

NBER WORKING PAPER SERIES

TAMING THE FACTOR ZOO:
A TEST OF NEW FACTORS

Guanhao Feng
Stefano Giglio
Dacheng Xiu

Working Paper 25481
<http://www.nber.org/papers/w25481>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2019

We appreciate insightful comments from Alex Belloni, John Campbell, John Cochrane, Chris Hansen, Lars Hansen, Bryan Kelly, Stefan Nagel and Chen Xue. We are also grateful for helpful comments from seminar and conference participants at the City University of Hong Kong, Peking University, Renmin University, University of British Columbia, Luxembourg School of Finance, AQR, Morgan Stanley, Two Sigma, 2018 Annual Meetings of the American Finance Association, the 2016 Financial Engineering and Risk Management Symposium in Guangzhou, 2017 EcoStat Conference at Hong Kong University of Science and Technology, and University of Oregon Summer Finance Conference. We acknowledge research support by the Fama-Miller Center for Research in Finance at Chicago Booth. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Guanhao Feng, Stefano Giglio, and Dacheng Xiu. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Taming the Factor Zoo: A Test of New Factors
Guanhao Feng, Stefano Giglio, and Dacheng Xiu
NBER Working Paper No. 25481
January 2019
JEL No. C01,C12,C23,C52,C55,C58,G00,G1,G10,G12

ABSTRACT

We propose a model-selection method to systematically evaluate the contribution to asset pricing of any new factor, above and beyond what a high-dimensional set of existing factors explains. Our methodology explicitly accounts for potential model-selection mistakes, unlike the standard approaches that assume perfect variable selection, which rarely occurs in practice and produces a bias due to the omitted variables. We apply our procedure to a set of factors recently discovered in the literature. While most of these new factors are found to be redundant relative to the existing factors, a few — such as profitability— have statistically significant explanatory power beyond the hundreds of factors proposed in the past. In addition, we show that our estimates and their significance are stable, whereas the model selected by simple LASSO is not.

Guanhao Feng
City University of Hong Kong
83 Tat Chee Avenue
Kowloon Tong
Hong Kong
gavin.feng@cityu.edu.hk

Dacheng Xiu
Booth School of Business
University of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637
dachxiu@chicagobooth.edu

Stefano Giglio
Yale School of Management
165 Whitney Avenue
New Haven, CT 06520
and NBER
stefano.giglio@yale.edu

An online appendix is available at <http://www.nber.org/data-appendix/w25481>

1 Introduction

The search for factors that explain the cross section of expected stock returns has produced hundreds of potential candidates, as noted by [Cochrane \(2011\)](#) and most recently by [Harvey et al. \(2015\)](#), [McLean and Pontiff \(2016\)](#), and [Hou et al. \(2017\)](#). A fundamental task facing the asset pricing field today is to bring more discipline to the proliferation of factors – a task that the literature has been trying to address using a variety of methods. In this paper, we approach this problem from the following angle: how can we judge whether a new factor adds explanatory power for asset pricing, relative to the existing set of hundreds of factors the literature has so far produced?

This paper provides a framework for systematically evaluating the contribution of individual factors relative to the myriad of existing factors, and conducting appropriate statistical inference in this high-dimensional setting. In particular, we show how to estimate and test the marginal importance of any factor g_t in pricing the cross section of expected returns *beyond* what is explained by a high-dimensional set of potential factors h_t , where g_t and h_t could be tradable or non-tradable factors. We assume the true asset pricing model is approximately low-dimensional; however, in addition to relevant asset pricing factors, g_t and h_t include redundant ones that add no explanatory power to the model, as well as useless ones that have no explanatory power at all. Selecting the relevant factors from h_t and conducting proper inference on the contribution of g_t above and beyond those factors is the aim of this paper. Our methodology can be thought of as a conservative test for new factors, which benchmarks them against a large-dimensional set of existing ones.

When h_t consists of a small number of factors, testing whether g_t is useful in explaining asset prices while controlling for the factors in h_t is straightforward: it simply requires estimating the loadings of the stochastic discount factor (SDF) on g_t and h_t , and testing whether the loading of g_t is different from zero (see [Cochrane \(2009\)](#)). This exercise not only tells us whether g_t is useful for pricing the cross section, but it also reveals how shocks to g_t affect marginal utility, which has a direct economic interpretation.

When h_t consists of potentially hundreds of factors, however, standard statistical methods to estimate and test the SDF loadings become infeasible or result in poor estimates and invalid inference, because of the curse of dimensionality. Although variable selection techniques (e.g., least absolute shrinkage and selection operator, LASSO) can be useful in selecting the correct variables under certain conditions and thereby reducing the dimensionality of h_t , relying on this result produces very poor approximations to the finite-sample distributions of the estimators, unless appropriate econometric methods are used to explicitly account for model-selection mistakes (see [Chernozhukov et al. \(2015\)](#)). This means that, for example, simply applying a model-selection tool like LASSO to a large set of factors and checking whether a particular factor g_t is significant (or even just checking

if it gets selected) is not a reliable way to determine whether g_t is actually one of the true factors.

The methodology we propose in this paper marries these new econometric methods (in particular, the *double-selection LASSO* method of Belloni et al. (2014b)) with two-pass regressions such as Fama-MacBeth to evaluate the contribution of a factor to explaining asset prices specifically in a high-dimensional setting. Without relying on prior knowledge about which factors to include as controls among a large number of factors in h_t , our procedure selects factors that are *either* useful in explaining the cross section of expected returns *or* are useful in mitigating the omitted variable bias problem due to potential model selection mistakes. We show that including both types of factors as controls is essential to conduct reliable inference on the SDF loading of g_t .

