Bernanke, Boivin, and Eliasz (2005), extends the standard VAR model by augmenting it with a small number of estimated factors that concisely summarize the information from a large set of macroeconomic variables. Therefore, the FAVAR-based expectations of macroeconomic variables encompass a much larger information set and are less likely to be misspecified than the AR(1)-based expectations.

To construct the FAVAR-based expectations, we apply the two-step procedure of Bernanke, Boivin, and Eliasz (2005) to our set of 120 macroeconomic variables. As before, the variables have been transformed to induce stationarity as indicated in Appendix B. In the first step, we find the three principal components (PCs) of the variables, which beforehand have been standardized to have zero means and unit standard deviations. In the second step, we estimate the predictive regression for each (non-standardized) macroeconomic variable using its lag and the lagged realizations of the three PCs obtained in the first step as predictors.[12] The difference between the realized macroeconomic variable and the fitted value from our model represents the macroeconomic shock. To avoid the look-ahead bias, in each period, both steps of the estimation use only the observations

[12]Because our objective is only to find the expectations of the macroeconomic variables, not to estimate the whole FAVAR model for the selected observable variables and latent factors, we do not orthogonalize the factors and variables as prescribed in Bernanke, Boivin, and Eliasz (2005).

available in that period.

The results are presented in Panel C of Table 7. Compared to Table 3, the FAVAR-based macroeconomic shocks produce zero R^2_{os} for a larger number of the models in the case of m_t^N, whereas R^2_{os} for $m_t^{\perp}$ are barely affected. Thus, the ability of macroeconomic shocks to explain the non-market components of the characteristics-based SDFs cannot be improved by enlarging the information set relative to which the shocks are defined.

5.4. Alternative specification of the elastic net regression

In our main analysis, we assume that the slopes of macroeconomic shocks are constant over time. However, as recognized by conditional asset pricing models, the contribution of each shock into the SDF may depend on the state of the economy. To entertain such a possibility, we repeat our analysis augmenting the macroeconomic shocks with the interactions of all shocks with potential state variables. As those variables, we choose (1) the yield spread between ten-year Treasury bond rate and the one-month Federal Fund Rate, (2) the yield spread between BAA and AAA corporate bonds from Moody's, (3) the one-month Treasury Bill rate, (4) the inflation rate, (5) the consumption growth rate, and (6) the industrial productivity growth rate. Those variables are stationary, and they are commonly used in the conditional asset pricing models (e.g., Ferson and Harvey, 1993; Jagannathan and Wang, 1996; Ferson and Harvey, 1999). The chosen state variables are available in FRED-MD, and we construct 720 interactions of each lagged state variable with 120 macroeconomic shocks. In total, we end up with 840 explanatory variables in the elastic net regression. Because the number of regressors exceeds the number of observations (which is around 700 in our sample), the OLS regression cannot be run, which underlines the importance of using elastic net in our estimation procedure.

The results are presented in Panel D of Table 7. They demonstrate that making the slopes of macroeconomic shocks state dependent does not improve their ability to explain the SDFs' non-market components: the R^2_{os} and p-values for m_t^N and $m_t^{\perp}$ are qualitatively, and for many models even quantitatively, comparable to their counterparts in Table 3.

5.5. Alternative training and testing samples

The next robustness test explores the sensitivity of our results to the particular split of the whole sample into the training and testing samples. In the main analysis, we assign the first two months of each quarter to the training sample and the last month of each quarter to the testing sample. Because of this convention, our results may not be representative for the whole sample if there is a seasonal variation in the ability of macroeconomic factors to proxy SDFs. To demonstrate that this is not the case, we repeat the main steps of our analysis after assigning the first month of each quarter to the testing sample and the other two months to the training sample. The results are reported in Panel E of Table 7. It shows that although the obtained R^2_{os} are slightly different compared to their counterparts from Table 3, there is no discernible pattern in those differences. Most importantly, all p-values are still large for both m_t^N and $m_t^{\perp}$ and for all models. Thus, the inability of macroeconomic shocks to explain SDFs cannot be attributed to the peculiarities of our allocation of time periods to the training and testing samples.

5.6. Alternative size of the training sample

In the main analysis, two thirds of the observations are allocated to the training sample, and the rest of them are used for testing. Although such a split is standard in the machine learning literature, it may not be optimal from the perspective of maximizing the efficiency of R^2_{os} as an estimator of the population R^2. In particular, it might be possible to improve the estimation precision and the power of our tests by increasing the size of the training sample, which would possibly allow us to better estimate the elastic net coefficients without compromising the assessment of the model's predictive ability in the testing sample. To ensure that our results are not produced by a suboptimally chosen split, we consider an alternative split in which we assign four fifths of the observations to the training sample. As in the main analysis, the split is deterministic: every four subsequent periods of the training sample are followed by a period assigned to the testing sample. The results obtained in this setting are presented in Panel F of Table 7.

Compared to the results from Table 3, a larger training sample on average leads to a smaller number of models for which the cross-validation excludes all shocks and R^2 in the testing sample is exactly zero. However, the inclusion of more shocks does not improve the explanatory power of elastic net regressions: many non-zero R^2_{os} are negative for both m_t^N and $m_t^{\perp}$. Accordingly, the p-values for all models and SDF components are high, and the hypothesis $R^2 = 0$ cannot be rejected. Thus, we conclude that our main results are not produced by a suboptimal allocation of observations to the training and testing samples.

6. Conclusion

The characteristics-based factor models are popular in empirical asset pricing because they are the most successful in explaining the cross-sectional variation in expected stock returns. However, their factors lack economic interpretation, and it is unclear which economic shocks they represent. In this paper, we explore whether a large number of macroeconomic shocks can at least partially approximate the SDFs of seven representative characteristics-based asset pricing models and use machine learning tools to mitigate the overfitting problem. Our paper makes both empirical and methodological contributions to the literature. On the empirical side, we find that the relations between the SDFs of the considered models and macroeconomic shocks are tenuous at best, and where detectable, they are produced by the correlation between market returns and the shocks. This result indicates that behavioral factors are likely to be responsible for the empirical success of the characteristics-based models, not macroeconomic risks. On the methodological side, we demonstrate how machine learning techniques can be used not only for forecasting but also for measuring the explainability of one variable by many others.

Our results suggest several directions for future research. First, an important methodological question is how to construct an estimator of the population R^2 that is better than the R^2_{os} of elastic net. As our simulations demonstrate, the latter clearly outperforms the in- and out-of-sample R^2 from the OLS regression when the number of regressors is large compared to the number of observations, but R^2_{os} still underestimates positive population R^2. Second, it might be interesting to further explore and validate the bootstrap-based computation of standard errors. Third and most importantly, our results imply that the search for an economic interpretation of the characteristics-based asset pricing models is far from being over, and additional work is needed to understand the pricing ability of the characteristics-based factors.

References

Adrian, Tobias, Erkko Etula, and Tyler Muir, 2014, Financial intermediaries and the cross-section of asset returns, *Journal of Finance* 69, 2557–2596.

Aretz, Kevin, Sohnke M. Bartram, and Peter F. Pope, 2010, Macroeconomic risks and characteristic-based factor models, *Journal of Banking and Finance* 34, 1383–1399.

Back, Kerry E., 2010, *Asset pricing and portfolio choice theory* (Oxford University Press: New York, NY).