We apply our methodology to a large set of factors proposed in the last 30 years; in particular, we collect and construct a large factor data library containing 150 risk factors. This factor zoo contains many potentially redundant factors, and is thus an ideal dataset to show our empirical results. As an example, consider the seasonality factor of Heston and Sadka (2008). This factor has a statistically significant alpha with respect to the Fama-French 3-factor model (t-stat 2.06) in our sample. So, if evaluated against this benchmark model, one would conclude that seasonality is a useful factor. But seasonality turns out to be highly correlated with momentum (for example, it has a correlation of 0.63 with Carhart momentum). And if one evaluates it against a model that includes momentum (like the Fama-French 4 factor model), the alpha becomes small and statistically insignificant (t-stat of -0.87). This example highlights the importance of the benchmark in evaluating new factors. Most papers in the literature that aim to produce new factors, nonetheless, choose the benchmark model somewhat arbitrarily, subject to a potential data-mining bias. Our procedure systematically constructs the best low-dimensional benchmark to evaluate new factors using the entire factor zoo.

We perform a variety of empirical exercises that illustrate the use of our procedure in the data. We start by evaluating the marginal contribution of recent factors proposed in the last five years (2012 - 2016) to the large set of factors proposed before then. The new factors include – among others – the two new factors introduced by Fama and French (2015) and Hou et al. (2015), and the intermediary-based factors from He et al. (2016). Note that our test is conservative: it requires a new factor g_t to contribute to explaining the cross-section relative to the *entire universe* of existing factors h_t . Given the large dimensionality of the factors produced in the literature, one might wonder whether, in practice, any additional factor could ever make a significant contribution. We show that indeed several of the newly proposed factors (e.g., profitability and investment) have significant marginal explanatory power for expected returns.

Second, we propose a recursive exercise in which factors are tested as they are introduced against previously proposed factors. The exercise shows that our procedure would have deemed factors as redundant or spurious in most cases, while finding significance for a small number of factors. Over

time, our procedure would have screened out many factors at the time of their introduction, thus helping address the proliferation of factors. Going forward, our test can be used to make inference about new factors that will be introduced in the future.

Third, we study the robustness of our procedure from different angles. We show that our results are robust to using alternative methods to reduce the dimensionality of h_t , like Elastic Net and PCA. We also show that the results are robust to alternative portfolio constructions. Most importantly, we explore in detail the robustness with respect to the tuning parameters. Like all machine learning methods, our procedure involves the choice of tuning parameters (in particular two, one for each selection step). In our analysis, we choose them by cross-validation; in the robustness section, we also show that our empirical findings are robust to varying the tuning parameters in the neighborhood of the values chosen by the cross-validation procedure.

The double-selection (DS) estimation procedure we propose, that combines cross-sectional asset pricing regressions with the double-selection LASSO of [Belloni et al. \(2014b\)](#) (designed originally for linear treatment effect models), starts by using a two-step selection method to select “control” factors from h_t , and then estimates the SDF loading of g_t from cross-sectional regressions that include g_t and the selected factors from h_t .

As the name implies, the “double selection” of factors from h_t happens in two stages; both stages are crucial to obtain correct inference on g_t . A first set of factors is selected from h_t based on their pricing ability for the cross-section of returns. Factors whose covariances appear to contribute little to pricing assets in the cross section are excluded from the set of controls. This first step – effectively an application of standard LASSO to the set of potential factors h_t – has the advantage of selecting factors based on their usefulness in pricing the cross section of assets, as opposed to other commonly used selection methods (e.g., principal components) that select factors based on their ability to explain the time-series variation of returns. Using a cross-sectional approach with factor covariances as inputs is expected to deliver more relevant factors for asset pricing.

This first step therefore chooses a low-dimensional model to explain the cross section using only factors in h_t . This model selection step corresponds closely to the approach taken in the current literature dealing with the proliferation of asset pricing factors (e.g., [Kozak et al. \(2017\)](#)): take a large set of factors (h_t), apply some dimension-reduction method (LASSO, Elastic net, PCA, etc.), and interpret the resulting low-dimensional model as the SDF. Importantly, the interpretation of the selected model in the literature has relied on the so-called “oracle property” of LASSO and other model-selection methods: an asymptotic property that guarantees that under certain assumptions, as the sample size goes to infinity, the procedures will eventually recover the true model. The first step in our procedure, therefore, is similar in spirit to what has been commonly applied in the recent literature.

In this paper, however, we make one step forward, and recognize that in practice the “oracle property” never holds. For instance, LASSO makes frequent and potentially important mistakes when recovering the SDF, as we show in simulations. To make things worse, the failure of the “oracle property” in finite samples is also a problem for addressing the question we focus on in this paper: whether a new factor g_t improves over the factors in h_t . Mistakes in selecting the reduced-dimension model from h_t also make inference on g_t invalid. The LASSO selection may exclude some factors that have small SDF loadings in sample, but whose covariance with returns are nonetheless highly cross-sectionally correlated with exposures to g_t . Any omission of relevant factors due to model-selection errors distorts the asymptotic distribution of the estimator, leading to incorrect inference on the significance – and even the sign – of g_t ’s SDF loading. This issue is well-known in the statistics literature (see, for example, [Leeb and Pötscher \(2005\)](#)), and it has spurred a large econometrics literature on uniformly valid inference, with important consequences for asset pricing tests that we explore in this paper.

The key contribution of our paper is to show that despite the mistakes that LASSO inevitably makes in selecting the model, correct inference *can* be made about the contribution to asset pricing of a factor g_t . To obtain reliable asymptotic inference for g_t , including a second stage of factor selection is crucial. The second step adds to the set of controls selected by the first-stage LASSO additional factors whose covariances with returns are highly correlated in the cross section with the covariance between returns and g_t (this step uses a second LASSO, since it still has to choose among many factors in h_t). Intuitively, we want to make sure to include even factors with small in-sample SDF loadings, if omitting them may still induce a large omitted variable bias due to the cross-sectional correlation between their risk exposures and the risk exposures to g_t . It is also possible that some variables selected from the second stage are redundant or even useless, but their inclusion only leads to a moderate loss in efficiency.

After selecting the set of controls from h_t (including all factors selected in either of the two selection stages), we conduct inference on g_t by estimating the coefficient of a standard two-pass regression using g_t and the selected control factors from h_t . This post-selection estimation step is also useful to remove biases arising from regularization in any LASSO procedure; see, for example, [Friedman et al. \(2009\)](#). We then conduct asymptotic inference on the SDF loading of g_t using a central-limit result we derive in this paper. We show in simulation that our estimator performs well in finite samples, and substantially outperforms alternative estimators.

Finally, it is worth pointing out an alternative motivation for the methodology proposed in this paper. Theoretical asset pricing models often predict that some factors (g_t) should be part of the SDF, i.e. they should enter the investors’ marginal utility. Theoretical models, however, are often very stylized, and their ability to explain the cross-section is limited. This suggests that, in reality,

investors may care about other risk factors that are not explicitly predicted by the model. This creates an omitted variable problem when testing for the SDF loading of g_t : if the true SDF contains additional factors not explicitly incorporated in the estimation, the estimate for the loading of g_t will be biased. Our methodology – that estimates the loading of g_t while taking a stand on the “omitted factors” by choosing them from the large set h_t – can then be seen as a way to address this omitted factor concern when estimating SDF loadings. In this sense, it is related to [Giglio and Xiu \(2016\)](#), that show how to make inference on risk premia in the presence of omitted factors. The crucial difference between the two approaches is that [Giglio and Xiu \(2016\)](#) focus on the estimation of risk *premia* (the compensation investors require for holding the g_t risk), whereas this paper makes inference on SDF *loadings* of observable factors in g_t . Both SDF loadings and risk premia have important, though very distinct, economic interpretations; they have different theoretical properties, and different tools need to be used to address the omitted factor problem in the two cases. Importantly, only SDF loadings, addressed in this paper, can speak to the contribution of factors to explaining asset prices (see [Cochrane \(2009\)](#)), and therefore SDF loadings are the appropriate concept to refer to for disciplining the zoo of factors.

Our paper builds on several strands of the asset pricing and econometrics literature. In addition to a large literature devoted to identifying asset pricing factors¹ and a vast econometrics literature on estimating factor models,² our paper is most closely related to the recent literature on the high dimensionality of cross-sectional asset pricing models. [Green et al. \(2016\)](#) test 94 firm characteristics through Fama-Macbeth regressions and find that 8-12 characteristics are significant independent determinants of average returns. [McLean and Pontiff \(2016\)](#) use an out-of-sample approach to study the post-publication bias of 97 discovered risk anomalies. [Harvey et al. \(2015\)](#) adopt a multiple testing framework to re-evaluate past research and suggest a new benchmark for current and future factor fishing. Following on this multiple-testing issue, [Harvey and Liu \(2016\)](#) provide a bootstrap technique to model selection. Recently, [Freyberger et al. \(2017\)](#) propose a group LASSO procedure to select characteristics and to estimate how they affect expected returns nonparametrically. [Kozak et al. \(2017\)](#) use model-selection techniques to approximate the SDF and the mean-variance efficient portfolio as a function of many test portfolios, and compare sparse models based on principal

¹Some of the factors proposed in the literature are based on economic theory (e.g., [Breedon \(1979\)](#), [Chen et al. \(1986\)](#), [Jagannathan and Wang \(1996\)](#), [Lettau and Ludvigson \(2001\)](#), [Yogo \(2006\)](#), [Pástor and Stambaugh \(2003a\)](#), [Adrian et al. \(2014\)](#), [He et al. \(2016\)](#)); others are constructed using firm characteristics, such as [Fama and French \(1993, 2015\)](#), [Carhart \(1997\)](#), and [Hou et al. \(2015\)](#).

²See, among the many papers, [Jensen et al. \(1972\)](#), [Fama and MacBeth \(1973\)](#), [Ferson and Harvey \(1991\)](#), [Shanken \(1992\)](#), [Jagannathan and Wang \(1996\)](#), [Welch \(2008\)](#), and [Lewellen et al. \(2010\)](#). These papers, along with the majority of the literature, rely on large T and fixed n asymptotic analysis for statistical inference and only deal with models in which all factors are specified and observable. Recent literature relies on alternative asymptotic designs, including [Bai and Zhou \(2015\)](#), [Gagliardini et al. \(2016\)](#), [Gagliardini et al. \(2017\)](#), [Connor et al. \(2012\)](#), [Giglio and Xiu \(2016\)](#), and [Raponi et al. \(2017\)](#), for better small-sample performance and robustness to model misspecification.

components of returns with sparse models based on characteristics.

A crucial distinction between our paper and the existing literature is that we focus on the evaluation of a *new factor*, rather than testing or estimating an entire reduced-form asset pricing model, e.g., in the GRS test of [Gibbons et al. \(1989\)](#). To the extent that our procedure is used to test a new factor g_t that is determined ex-ante and motivated by theory, it is not directly subject to the multiple testing concern that [Harvey and Liu \(2016\)](#) aim to address.³ Our procedure also helps alleviate the concern of data-snooping, another form of multiple testing (see e.g., [Lo and MacKinlay \(1990\)](#), [Harvey et al. \(2015\)](#)), because we suggest imposing discipline to the selection of controls as opposed to the conventional practice of selecting arbitrary controls that leaves the researcher much more freedom.