Barillas, Francisco, and Jay Shanken, 2018, Comparing asset pricing models, *Journal of Finance* 73, 715–754.

Bernanke, Ben S., Jean Boivin, and Piotr Eliasz, 2005, Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach, *Quarterly Journal of Economics* 120, 387–422.

Bianchi, Daniele, Matthias Büchner, and Andrea Tamoni, 2020, Bond risk premia with machine learning, *Working paper.*

Boons, Martijn, 2016, State variables, macroeconomic activity, and the cross section of individual stocks, *Journal of Financial Economics* 119, 489–511.

Borochin, Paul, and Yanhui Zhao, 2020, The economic value of equity implied volatility forecasting with machine learning, *Working paper.*

Breeden, Douglas T., 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, *Journal of Financial Economics* 7, 265–296.

Bryzgalova, Svetlana, 2016, Spurious factors in linear asset pricing models, *Working paper.*

Chen, Luyang, Markus Pelger, and Jason Zhu, 2020, Deep learning in asset pricing, *Working paper.*

Chen, Nai-Fu, Richard Roll, and Stephen A. Ross, 1986, Economic forces and the stock market, *Journal of Business* 59, 383–403.

Chinco, Alex, Adam D. Clark-Joseph, and Mao Ye, 2019, Sparse signals in the cross-section of returns, *Journal of Finance* 74, 449–492.

Cochrane, John H., 2005, *Asset Pricing* (Princeton University Press: Princeton, NJ).

Daniel, Kent, David Hirshleifer, and Lin Sun, 2020, Short- and long-horizon behavioral factors, *Review of Financial Studies* 33, 1673–1736.

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

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

——— , 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Faust, Jon, and Jonathan H. Wright, 2009, Comparing greenbook and reduced form forecasts using a large realtime dataset, *Journal of Business & Economic Statistics* 27, 468–479.

Feng, Guanhao, Andras Fulop, and Junye Li, 2020, Real-time macro information and bond return predictability: Does deep learning help?, *Working paper*.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu, 2020, Taming the factor zoo: a test of new factors, *Journal of Finance* 75, 1327–1370.

Feng, Guanhao, Nicholas G. Polson, and Jianeng Xu, 2020, Deep learning in characteristics-sorted factor models, *Working paper*.

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

———, 1999, Conditioning variables and the cross section of stock returns, *Journal of Finance* 54, 1325–1360.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber, 2020, Dissecting characteristics non-parametrically, *Review of Financial Studies* 33, 2326–2377.

Ghysels, Eric, Casidhe Horan, and Emanuel Moench, 2018, Forecasting through the rearview mirror: Data revisions and bond return predictability, *Review of Financial Studies* 31, 678–714.

Griffin, John M., Xiuqing Ji, and J. Spencer Martin, 2003, Momentum investing and business cycle risk: Evidence from pole to pole, *Journal of Finance* 58, 2515–2547.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020a, Autoencoder asset pricing models, *Journal of Econometrics* (forthcoming).

———, 2020b, Empirical asset pricing via machine learning, *Review of Financial Studies* 33, 2223–2273.

Hahn, Jaehoon, and Hangyong Lee, 2006, Yield spreads as alternative risk factors for size and book-to-market, *Journal of Financial and Quantitative Analysis* 41, 245–269.

Han, Yufeng, Ai He, David E. Rapach, and Guofu Zhou, 2019, Firm characteristics and expected stock returns, *Working paper*.

Hansen, Lars Peter, and Ravi Jagannathan, 1991, Implications of security market data for models of dynamic economics, *Journal of Political Econonomy* 99, 225–262.

Hansen, Peter Reinhard, and Allan Timmermann, 2015, Comment, *Journal of Business & Economic Statistics* 33, 17–21.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman, 2016, *The elements of statistical learning: Data mining, inference, and prediction* (Springer: New York, NY) 2nd edn.

He, Zhiguo, Bryan Kelly, and Asaf Manela, 2017, Intermediary asset pricing: New evidence from many asset classes, *Journal of Financial Economics* 126, 1–35.

Hoerl, Arthur E., and Robert W. Kennard, 1970, Ridge regression: Biased estimation for nonorthogonal problems, *American Statistical Association* 12, 55–67.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 650–705.

Huang, Dashan, Fuwei Jiang, Guoshi Tong, and Guofu Zhou, 2020, Are bond returns predictable with real-time macro data?, *Working paper.*

Huang, Jing-Zhi, and Zhan Shi, 2019, Determinants of bond risk premia: A machine-learning-mased resolution of the spanning controversy, *Working paper.*

Jagannathan, Ravi, and Yong Wang, 2007, Lazy investors, discretionary consumption, and the cross-section of stock returns, *Journal of Finance* 62, 1623–1661.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional capm and the cross-section of expected returns, *Journal of Finance* 51, 3–53.

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

Kozak, Serhiy, 2019, Kernel trick for the cross section, *Working paper.*

———, Stefan Nagel, and Shrihari Santosh, 2018, Interpreting factor models, *Journal of Finance* 73, 1183–1223.

———, 2020, Shrinking the cross-section, *Journal of Financial Economics* 135, 271–292.

Lettau, Martin, and Sydney Ludvigson, 2001, Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238–1287.

Liew, Jimmy, and Maria Vassalou, 2000, Can book-to-market, size and momentum be risk factors that predict economic growth?, *Journal of Financial Economics* 57, 221–245.

Liu, Laura Xiaolei, and Lu Zhang, 2008, Momentum profits, factor pricing, and macroeconomic risk, *Review of Financial Studies* 21, 2417–2448.

Maio, Paulo, and Pedro Santa-Clara, 2012, Multifactor models and their consistency with the ICAPM, *Journal of Financial Economics* 106, 586–613.

McCracken, Michael W., and Serena Ng, 2016, FRED-MD: A monthly database for macroeconomic research, *Journal of Business & Economic Statistics* 34, 574–589.

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

Moritz, Benjamin, and Tom Zimmermann, 2016, Tree-based conditional portfolio sorts: The relation between past and future stock returns, *Working paper*.

Muller, Mervin E, 1959, A note on a method for generating points uniformly on n-dimensional spheres, *Communications of the ACM* 2, 19–20.

Nagel, Stefan, 2013, Empirical cross-sectional asset pricing, *Annual Review of Financial Economics* 5, 167–199.

Petkova, Ralitsa, 2006, Do the Fama-French factors proxy for innovations in predictive variables, *Journal of Finance* 61, 581–612.

Rapach, David E., Jack K. Strauss, Jun Tu, and Guofu Zhou, 2019, Industry return predictability: A machine learning approach, *Journal of Financial Data Science* 1, 9–28.

Rapach, David E., Jack K. Strauss, and Guofu Zhou, 2013, International stock return predictability: What is the role of the United States?, *Journal of Finance* 68, 1633–1662.

Romano, Joseph P., and Cyrus DiCiccio, 2019, Multiple data splitting for testing, *Working paper*.

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

Stambaugh, Robert F., and Yu Yuan, 2017, Mispricing factors, *Review of Financial Studies* 30, 1270–1315.

Tibshirani, Robert, 1996, Regression shrinkage and selection via the Lasso, *Journal of the Royal Statistical Society* 58, 267–288.

van de Wiel, Mark A., Johannes Berkhof, and Wessel N. van Wieringen, 2009, Testing the prediction error difference between 2 predictors, *Biostatistics* 10, 550–560.

Vassalou, Maria, 2003, News related to future GDP growth as a risk factor in equity returns, *Journal of Financial Economics* 68, 47–73.

Zou, Hui, and Trevor Hastie, 2005, Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society* 67, 301–320.

Table 1
The table presents the GMM estimates of risk prices and their standard errors for six characteristics-based factor models: FF3, FF5, Q4, BS6, M4, and BF3. The models are described in Section 3.1. The samples are from July 1963 to December 2019 for FF3 and FF5, from January 1967 to December 2019 for Q4 and BS6, from July 1963 to December 2016 for M4, and from July 1972 to December 2019 for BF3.

			FF3			
	MKT	SMB	HML			
$\hat{b}$	3.43	1.55	5.68			
s.e.	1.04	1.33	1.45			
			FF5			
	MKT	SMB	HML	RMW	CMA	
$\hat{b}$	5.39	3.68	-0.63	10.01	12.97	
s.e.	1.14	1.48	2.09	2.15	3.06	
			Q4			
	MKT	ME	IA	ROE		
$\hat{b}$	5.78	5.14	16.09	12.41		
s.e.	1.19	1.64	2.71	2.06		
			BS6			
	MKT	SMB	$\mathrm{HML}^{\mathrm{m}}$	IA	ROE	UMD
$\hat{b}$	6.21	6.54	12.25	6.11	16.09	6.43
s.e.	1.29	1.77	2.56	3.63	2.46	1.77
			M4			
	MKT	SMB	MGMT	PERF		
$\hat{b}$	8.55	7.40	16.53	7.63		
s.e.	1.44	1.62	2.18	1.50		
			BF3			
	MKT	FIN	PEAD			
$\hat{b}$	8.01	10.09	20.50			
s.e.	1.45	1.50	2.52			

Table 2
The table shows the summary statistics of the SDFs m_t (Panel A), their non-market components m_t^N (Panel B), and the orthogonalized non-market components $m_t^{\perp}$ (Panel C) for seven characteristics-based factor models: FF3, FF5, Q4, BS6, M4, BF3, and KNS. The models are described in Section 3.1. The samples are from July 1963 to December 2019 for FF3 and FF5, from January 1967 to December 2019 for Q4 and BS6, from July 1963 to December 2016 for M4, from July 1972 to December 2019 for BF3, and from November 1973 to December 2019 for KNS. $\sigma(m_t)$, $\sigma(m_t^N)$, and $\sigma(m_t^{\perp})$ are the standard deviations of the corresponding variables; $\rho(m_t^M, m_t^N)$ and $\rho(m_t^N, m_t^{\perp})$ are the correlations between the corresponding components, and m_t^M is defined in equation (3). The bottom parts of the panels report the correlations of the SDFs and their components across the models.

Panel A: SDFs m_t							
	FF3	FF5	Q4	BS6	M4	BF3	KNS
$\sigma(m_t)$	0.20	0.31	0.41	0.48	0.47	0.50	–
Correlations							
FF3	1.00						
FF5	0.63	1.00					
Q4	0.43	0.74	1.00				
BS6	0.55	0.70	0.89	1.00			
M4	0.46	0.68	0.63	0.64	1.00		
BF3	0.29	0.40	0.47	0.47	0.56	1.00	

Panel B: non-market components m_t^N							
	FF3	FF5	Q4	BS6	M4	BF3	KNS
$\sigma(m_t^N)$	0.16	0.31	0.42	0.49	0.53	0.53	–
$\rho(m_t^M, m_t^N)$	-0.17	-0.37	-0.33	-0.32	-0.50	-0.44	–
Correlations							
FF3	1.00						
FF5	0.55	1.00					
Q4	0.36	0.74	1.00				
BS6	0.53	0.71	0.90	1.00			
M4	0.44	0.71	0.66	0.67	1.00		
BF3	0.24	0.45	0.50	0.51	0.64	1.00	

Panel C: orthogonalized non-market components $m_t^{\perp}$							
	FF3	FF5	Q4	BS6	M4	BF3	KNS
$\sigma(m_t^{\perp})$	0.15	0.29	0.40	0.46	0.46	0.48	0.16
$\rho(m_t^N, m_t^{\perp})$	0.99	0.93	0.94	0.95	0.87	0.90	–
Correlations							
FF3	1.00						
FF5	0.53	1.00					
Q4	0.33	0.71	1.00				
BS6	0.51	0.68	0.88	1.00			
M4	0.41	0.65	0.61	0.62	1.00		
BF3	0.18	0.34	0.43	0.44	0.53	1.00	
KNS	0.46	0.65	0.63	0.70	0.64	0.46	1.00

Table 3
This table presents R^2_{os} produced by elastic net regressions of SDFs and their components on macroeconomic shocks for seven characteristics-based factor models. The p-values correspond to the van de Wiel, Berkhof, and van Wieringen (2009) test of the null hypothesis $R^2 = 0$. The table also reports the means and standard deviations of R^2_{os} computed for bootstrapped training samples, bootstrapped testing samples, and bootstrapped both training and testing samples.

Panel A: SDFs m_t								
		FF3	FF5	Q4	BS6	M4	BF3	KNS
Empirical sample	R^2_{os},%	19.06	4.39	0.00	0.00	1.79	1.74	–
	p-value	0.00	0.05	0.72	0.44	0.28	0.47	–
Bootstrapped training sample	Mean(R^2_{os}),%	15.53	-0.68	-0.06	-0.04	-0.01	-0.03	–
	Std(R^2_{os}),%	2.99	5.44	1.15	1.60	2.53	2.55	–
Bootstrapped testing sample	Mean(R^2_{os}),%	19.10	4.51	0.00	0.00	1.82	1.60	–
	Std(R^2_{os}),%	4.58	2.53	0.00	0.00	1.07	2.52	–
Bootstrapped training and testing sample	Mean(R^2_{os}),%	15.49	-0.43	-0.02	-0.05	-0.02	-0.32	–
	Std(R^2_{os}),%	6.24	6.85	1.55	1.97	3.46	4.41	–
Panel B: non-market components m^N_t								
		FF3	FF5	Q4	BS6	M4	BF3	KNS
Empirical sample	R^2_{os},%	0.00	-6.18	-0.36	0.00	1.72	8.45	–
	p-value	1.00	0.84	0.59	1.00	0.30	0.09	–
Bootstrapped training sample	Mean(R^2_{os}),%	-0.93	-12.88	-7.13	-2.59	-1.92	6.61	–
	Std(R^2_{os}),%	2.29	7.90	5.01	2.84	4.31	3.39	–
Bootstrapped testing sample	Mean(R^2_{os}),%	0.00	-6.21	-0.38	-0.04	1.63	8.27	–
	Std(R^2_{os}),%	0.00	3.02	3.33	0.80	2.08	2.73	–
Bootstrapped training and testing sample	Mean(R^2_{os}),%	-0.90	-12.83	-7.19	-2.60	-1.92	6.26	–
	Std(R^2_{os}),%	2.88	9.44	6.87	3.48	5.59	4.92	–
Panel C: orthogonalized non-market components $m^{\perp}_t$								
		FF3	FF5	Q4	BS6	M4	BF3	KNS
Empirical sample	R^2_{os},%	0.00	-2.86	0.00	0.00	0.00	-0.04	0.00
	p-value	0.89	0.99	0.98	1.00	0.82	0.59	0.92
Bootstrapped training sample	Mean(R^2_{os}),%	0.23	-6.90	-0.66	-1.17	-0.36	-0.11	0.19
	Std(R^2_{os}),%	2.09	5.12	1.20	1.74	0.95	1.71	1.71
Bootstrapped testing sample	Mean(R^2_{os}),%	0.00	-2.79	0.00	0.00	0.00	-0.09	0.00
	Std(R^2_{os}),%	0.00	1.57	0.00	0.00	0.00	0.69	0.00
Bootstrapped training and testing sample	Mean(R^2_{os}),%	0.28	-6.65	-0.62	-1.16	-0.35	-0.30	0.17
	Std(R^2_{os}),%	2.64	5.99	1.57	2.04	1.23	2.56	2.28