Of course, the existing literature has routinely attempted to evaluate the contribution of new factors relative to some benchmark model, typically by estimating and testing the alpha of a regression of the new factor onto existing factors (e.g., [Barillas and Shanken \(2018\)](#) and [Fama and French \(2016\)](#)). Our methodology differs from the existing procedures in several ways. First, we do not use as control an arbitrary set of factors from h_t (e.g., the three Fama-French factors), but rather we select from h_t the control model that best explains the cross section of returns. In addition, our procedure aims to minimize the potential omitted variable bias while enhancing statistical efficiency. Second, we not only test whether the factor of interest g_t is useful in explaining asset prices, but we also estimate its role in driving marginal utility (its coefficient in the stochastic discount factor); this is important to be able to interpret the results in economic terms and relate them to the models that motivated the choice of g_t . Third, our procedure handles both traded and non-traded factors. Fourth, our procedure leverages information from the cross section of the test assets in addition to the times-series of the factors. Lastly, our inference is valid given a large dimensional set of controls and test assets in addition to an increasing span of time series.

Finally, our paper is related to a large statistical and machine-learning literature on variable selection and regularization using LASSO and post-selection inference. For theoretical properties of LASSO, see [Bickel et al. \(2009\)](#), [Meinshausen and Yu \(2009\)](#), [Tibshirani \(2011\)](#), [Wainwright \(2009\)](#), [Zhang and Huang \(2008\)](#), [Belloni and Chernozhukov \(2013\)](#). For the post-selection-inference method, see, for example, [Belloni et al. \(2012\)](#), [Belloni et al. \(2014b\)](#), and review articles by [Belloni et al. \(2014a\)](#) and [Chernozhukov et al. \(2015\)](#). Our asymptotic results are new to the existing literature in two important respects. First, our setting is a large panel regression with a large number of factors, in which both cross-sectional and time-series dimensions increase. Second, our procedure in fact

³The two methodologies could potentially be combined to produce more conservative inference that also deals with the possibility that the set of test factors g_t is selected ex-post after looking at the inference results, raising concerns about multiple testing. We leave this for future research. Relatedly, [Giglio et al. \(2018\)](#) tackle the multiple testing of alphas in a linear asset pricing model.

selects covariances between factors and returns, which are contaminated by estimation errors, rather than factors themselves that are immediately observable.

The rest of the paper is organized as follows. In Section 2, we set up the model, present our methodology, and develop relevant statistical inference. In Section 3, we show several empirical applications of the procedure, and explore the robustness of the results. Section 4 concludes. The appendix contains technical details and Monte Carlo simulations.

2 Methodology

2.1 Model Setup

We start from a linear specification for the SDF:

$$m_t := \gamma_0^{-1} - \gamma_0^{-1} \lambda_v^\top v_t := \gamma_0^{-1} (1 - \lambda_g^\top g_t - \lambda_h^\top h_t), \quad (1)$$

where γ_0 is the zero-beta rate, g_t is a $d \times 1$ vector of factors to be tested, and h_t is a $p \times 1$ vector of potentially confounding factors. Without loss of generality, both g_t and h_t are de-measured; that is, they are factor innovations satisfying $E(g_t) = 0$ and $E(h_t) = 0$. λ_g and λ_h are $d \times 1$ and $p \times 1$ vectors of parameters, respectively. We refer to λ_g and λ_h as the *SDF loadings* of the factors g_t and h_t .

Our goal in this paper is to make inference on the SDF loadings of a small set of factors g_t while accounting for the explanatory power of a large number of existing factors, collected in h_t . That is, the main question in this paper is to evaluate the *marginal contribution* of g_t relative to a high-dimensional benchmark model h_t .

Note that the factors in h_t are not necessarily all useful factors: their corresponding SDF loadings may be equal to zero. This framework therefore potentially includes redundant factors (factors that have zero SDF loadings but whose covariances with returns are correlated in the cross section with the covariance between returns and the SDF), as well as completely useless factors (factors that have zero SDF loadings and whose covariances with returns are uncorrelated with the covariances of returns with the SDF). So part of the procedure we propose will reduce the dimensionality of h_t , trying to eliminate the useless and redundant factors, obtaining a low-dimensional benchmark model.

In addition to g_t and h_t , we observe a $n \times 1$ vector of test asset returns, r_t . Because of (1), expected returns satisfy:

$$E(r_t) = \iota_n \gamma_0 + C_v \lambda_v = \iota_n \gamma_0 + C_g \lambda_g + C_h \lambda_h, \quad (2)$$

where ι_n is a $n \times 1$ vector of 1s, $C_a = \text{Cov}(r_t, a_t)$, for $a = g, h$, or v . Furthermore, we assume the

dynamics of r_t follow a standard linear factor model:

$$r_t = \mathbb{E}(r_t) + \beta_g g_t + \beta_h h_t + u_t, \quad (3)$$

where β_g and β_h are $n \times d$ and $n \times p$ factor-loading matrices, u_t is a $n \times 1$ vector of idiosyncratic components with $\mathbb{E}(u_t) = 0$ and $\text{Cov}(u_t, v_t) = 0$.