Table 4
This table presents R^2_{os} produced by elastic net regressions of individual asset pricing factors (Panel A) and the factors orthogonalized to the market factor (Panel B) on 120 macroeconomic shocks. The p-values correspond to the van de Wiel, Berkhof, and van Wieringen (2009) test of the null hypothesis $R^2 = 0$. The table also reports the means and standard deviations of bootstrapped R^2_{os} computed for bootstrapped training samples, bootstrapped testing samples, and bootstrapped both training and testing samples.

Panel A: original factors

		SMB	HML	HML$^{\mathrm{m}}$	RMW	CMA	ME	IA	ROE	UMD	MGMT	PERF	FIN	PEAD
Empirical sample	R^2_{os},%	4.23	0.00	-0.18	0.64	1.03	4.58	0.08	2.41	-0.37	3.37	1.07	3.36	0.00
	p-value	0.08	1.00	0.85	0.44	0.46	0.07	0.52	0.31	0.96	0.09	0.71	0.10	0.82
Bootstrapped training sample	Mean(R^2_{os}),%	1.97	-0.14	-1.01	-0.22	-0.97	0.17	-2.96	-0.21	-1.81	2.01	0.92	0.80	0.24
	Std(R^2_{os}),%	3.43	0.59	1.76	1.32	2.07	4.01	3.09	3.49	2.13	2.09	0.77	3.23	0.37
Bootstrapped testing sample	Mean(R^2_{os}),%	4.42	0.00	-0.15	0.72	0.98	4.74	0.08	2.28	-0.35	3.35	1.02	3.35	0.00
	Std(R^2_{os}),%	5.43	0.75	2.86	2.05	3.41	6.18	4.47	5.68	3.40	3.03	1.54	4.28	0.51
Bootstrapped both samples	Mean(R^2_{os}),%	1.90	-0.13	-1.04	-0.23	-1.05	-0.01	-2.90	-0.39	-1.89	2.03	0.83	0.72	0.20
	Std(R^2_{os}),%	5.43	0.75	2.86	2.05	3.41	6.18	4.47	5.68	3.40	3.03	1.54	4.28	0.51

Panel B: orthogonalized factors

		SMB	HML	HML$^{\mathrm{m}}$	RMW	CMA	ME	IA	ROE	UMD	MGMT	PERF	FIN	PEAD
Empirical sample	R^2_{os},%	-1.63	0.00	0.41	-0.29	0.91	-1.63	-1.70	-0.59	0.00	0.00	0.00	0.00	0.00
	p-value	0.51	0.72	0.39	0.80	0.48	0.49	0.71	0.44	1.00	0.64	1.00	0.95	0.95
Bootstrapped training sample	Mean(R^2_{os}),%	-4.25	0.20	-0.47	-1.14	-0.81	-5.14	-4.54	-1.94	-0.01	0.29	0.03	-0.19	0.19
	Std(R^2_{os}),%	3.45	0.65	1.74	1.45	2.21	2.64	3.45	3.22	0.20	0.61	0.12	0.56	0.30
Bootstrapped testing sample	Mean(R^2_{os}),%	-1.60	0.00	0.46	-0.22	0.95	-1.64	-1.67	-0.65	0.00	0.00	0.00	0.00	0.00
	Std(R^2_{os}),%	4.77	0.86	2.96	2.23	3.20	4.76	4.87	4.58	0.47	0.84	0.24	0.77	0.44
Bootstrapped both samples	Mean(R^2_{os}),%	-4.43	0.22	-0.49	-1.19	-0.78	-5.49	-4.43	-2.06	-0.01	0.31	0.03	-0.21	0.15
	Std(R^2_{os}),%	4.77	0.86	2.96	2.23	3.20	4.76	4.87	4.58	0.47	0.84	0.24	0.77	0.44

Table 5
This table presents R^2_{os} produced by elastic net regressions of SDFs m_t and their non-market components m^N_t and $m^{\perp}_t$ on the consumption growth and 120 macroeconomic shocks for seven characteristics-based factor models. The p-values correspond to the van de Wiel, Berkhof, and van Wieringen (2009) test of the null hypothesis $R^2 = 0$. The table also reports the means and standard deviations of R^2_{os} computed for bootstrapped training and testing samples.

Panel A: SDFs m_t

		consumption growth							consumption growth + macro shocks						
		FF3	FF5	Q4	BS6	M4	BF3	KNS	FF3	FF5	Q4	BS6	M4	BF3	KNS
Empirical sample	R^2_{os},%	2.29	1.38	0.75	0.43	0.45	-0.25	–	20.46	4.56	0.00	0.00	1.79	1.74	–
	p-value	0.45	0.39	0.39	0.37	0.53	0.72	–	0.00	0.05	0.73	0.44	0.27	0.47	–
Bootstrapped	Mean(R^2_{os}),%	1.82	0.98	0.44	0.00	-0.08	-0.83	–	15.25	0.25	-0.01	-0.04	-0.02	-0.34	–
samples	Std(R^2_{os}),%	2.27	1.86	1.49	1.43	1.47	1.85	–	7.24	6.65	1.55	1.97	3.46	4.40	–

Panel B: non-market components m^N_t

		consumption growth							consumption growth + macro shocks						
		FF3	FF5	Q4	BS6	M4	BF3	KNS	FF3	FF5	Q4	BS6	M4	BF3	KNS
Empirical sample	R^2_{os},%	0.03	0.41	0.24	0.35	0.96	1.58	–	0.00	-6.18	-0.36	0.00	1.72	8.97	–
	p-value	0.76	0.62	0.53	0.44	0.39	0.54	–	1.00	0.84	0.58	1.00	0.29	0.10	–
Bootstrapped	Mean(R^2_{os}),%	-0.48	-0.07	-0.13	-0.12	0.37	0.89	–	-0.90	-12.80	-7.18	-2.58	-1.80	6.70	–
samples	Std(R^2_{os}),%	1.35	1.40	1.17	1.42	1.92	3.08	–	2.87	9.41	6.85	3.48	5.55	4.93	–