Equation (2) represents expected returns in terms of (univariate) covariances with the factors, multiplied by λ_g and λ_h . An equivalent representation of expected returns can be obtained in terms of multivariate betas:

$$\mathbb{E}(r_t) = \iota_n \gamma_0 + \beta_g \gamma_g + \beta_h \gamma_h, \quad (4)$$

where β_g and β_h are the factor exposures (i.e., multivariate betas) and γ_g and γ_h are the *risk premia* of the factors. SDF loadings λ and risk premia γ are directly related through the covariance matrix of the factors, but they differ substantially in their interpretation. The risk premium of a factor tells us whether investors are willing to pay to hedge a certain risk factor, but it does not tell us whether that factor is useful in pricing the cross section of returns. For example, a factor could command a nonzero risk premium without even appearing in the SDF, simply because it is correlated with the true factors. As discussed extensively in [Cochrane \(2009\)](#), to understand whether a factor is useful in pricing the cross section of assets, we should look at its SDF loading instead of its risk premium.

Our model assumes constant risk exposure and risk premia. In the empirical analysis, we thereby recommend using characteristic-sorted portfolios instead of individual stocks. The main advantage of using portfolios is that their risk exposures are more stable over time, as discussed at length in the asset pricing literature. [Gagliardini et al. \(2016\)](#) and [Kelly et al. \(2017\)](#) allow for stock specific and time-varying betas as well as time-varying risk premia, by modeling these quantities as functions of characteristics or macro time series. Our framework can be extended to a similar setting, see a detailed discussion in [Giglio and Xiu \(2016\)](#). In particular, the estimated SDF loadings can be interpreted as estimates of their time-series averages, if the SDF loadings are time-varying.

Because the link between SDF loadings and risk premia depends on the covariances among factors, it is useful to write explicitly the projection of g_t on h_t as

$$g_t = \eta h_t + z_t, \quad \text{where} \quad \text{Cov}(z_t, h_t) = 0. \quad (5)$$

Finally, for the estimation of λ_g , it is essential to characterize the cross-sectional dependence between C_g and C_h , so we write the cross-sectional projection of C_g onto C_h as:

$$C_g = \iota_n \xi^\top + C_h \chi^\top + C_e, \quad (6)$$

where ξ is a $d \times 1$ vector, χ is a $d \times p$ matrix, and C_e is a $n \times d$ matrix of cross-sectional regression residuals.

2.2 Challenges with Standard Two-Pass Methods in High-Dimensional Settings

Using two-pass regressions to estimate empirical asset pricing models dates back to [Jensen et al. \(1972\)](#) and [Fama and MacBeth \(1973\)](#). Partly because of its simplicity, this approach is widely used in practice. The procedure involves two steps, including one asset-by-asset time-series regression to estimate individual factor loadings β s, and one cross-sectional regression of expected returns on the estimated factor loadings to estimate risk *premia* γ . Because our parameter of interest is λ_g , the first step needs to be modified to use covariances between returns and factors rather than multivariate betas. In a low-dimensional setting, this method would work smoothly, as pointed out by [Cochrane \(2009\)](#).

However, the empirical asset pricing literature has created hundreds of factors, which can include useless and redundant factors in addition to useful factors; all the useful ones should be used as controls in estimating λ_g and testing for its significance. Over time, the number of potential factors p discovered in the literature has increased to the same scale as, if not greater than, n or T . In such a scenario, the standard cross-sectional regression with all factor covariances included is at best highly inefficient. Moreover, when p is larger than n , the standard Fama-MacBeth approach becomes infeasible because the number of parameters exceeds the sample size.

Standard methodologies therefore do not work well if at all in a high-dimensional setting due to the curse of dimensionality, so that dimension-reduction and regularization techniques are inevitable for valid inference. The existing literature has so far employed ad hoc solutions to this dimensionality problem. In particular, in testing for the contribution of a new factor, it is common to cherry-pick a handful of control factors, such as the prominent Fama-French three factors, effectively imposing an assumption that the selected model is the true one and is not missing any additional factors. However, this assumption is clearly unrealistic. These standard models have generally poor performance in explaining a large available cross section of expected returns beyond 25 size- and value-sorted portfolios, indicating omitted factors are likely to be present in the data. The stake of selecting an incorrect model is high, because it leads to an omitted variable bias when useful factors are not included, or an efficiency loss when many useless or redundant factors are included.

2.3 Sparsity

This high-dimensionality issue is not unique to asset pricing. To address it, we need to impose a certain low-dimensional structure on the model. In this paper, like in much of the recent asset pricing literature, we impose a sparsity assumption that has a natural economic interpretation and has recently been studied at length in the machine-learning literature. Imposing sparsity in our setting means that a relatively small number of factors exist in h_t , whose linear combinations along

with g_t nest the SDF m_t , and those alone are relevant for the estimation of λ_g . More specifically, sparsity in our setting means there are only s non-zero entries in λ_h , and in each row of η and χ , where s is small relative to n and T . The sparsity assumption allows us to extract the most influential factors, while making valid inference on the parameters of interest, without *prior knowledge* or even *perfect recovery* of the useful factors that determine m_t .

Does sparsity make sense in asset pricing? In fact, the asset pricing literature has adopted the concept of sparsity without always explicitly acknowledging it. In addition to the proposed factor or the factor of interest, almost all empirical asset pricing models include only a handful of control factors, such as the Fama-French three or five factors, the momentum factor, etc. Such models provide a parsimonious representation of the cross section of expected returns, hence they typically outperform models with many factors in out-of-sample settings. This is a form of sparsity where the few factors allowed to have non-zero SDF loadings are chosen *ex ante*. Moreover, sparse models are easier to interpret and to link to economic theories, compared to alternative latent factor models, which often use the principal components as factors. Last but not least, as advocated in [Friedman et al. \(2009\)](#), one should “bet on sparsity” since no procedure does well in dense problems. The notion of sparse versus dense is relative to the sample size, the number of covariates, the signal to noise ratio, etc. Sparsity does not necessarily mean that the true model should always only involve a very small number of factors in absolute terms, say 3 or 5. More non-zero coefficients can be identified given better conditions (e.g., larger sample size).