Panel C: orthogonalized non-market components $m^{\perp}_t$

		consumption growth							consumption growth + macro shocks						
		FF3	FF5	Q4	BS6	M4	BF3	KNS	FF3	FF5	Q4	BS6	M4	BF3	KNS
Empirical sample	R^2_{os},%	-0.07	0.04	0.06	0.01	0.00	-0.01	-0.01	0.00	-2.86	0.00	0.00	0.00	-0.04	0.00
	p-value	0.59	0.70	0.60	0.74	0.73	0.67	0.71	0.88	0.98	0.99	1.00	0.82	0.61	0.92
Bootstrapped	Mean(R^2_{os}),%	-0.54	-0.36	-0.25	-0.43	-0.53	-0.63	-0.47	0.27	-6.65	-0.62	-1.16	-0.35	-0.31	0.17
samples	Std(R^2_{os}),%	1.42	1.25	1.11	1.27	1.33	2.30	1.51	2.63	5.99	1.57	2.05	1.23	2.57	2.28

Table 6

This table presents R^2_{os} produced by elastic net regressions of the non-market components of SDFs m_t^N and $m_t^{\perp}$ on the intermediary asset pricing factor and 120 macroeconomic shocks for seven characteristics-based factor models. The p-values correspond to the van de Wiel, Berkhof, and van Wieringen (2009) test of the null hypothesis $R^2 = 0$. The table also reports the means and standard deviations of R^2_{os} computed for bootstrapped training and testing samples.

Panel A: non-market components m_t^N															
		intermediary factor							intermediary factor + macro shocks						
		FF3	FF5	Q4	BS6	M4	BF3	KNS	FF3	FF5	Q4	BS6	M4	BF3	KNS
Empirical sample	R^2_{os},%	0.25	4.81	0.46	2.00	11.18	10.27	–	0.00	3.84	1.58	2.06	12.27	11.50	–
	p-value	0.64	0.24	0.37	0.24	0.01	0.10	–	1.00	0.13	0.27	0.16	0.00	0.06	–
Bootstrapped samples	Mean(R^2_{os}),%	-0.33	4.33	0.16	1.80	10.63	9.69	–	-0.46	1.29	-0.19	-0.57	8.94	10.46	–
	Std(R^2_{os}),%	1.58	4.10	4.25	3.76	6.52	5.40	–	2.52	4.39	3.74	4.05	7.56	4.94	–

Panel B: orthogonalized non-market components $m_t^{\perp}$															
		intermediary factor							intermediary factor + macro shocks						
		FF3	FF5	Q4	BS6	M4	BF3	KNS	FF3	FF5	Q4	BS6	M4	BF3	KNS
Empirical sample	R^2_{os},%	3.13	-0.43	0.34	0.05	0.28	-0.10	-1.84	0.00	0.00	0.00	0.00	-0.93	0.60	0.00
	p-value	0.17	0.15	0.21	0.38	0.79	0.80	0.11	0.85	1.00	1.00	1.00	1.00	0.73	0.90
Bootstrapped samples	Mean(R^2_{os}),%	2.46	-0.89	-0.02	-0.30	-0.19	-0.47	-2.23	0.86	-1.19	-0.94	-1.28	-2.49	0.11	0.18
	Std(R^2_{os}),%	2.89	1.68	1.14	1.22	1.23	1.45	3.04	2.48	2.20	1.84	2.23	3.61	3.07	2.32

Table 7
The table reports the results of robustness tests. In Panel A, the SDFs of the models are estimated using excess returns on the Fama-French 25 size and book-to-market portfolios, 10 price-to-earnings portfolios, and 10 industry portfolios as additional priced assets. In Panel B, the macroeconomic shocks are constructed from the real-time data. In Panel C, the macroeconomic shocks are obtained from the FAVAR model. In Panel D, we include the interactions of all macroeconomic shocks with six state variables as additional regressors in the elastic net. In Panel E, the first month of each quarter is assigned to the testing sample and the other two months represent the training sample. In Panel F, every fifth month is assigned to the testing sample and all other months constitute the training sample.

Panel A: Alternative estimation of the SDFs

		non-market components m_t^N						orthogonalized non-market components $m_t^\perp$						
		FF3	FF5	Q4	BS6	M4	BF3	FF3	FF5	Q4	BS6	M4	BF3	KNS
Empirical sample	R^2_{os},%	0.00	0.00	-0.91	-0.11	1.37	8.54	0.00	0.00	0.00	0.00	0.00	-0.02	0.00
	p-value	1.00	0.99	0.76	0.97	0.31	0.10	0.85	1.00	0.98	1.00	0.76	0.53	0.92
Bootstrapped samples	Mean(R^2_{os}),%	-0.82	-1.21	-7.63	-3.70	-2.28	6.23	0.24	-1.30	-0.51	-1.00	-0.33	-0.31	0.17
	Std(R^2_{os}),%	2.60	2.09	6.83	4.28	5.62	4.95	2.44	2.14	1.50	1.81	1.18	2.56	2.28

Panel B. Real-time macroeconomic shocks

		non-market components m_t^N						orthogonalized non-market components $m_t^\perp$						
		FF3	FF5	Q4	BS6	M4	BF3	FF3	FF5	Q4	BS6	M4	BF3	KNS
Empirical sample	R^2_{os},%	0.00	-0.56	3.42	2.42	2.13	2.10	0.00	-1.02	1.53	2.04	0.17	1.91	4.47
	p-value	0.76	0.66	0.24	0.05	0.34	0.09	0.66	0.75	0.12	0.03	0.82	0.15	0.00
Bootstrapped samples	Mean(R^2_{os}),%	-0.17	-3.73	-0.36	1.64	0.05	0.88	0.06	-4.03	-1.43	-0.02	-2.47	1.23	1.05
	Std(R^2_{os}),%	1.08	5.82	4.25	2.73	4.42	5.12	0.95	5.36	3.69	3.26	4.96	2.75	6.36

Panel C. FAVAR-based macroeconomic shocks

		non-market components m_t^N						orthogonalized non-market components $m_t^\perp$						
		FF3	FF5	Q4	BS6	M4	BF3	FF3	FF5	Q4	BS6	M4	BF3	KNS
Empirical sample	R^2_{os},%	0.00	0.00	0.00	0.00	2.96	3.23	0.69	-0.61	0.00	0.00	0.00	0.41	-0.38
	p-value	1.00	0.96	0.70	0.99	0.23	0.10	0.89	1.00	0.99	1.00	1.00	0.46	0.98
Bootstrapped samples	Mean(R^2_{os}),%	-0.09	-0.53	0.04	-1.19	1.37	1.81	-0.05	-1.62	-0.40	-1.20	-0.04	0.13	-1.92
	Std(R^2_{os}),%	0.95	1.74	1.93	2.01	3.05	4.36	2.12	2.17	0.97	1.71	1.02	2.42	4.04

Table 7 (continuation)

Panel D: Alternative set of regressors

		non-market components m_t^N						orthogonalized non-market components $m_t^{\perp}$						
		FF3	FF5	Q4	BS6	M4	BF3	FF3	FF5	Q4	BS6	M4	BF3	KNS
Empirical sample	R^2_{os},%	0.03	0.00	-0.97	-0.62	1.60	5.02	0.73	0.00	-0.14	0.00	0.00	-0.34	0.00
	p-value	0.90	0.81	0.87	0.99	0.20	0.16	0.84	0.84	0.99	1.00	1.00	0.95	1.00
Bootstrapped samples	Mean(R^2_{os}),%	-1.49	-0.12	-6.42	-3.84	0.09	-0.41	-1.00	-0.75	-1.86	-0.42	-0.71	-1.65	-0.85
	Std(R^2_{os}),%	2.98	1.33	5.05	3.40	3.37	6.84	2.79	1.71	2.65	1.03	2.24	3.26	2.59

Panel E: Alternative training and testing samples

		non-market components m_t^N						orthogonalized non-market components $m_t^{\perp}$						
		FF3	FF5	Q4	BS6	M4	BF3	FF3	FF5	Q4	BS6	M4	BF3	KNS
Empirical sample	R^2_{os},%	0.00	0.00	0.00	0.00	3.05	0.37	0.55	-0.51	0.00	0.00	0.51	-1.14	0.71
	p-value	1.00	0.84	0.59	1.00	0.30	0.09	0.89	0.99	0.98	1.00	0.82	0.59	0.92
Bootstrapped samples	Mean(R^2_{os}),%	-0.22	-0.49	0.07	-1.46	0.84	-0.61	-0.11	-1.17	-0.36	-1.56	-0.33	-1.19	-0.98
	Std(R^2_{os}),%	1.01	1.79	1.66	2.37	3.53	5.07	1.92	2.15	0.99	2.04	1.91	4.10	3.48

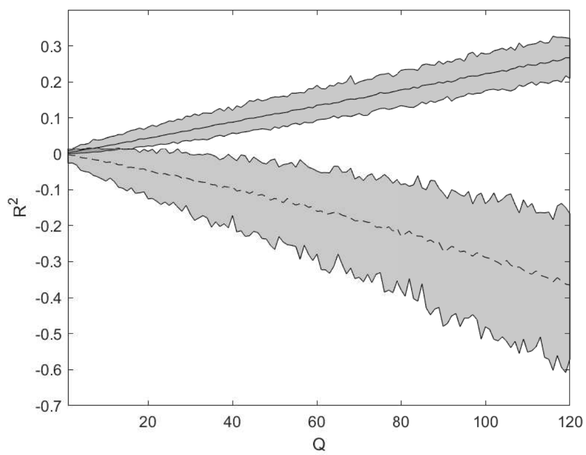
Panel F: Alternative size of the training sample

		non-market components m_t^N						orthogonalized non-market components $m_t^{\perp}$						
		FF3	FF5	Q4	BS6	M4	BF3	FF3	FF5	Q4	BS6	M4	BF3	KNS
Empirical sample	R^2_{os},%	0.00	-0.08	-0.20	0.29	2.34	-4.81	-0.69	0.00	-1.39	0.00	-0.65	-7.52	0.00
	p-value	1.00	0.84	0.50	0.94	0.21	0.18	0.75	0.85	0.92	1.00	0.71	0.63	0.98
Bootstrapped samples	Mean(R^2_{os}),%	-0.30	-0.76	-5.77	-3.27	-1.26	-12.46	-3.72	-0.11	-3.69	-1.05	-3.92	-12.65	-1.07
	Std(R^2_{os}),%	2.05	2.31	6.37	4.86	5.52	13.45	7.16	1.92	4.36	1.89	5.88	8.68	2.97

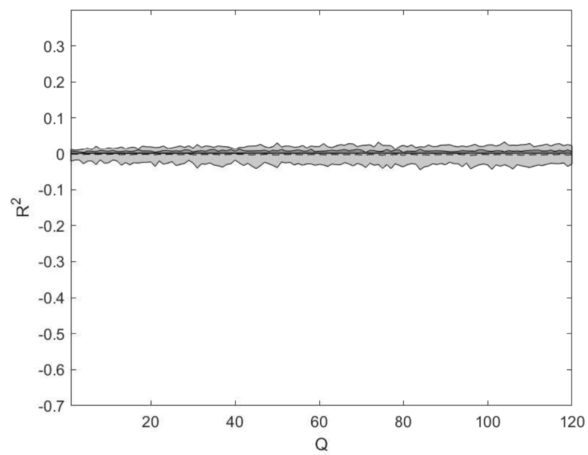
Figure 1
This figure plots the simulated average in-sample R^2, R^2_{is}, (solid line) and the simulated average out-of-sample R^2, R^2_{os}, (dashed line) estimated by the OLS and elastic net regressions for different numbers of explanatory variables Q. The figure also shows the 95% confidence bands for R^2_{is} and R^2_{os}.