2.4 LASSO and Model Selection Mistakes

To leverage sparsity, [Tibshirani \(1996\)](#) proposes the so-called LASSO estimator, which incorporates into the least-squares optimization a penalty function on the \mathbb{L}_1 norm of parameters, which leads to an estimator that has many zero coefficients in the parameter vector. The LASSO estimator has appealing properties in particular for prediction purposes. With respect to parameter estimation, however, a well-documented bias is associated with the non-zero coefficients of the LASSO estimate because of the regularization. For these reasons, [Belloni and Chernozhukov \(2013\)](#) and [Belloni et al. \(2012\)](#) suggest the use of a “Post-LASSO” estimator, which they have shown more desirable statistical properties. The Post-LASSO estimator runs LASSO as a model selector, and then re-fits the least-squares problem without penalty, using only variables that have non-zero coefficients in the first step.

In the asset pricing context, the LASSO and Post-LASSO procedures could theoretically be used to select the factors in h_t with non-zero SDF loadings as controls for g_t , therefore accounting for the possibility that h_t contains useless or redundant factors.

Unfortunately, these procedures are not appropriate when we conduct *inference*, because funda-

mentally, LASSO and other machine learning methods aim for better *prediction*. LASSO is designed to minimize the out-of-sample prediction error. Certain variables, even if they are part of the true model, may be not worth of including for prediction purpose, because their contribution to prediction is too small relative to the cost of inclusion. In fact, in any finite sample, we can never be sure LASSO or Post-LASSO will select the correct model, just like we cannot claim the estimated parameter values in a given finite sample are equal to their population counterparts. But if the model is misspecified, that is, if important factors are mistakenly excluded from the control, inference about the SDF loadings will be affected by an omitted variable bias. Therefore, standard LASSO or Post-LASSO regressions will generally yield erroneous inference, as we confirm in simulations in [Appendix A](#).

This omitted variable bias due to model-selection mistakes is exacerbated if risk exposures to the omitted factors are highly correlated in the cross section with the exposures to g_t , even though these factors may have small SDF loadings (which is why they are likely omitted by LASSO). We will therefore need to ensure that these factors are included in the set of controls *even if LASSO would suggest excluding them*. Note this issue is not unique to high-dimensional problems – see, for example, [Leeb and Pötscher \(2005\)](#) – but it is arguably more severe in such a scenario because model selection is inevitable.

2.5 Two-Pass Regression with Double-Selection LASSO

To guard against omitted variable biases due to selection mistakes, we therefore adopt a double-selection strategy in the same spirit as what [Belloni et al. \(2014b\)](#) propose for estimating the treatment effect. The first selection (basically, standard LASSO) searches for factors in h_t whose covariances with returns are useful for explaining the cross section of expected returns. A second selection (also using LASSO) is then added to search for factors in h_t potentially missed from the first step, but that, if omitted, would induce a large omitted variable bias. Factors excluded from both stages of the double-selection procedure must have small SDF loadings and have covariances that correlate only mildly in the cross section with the covariance between factors of interest g_t and returns – these factors can be excluded with minimal omitted variable bias. This strategy results in a parsimonious model that minimizes the omitted factor bias ex ante when estimating and testing λ_g .

The regularized two-pass estimation proceeds as follows:

(1) Two-Pass Variable Selection

(1.a) Run a cross-sectional LASSO regression of average returns on sample covariances between

factors in h_t and returns:⁴

$$\min_{\gamma, \lambda} \left\{ n^{-1} \left\| \bar{r} - \iota_n \gamma - \widehat{C}_h \lambda \right\|^2 + \tau_0 n^{-1} \|\lambda\|_1 \right\}, \quad (7)$$

where $\widehat{C}_h = \widehat{\text{Cov}}(r_t, h_t) = T^{-1} \bar{R} \bar{H}^\top$.⁵ This step selects among the factors in h_t those that best explain the cross section of expected returns. Denote $\{\widehat{I}_1\}$ as the set of indices corresponding to the selected factors in this step.

- (1.b) For each factor j in g_t (with $j = 1, \dots, d$), run a cross-sectional LASSO regression of $\widehat{C}_{g, \cdot, j}$ (the covariance between returns and the j th factor of g_t) on \widehat{C}_h (the covariance between returns and all factors h_t):⁶

$$\min_{\xi_j, \chi_{j, \cdot}} \left\{ n^{-1} \left\| (\widehat{C}_{g, \cdot, j} - \iota_n \xi_j - \widehat{C}_h \chi_{j, \cdot}^\top) \right\|^2 + \tau_j n^{-1} \|\chi_{j, \cdot}^\top\|_1 \right\}. \quad (8)$$

This step identifies factors whose exposures are highly correlated to the exposures to g_t in the cross-section. This is the crucial second step in the double-selection algorithm, that searches for factors that may be missed by the first step but that may still induce large omitted variable bias in the estimation of λ_g if omitted, due to their covariance properties. Denote $\{\widehat{I}_{2, j}\}$ as the set of indices corresponding to the selected factors in the j th regression, and $\widehat{I}_2 = \bigcup_{j=1}^d \widehat{I}_{2, j}$.

(2) Post-selection Estimation

Run an OLS cross-sectional regression using covariances between the selected factors from *both* steps and returns:

$$(\widehat{\gamma}_0, \widehat{\lambda}_g, \widehat{\lambda}_h) = \arg \min_{\gamma_0, \lambda_g, \lambda_h} \left\{ \left\| \bar{r} - \iota_n \gamma_0 - \widehat{C}_g \lambda_g - \widehat{C}_h \lambda_h \right\|^2 : \lambda_{h, j} = 0, \quad \forall j \notin \widehat{I} = \widehat{I}_1 \cup \widehat{I}_2 \right\}. \quad (9)$$

We refer to this procedure as a double-selection (DS) approach, as opposed to the single-selection (SS) approach which only involves (1.a) and (2).

The LASSO estimators involve only convex optimizations, so that the implementation is quite fast. Statistical software such as R, Python, and Matlab have existing packages that implement LASSO using efficient algorithms. Note that other variable-selection procedures are also applicable. Either (1.a) or (1.b) can be replaced by other machine-learning methods such as regression tree, random forest, boosting, and neural network, as shown in [Chernozhukov et al. \(2016\)](#) for treatment-effect estimation, or by subset selection, partial least squares, and PCA regressions (or with LASSO

⁴We use $\|A\|$ and $\|A\|_1$ to denote the operator norm and the \mathbb{L}_1 norm of a matrix $A = (a_{ij})$, that is, $\sqrt{\lambda_{\max}(A^\top A)}$, $\max_j \sum_i |a_{ij}|$, where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of a matrix.

⁵For any matrix $A = (a_1 : a_2 : \dots : a_T)$, we write $\bar{a} = T^{-1} \sum_{t=1}^T a_t$, $\bar{A} = A - \iota_T \bar{a}$.

⁶For any matrix A , we use $A_{i, \cdot}$ and $A_{\cdot, j}$ to denote the i th row and j th column of A , respectively.

selection on top of PCs similar to [Kozak et al. \(2017\)](#)). They call this general procedure double machine learning. We advocate LASSO because the underlying asset pricing model is linear, the selected model is more interpretable, and its theoretical properties are more tractable.

It is useful to relate our approach to the recent model selection method by [Harvey and Liu \(2016\)](#). Their model selection procedure is an algorithm that resembles the forward stepwise regression in [Friedman et al. \(2009\)](#) (a so-called “greedy” algorithm). Their algorithm evaluates the contribution of each factor relative to a pre-selected best model through model comparison, and builds up the best model sequentially. Just like LASSO cannot deliver the true model with certainty, this algorithm cannot do so either, because it makes commitments to certain variables too early which prevent the algorithm from finding the best overall solution later. Specifically, if one of the factors in the pre-selected model is redundant relative to the factor under consideration (i.e., the latter factor is in the DGP and the former one is a noisy version of it), the latter factor could either be added or discarded depending on how noisy the former factor is. Neither scenario, however, yields a model that is closer to the truth. In any case, if this algorithm were preferred to LASSO for any reasons, we could easily substitute it in place of LASSO and still obtain correct inference, because the double machine learning procedure explicitly accounts for model selection mistakes.

Our LASSO regression contains nonnegative regularization parameters, for example, τ_j ($j = 0, 1, \dots, d$), to control the level of penalty. A higher τ_j indicates a greater penalty and hence results in a smaller model. The optimization becomes a least-squares problem if τ_j is 0. In practice, we typically test one factor each time, so that this procedure involves two regularization parameters τ_0 and τ_1 . To determine these parameters, we adopt the widely used cross-validation (CV) procedure, see [Friedman et al. \(2009\)](#).

We can also give different weights to λ_h . [Belloni et al. \(2012\)](#) recommend a data-driven method for choosing a penalty that allows for non-Gaussian and heteroskedastic disturbances. We adopt a strategy in the spirit of [Bryzgalova \(2015\)](#), which assigns weights to λ_h proportional to the inverse of the operator norm of the univariate betas of the corresponding factor in h_t . This strategy helps remove spurious factors in h_t because of a higher penalty assigned on those factors with smaller univariate betas.

2.6 Statistical Inference

We derive the asymptotic distribution of the estimator for λ_g under a jointly large n and T asymptotic design. Whereas d is fixed throughout, s and p can either be fixed or increasing. In the appendix, we prove the following theorem:

Theorem 1. *Under Assumptions [B.1 - B.6](#) in [Appendix B.2](#), if $s^2 T^{1/2} (n^{-1} + T^{-1}) \log(n \vee p \vee T) =$*

$o(1)$, we have

$$T^{1/2}(\widehat{\lambda}_g - \lambda_g) \xrightarrow{\mathcal{L}} \mathcal{N}_d(0, \Pi),$$

where the asymptotic variance is given by

$$\Pi = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \mathbb{E} \left((1 - \lambda^\top v_t)(1 - \lambda^\top v_s) \Sigma_z^{-1} z_t z_s^\top \Sigma_z^{-1} \right), \quad \Sigma_z = \text{Var}(z_t).$$

Note the asymptotic distribution of $\widehat{\lambda}_g$ does not rely on covariances (C_g, C_h) or factor loadings (β_g, β_h) of g_t and h_t , because they appear in strictly higher-order terms, which further facilitates our inference. The next theorem provides a Newey-West-type estimator of the asymptotic variance Π .