Panel A: $R^2_p = 0$

OLS

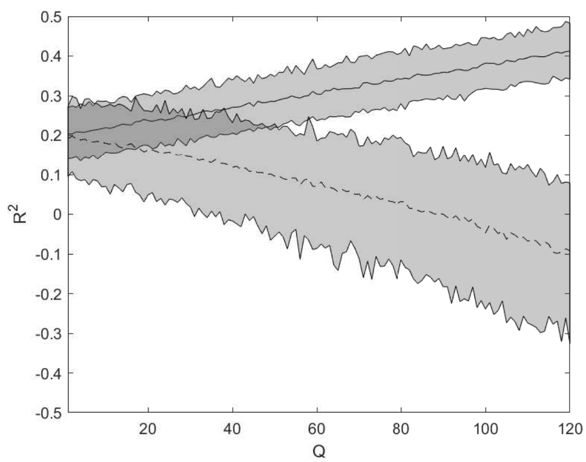


Elastic Net



Panel B: $R^2_p = 0.2$

OLS



Elastic Net

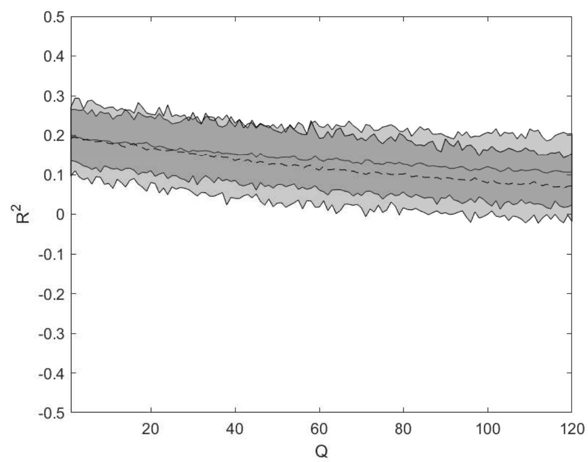
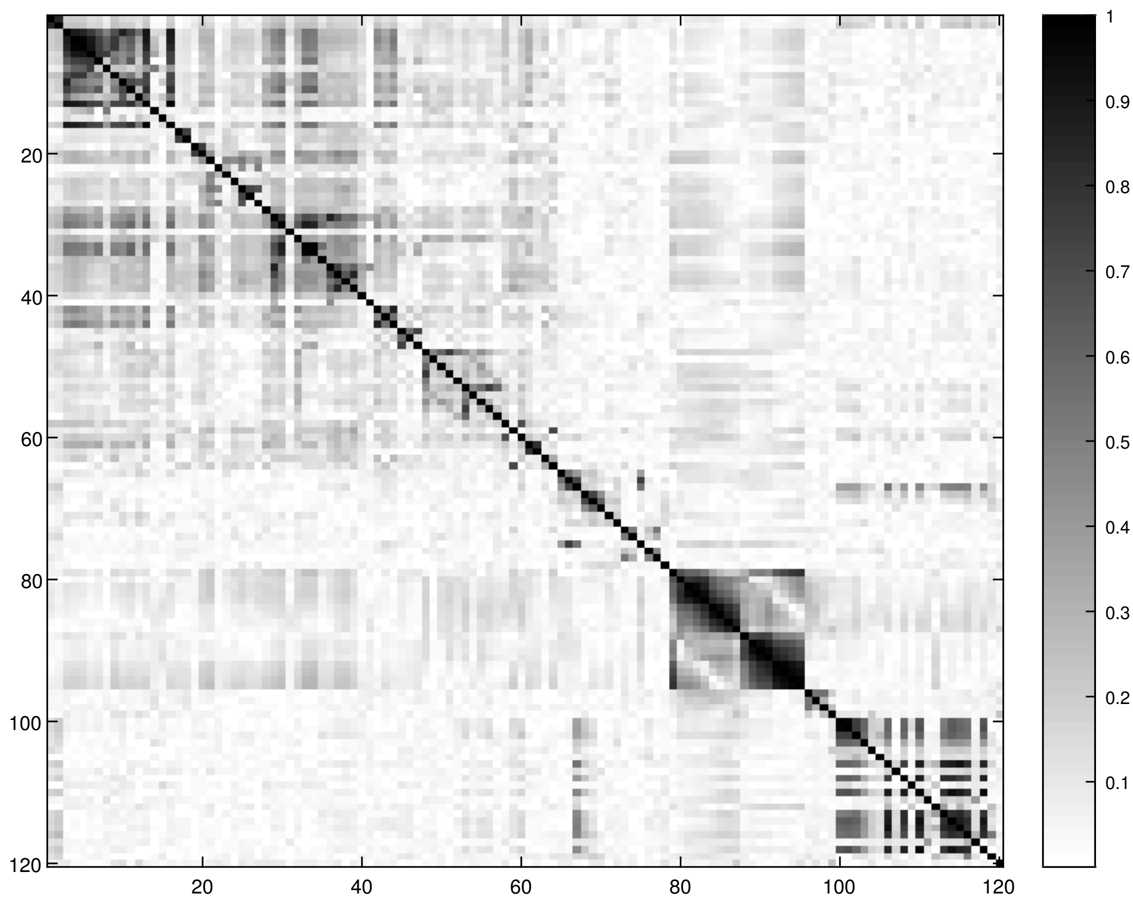


Figure 2
This figure provides a heat map of absolute correlations between macroeconomic shocks. The darker shade represents a stronger correlation. The macroeconomic shocks are ordered as in Appendix B. The sample is from January 1963 to December 2019.



Appendix A. Asset pricing factors

This appendix provides definitions of the asset pricing factors from the considered asset pricing models.

CMA is an investment factor, which is constructed using the 2-by-3 value-weighted portfolios formed on size and investment. The factor realizations are the average returns on the two conservative investment portfolios minus the average returns on the two aggressive investment portfolios.

FIN is a long-term financing factor, which is constructed using the 2-by-3 value-weighted portfolios formed on size and financing characteristics. The factor realizations are the average returns on the two high financing characteristics portfolios minus the the average returns on the two low financing characteristics portfolios.

HML is a value factor, which is constructed using the 2-by-3 value-weighted portfolios formed on size and book-to-market ratio. The factor realizations are the average returns on the two high book-to-market portfolios minus the average returns on the two low book-to-market portfolios.

HML^{m} is a modified value factor constructed similar to HML but using book-to-market ratios with the most recent monthly stock price in the denominator.

IA is an investment factor, which is constructed using the 2-by-3-by-3 value-weighted portfolios formed on size, investment-to-assets, and return-on-equity. The factor realizations are the average returns on the six low investment-to-asset portfolios minus average returns on the six high investment-to-asset portfolios.

ME is a size factor, which is constructed using the value-weighted returns on 18 portfolios that are sorted 2-by-3-by-3 independently based on size, investment-to-asset, and return-on-equity. The factor realizations are the average returns on the nine small size portfolios minus the average returns on the nine big portfolios.

MGMT is a management anomaly factor, which is constructed using the 2-by-3 value-weighted portfolios formed on size and firms' average ranking based on a cluster of six asset pricing anomalies that represent quantities that firms' managements can affect directly. The factor realizations are the average returns on the two high management anomalies portfolios minus the average returns on the two low management anomalies portfolios.

MKT is the market factor defined as the value-weighted excess return on all CRSP firms incorporated in the U.S. and listed on the NYSE, AMEX, or NASDAQ.

PEAD is a short-term post earning announcement drift factor, which is constructed using the 2-by-2 value-weighted portfolios formed on size and the sign of post earnings announcement drift. The factor realizations are the returns on the two positive post earnings announcement drift portfolios minus the average returns on the two negative post earnings announcement drift portfolios.

PERM is a performance anomaly factor, which is constructed using the 2-by-3 value-weighted portfolios formed on size and firms' average ranking based on a cluster of five asset pricing anomalies that represent quantities that are related to firm performance. The factor realizations are the average returns on the two high performance anomalies portfolios minus the average returns on the two low performance anomalies portfolios.

ROE is a profitability factor, which is constructed using value-weighted returns on 18 portfolios that are sorted 2-by-3-by-3 independently based on size, investment-to-asset, and return-on-equity. The factor realizations are the average returns on the six high return-on-equity portfolios minus the average returns on the six low return-on-equity portfolios.

RMW is a profitability factor, which is constructed using the 2-by-3 value-weighted portfolios formed on size and operating profitability. The factor realizations are the average returns on the two robust operating profitability portfolios minus the average returns on the two weak operating profitability portfolios.

SMB is a size factor, which is constructed using the 2-by-3 value-weighted portfolios formed on size and book-to-market ratio. The factor realizations are the average returns on the three small portfolios minus the average returns on the three big portfolios.

UMD is a momentum factor, which is constructed using the 2-by-3 value-weighted portfolios formed on size and prior 2-12 month stock returns. The factor realizations are the average returns on the two high return portfolios minus the average returns on the two low return portfolios.

Appendix B. Macroeconomic variables

The table below lists the 120 macroeconomic variables used in our analysis. The names and descriptions directly follow McCracken and Ng (2016). "T-Code" indicates how each variable has been transformed: (1) = no transformation, (2) = first-order difference, (3) = second-order difference, (4) = logarithm, (5) = first-order difference in the log value, (6) = second-order difference in the log value, (7) = first-order difference in the relative change.

	Variable	T-Code	Description
			Category 1: Output and income
1	RPI	5	Real Personal Income
2	W875RX1	5	Real Personal Income Ex Transfer Receipts
3	INDPRO	5	IP Index

	Variable	T-Code	Description
4	IPFPNSS	5	IP: Final Products and Nonindustrial Supplies
5	IPFINAL	5	IP: Final Products (Market Group)
6	IPCONGD	5	IP: Consumer Goods
7	IPDCONGD	5	IP: Durable Consumer Goods
8	IPNCONGD	5	IP: Nondurable Consumer Goods
9	IPBUSEQ	5	IP: Business Equipment
10	IPMAT	5	IP: Materials
11	IPDMAT	5	IP: Durable Materials
12	IPNMAT	5	IP: Nondurable Materials
13	IPMANSICS	5	IP: Manufacturing (SIC)
14	IPB51222S	5	IP: Residential Utilities
15	IPFUELS	5	IP: Fuels
16	CUMFNS	2	Capacity Utilization: Manufacturing
		Category 2: Labor market	
17	HWI	2	Help-Wanted Index for United States
18	HWIURATIO	2	Ratio of Help Wanted/No. Unemployed
19	CLF16OV	5	Civilian Labor Force
20	CE16OV	5	Civilian Employment
21	UNRATE	2	Civilian Unemployment Rate
22	UEMPMEAN	2	Average Duration of Unemployment (Weeks)
23	UEMPLT5	5	Civilians Unemployed - Less Than 5 Weeks
24	UEMP5TO14	5	Civilians Unemployed for 5-14 Weeks
25	UEMP15OV	5	Civilians Unemployed - 15 Weeks & Over
26	UEMP15T26	5	Civilians Unemployed for 15-26 Weeks
27	UEMP27OV	5	Civilians Unemployed for 27 Weeks and Over
28	CLAIMSx	5	Initial Claims
29	PAYEMS	5	All Employees: Total nonfarm
30	USGOOD	5	All Employees: Goods-Producing Industries
31	CES1021000001	5	All Employees: Mining and Logging: Mining
32	USCONS	5	All Employees: Construction
33	MANEMP	5	All Employees: Manufacturing
34	DMANEMP	5	All Employees: Durable goods
35	NDMANEMP	5	All Employees: Nondurable goods
36	SRVPRD	5	All Employees: Service-Providing Industries
37	USTPU	5	All Employees: Trade, Transportation & Utilities
38	USWTRADE	5	All Employees: Wholesale Trade
39	USTRADE	5	All Employees: Retail Trade
40	USFIRE	5	All Employees: Financial Activities
41	USGOVT	5	All Employees: Government
42	CES0600000007	1	Avg Weekly Hours : Goods-Producing
43	AWOTMAN	2	Avg Weekly Overtime Hours : Manufacturing
44	AWHMAN	1	Avg Weekly Hours : Manufacturing
45	CES0600000008	6	Avg Hourly Earnings : Goods-Producing
46	CES2000000008	6	Avg Hourly Earnings : Construction
47	CES3000000008	6	Avg Hourly Earnings : Manufacturing

	Variable	T-Code	Description
			Category 3: Housing
48	HOUST	4	Housing Starts: Total New Privately Owned
49	HOUSTNE	4	Housing Starts, Northeast
50	HOUSTMW	4	Housing Starts, Midwest
51	HOUSTS	4	Housing Starts, South
52	HOUSTW	4	Housing Starts, West
53	PERMIT	4	New Private Housing Permits (SAAR)
54	PERMITNE	4	New Private Housing Permits, Northeast (SAAR)
55	PERMITMW	4	New Private Housing Permits, Midwest (SAAR)
56	PERMITS	4	New Private Housing Permits, South (SAAR)
57	PERMITW	4	New Private Housing Permits, West (SAAR)
			Category 4: Consumption, orders, and inventories
58	DPCERA3M086SBEA	5	Real personal consumption expenditures
59	CMRMTSPLx	5	Real Manu. and Trade Industries Sales
60	RETAILx	5	Retail and Food Services Sales
61	AMDMNOx	5	New Orders for Durable Goods
62	AMDMUOx	5	Unfilled Orders for Durable Goods
63	BUSINVx	5	Total Business Inventories
64	ISRATIOx	2	Total Business: Inventories to Sales Ratio
			Category 5: Money and credit
65	M1SL	6	M1 Money Stock
66	M2SL	6	M2 Money Stock
67	M2REAL	5	Real M2 Money Stock
68	AMBSL	6	St. Louis Adjusted Monetary Base
69	TOTRESNS	6	Total Reserves of Depository Institutions
70	NONBORRES	7	Reserves Of Depository Institutions
71	BUSLOANS	6	Commercial and Industrial Loans
72	REALLN	6	Real Estate Loans at All Commercial Banks
73	NONREVSL	6	Total Nonrevolving Credit
74	CONSPI	2	Nonrevolving consumer credit to Personal Income
75	MZMSL	6	MZM Money Stock
76	DTCOLNVHFNM	6	Consumer Motor Vehicle Loans Outstanding
77	DTCTHFNM	6	Total Consumer Loans and Leases Outstanding
78	INVEST	6	Securities in Bank Credit at All Commercial Banks N.A.
			Category 6: Interest rates and exchange rates
79	FEDFUNDS	2	Effective Federal Funds Rate
80	CP3Mx	2	3-Month AA Financial Commercial Paper Rate
81	TB3MS	2	3-Month Treasury Bill
82	TB6MS	2	6-Month Treasury Bill
83	GS1	2	1-Year Treasury Rate
84	GS5	2	5-Year Treasury Rate
85	GS10	2	10-Year Treasury Rate
86	AAA	2	Moody's Seasoned Aaa Corporate Bond Yield

	Variable	T-Code	Description
87	BAA	2	Moody's Seasoned Baa Corporate Bond Yield
88	COMPAPFFx	1	3-Month Commercial Paper Minus FEDFUNDS
89	TB3SMFFM	1	3-Month Treasury C Minus FEDFUNDS
90	TB6SMFFM	1	6-Month Treasury C Minus FEDFUNDS
91	T1YFFM	1	1-Year Treasury C Minus FEDFUNDS
92	T5YFFM	1	5-Year Treasury C Minus FEDFUNDS
93	T10YFFM	1	10-Year Treasury C Minus FEDFUNDS
94	AAAFFM	1	Moody's Aaa Corporate Bond Minus FEDFUNDS
95	BAAFFM	1	Moody's Baa Corporate Bond Minus FEDFUNDS
96	EXSZUSx	5	Switzerland / U.S. Foreign Exchange Rate
97	EXJPUSx	5	Japan / U.S. Foreign Exchange Rate
98	EXUSUKx	5	U.S. / U.K. Foreign Exchange Rate
99	EXCAUSx	5	Canada / U.S. Foreign Exchange Rate
		Category 7: Prices	
100	WPSFD49207	6	PPI: Finished Goods
101	WPSFD49502	6	PPI: Finished Consumer Goods
102	WPSID61	6	PPI: Intermediate Materials
103	WPSID62	6	PPI: Crude Materials
104	OILPRICEx	6	Crude Oil, spliced WTI and Cushing
105	PPICMM	6	PPI: Metals and metal products
106	CPIAUCSL	6	CPI : All Items
107	CPIAPPSL	6	CPI : Apparel
108	CPITRNSL	6	CPI : Transportation
109	CPIMEDSL	6	CPI : Medical Care
110	CUSR0000SAC	6	CPI : Commodities
111	CUUR0000SAD	6	CPI : Durables
112	CUSR0000SAS	6	CPI : Services
113	CPIULFSL	6	CPI : All Items Less Food
114	CUUR0000SA0L2	6	CPI : All items less shelter
115	CUSR0000SA0L5	6	CPI : All items less medical care
116	PCEPI	6	Personal Cons. Expend.: Chain Index
117	DDURRG3M086SBEA	6	Personal Cons. Exp: Durable goods
118	DNDGRG3M086SBEA	6	Personal Cons. Exp: Nondurable goods
119	DSERRG3M086SBEA	6	Personal Cons. Exp: Services
		Category 8: Stock market	
120	VXOCLSx	1	VXO