Theorem 2. *Suppose the same assumptions as in Theorem 1 hold. In addition, Assumption B.7 holds. If $qs^{3/2}(T^{-1/2} + n^{-1/2}) \|V\|_{\text{MAX}} \|Z\|_{\text{MAX}} = o_p(1)$,⁷ we have*

$$\widehat{\Pi} \xrightarrow{p} \Pi,$$

where $\widehat{\lambda} = (\widehat{\lambda}_g : \widehat{\lambda}_h)$ is given by (9), and

$$\begin{aligned} \widehat{\Pi} &= \frac{1}{T} \sum_{t=1}^T (1 - \widehat{\lambda}^\top v_t)^2 \widehat{\Sigma}_z^{-1} \widehat{z}_t \widehat{z}_t^\top \widehat{\Sigma}_z^{-1} \\ &\quad + \frac{1}{T} \sum_{k=1}^q \sum_{t=k+1}^T \left(1 - \frac{k}{q+1} \right) \left((1 - \widehat{\lambda}^\top v_t)(1 - \widehat{\lambda}^\top v_{t-k}) \widehat{\Sigma}_z^{-1} (\widehat{z}_t \widehat{z}_{t-k}^\top + \widehat{z}_{t-k} \widehat{z}_t^\top) \widehat{\Sigma}_z^{-1} \right), \\ \widehat{\Sigma}_z &= \frac{1}{T} \sum_{t=1}^T \widehat{z}_t \widehat{z}_t^\top, \quad \widehat{z}_t = g_t - \widetilde{\eta}_I^\top h_t, \quad \widetilde{\eta}_I = \arg \min_{\eta} \left\{ \|G - \eta H\|^2 : \eta_{\cdot,j} = 0, \quad j \notin \widetilde{I} \right\}, \end{aligned}$$

and \widetilde{I} is the union of selected variables using a LASSO regression of each factor in g_t on h_t :

$$\min_{\eta_j} \left\{ T^{-1} \|G_{j,\cdot} - \eta_j H\|^2 + \bar{\tau}_j T^{-1} \|\eta_j\|_1 \right\}, \quad j = 1, 2, \dots, d. \quad (10)$$

We stress that the inference procedure is valid even with imperfect model selection. That is, the selected models from (7) and (8) may omit certain useful factors and include redundant ones, which nonetheless has a negligible effect on the inference of λ_g . Using analysis similar to Belloni et al. (2014b), the results can be strengthened to hold uniformly over a sequence of data-generating processes that may vary with the sample size and only under approximately sparse conditions, so that our inference is valid without relying on perfect recovery of the correct model in finite sample.

We provide in Appendix A an extensive set of simulations that show the finite-sample performance of our estimator.

⁷We use $\|A\|_{\text{MAX}}$ to denote the \mathbb{L}_∞ -norm of A in the vector space.

3 Empirical Analysis

3.1 Data

3.1.1 The Zoo of Factors

Our factor library contains 150 risk factors at the monthly frequency for the period from July 1976 to December 2017, obtained from multiple sources. First, we download all workhorse factors in the U.S. equity market from Ken French’s data library. Then we add several published factors directly from the authors’ websites, including liquidity from [Pástor and Stambaugh \(2003a\)](#), the q-factors from [Hou et al. \(2015\)](#), and the intermediary asset pricing factors from [He et al. \(2016\)](#). We also include factors from the AQR data library, such as Betting-Against-Beta, HML Devil, and Quality-Minus-Junk. In addition to these 15 publicly available factors, we follow [Fama and French \(1993\)](#) to construct 135 long-short value-weighted portfolios as factor proxies, using firm characteristics surveyed in [Hou et al. \(2017\)](#) and [Green et al. \(2016\)](#).

To construct these factors, we include only stocks for companies listed on the NYSE, AMEX, or NASDAQ that have a CRSP share code of 10 or 11. Moreover, we exclude financial firms and firms with negative book equity. For each characteristic, we sort stocks using NYSE breakpoints based on their previous year-end values, then build and rebalance a long-short value-weighted portfolio (top 30% - bottom 30% or 1-0 dummy difference) every June for a 12-month holding period. Both [Fama and French \(2008\)](#) and [Hou et al. \(2017\)](#) have discussed the importance of using NYSE breakpoints and value-weighted portfolios. Microcaps, i.e., stocks with market equity smaller than the 20th percentile, have the largest cross-sectional dispersion in most anomalies, while accounting for only 3% of the total market equity. Equal-weighted returns overweight microcaps, despite their small economic importance.

In [Table 4](#) we report a complete list of the 150 factors and various descriptive statistics (publication years, the ending years of their sample used in the papers, monthly average returns, and annualized Sharpe ratios), as well as the references.

3.1.2 Test Portfolios

We conduct our empirical analysis on a large set of standard portfolios of U.S. equities. We target U.S. equities because of their better data quality and because they are available for a long period; however, our methodology could be applied to any set of countries or asset classes. We focus on portfolios rather than individual assets because characteristic-sorted portfolios have more stable betas, higher signal-to-noise ratios, and they are less prone to missing data issues, despite the existence of a bias-variance trade-off between the choice of portfolios and individual assets. Selecting few portfolios

based on sorts of a handful characteristics is likely to tilt the results in favor of these factors, see [Harvey and Liu \(2016\)](#). There might also be a loss in efficiency in using too few portfolios, e.g., [Litzenberger and Ramaswamy \(1979\)](#). In line with the suggestion of [Lewellen et al. \(2010\)](#), we base our analysis on a large cross section of characteristic-sorted portfolios, which helps strike a balance between having many individual stocks or a handful of portfolios.

We use a total of 750 portfolios as test assets. We start from a set of 36 portfolios: 3×2 portfolios sorted by size and book-to-market ratio, 3×2 portfolios sorted by size and operating profitability, 3×2 portfolios sorted by size and investment, 3×2 portfolios sorted by size and short-term reversal on prior (1-1) return, 3×2 portfolios sorted by size and momentum on prior (2-12) return, and 3×2 portfolios sorted by size and long-term reversal on prior (13-60) return. This set of test assets – all available from Kenneth French’s website – captures a vast cross-section of anomalies and exposures to different factors.⁸

We add to these 36 portfolios 714 additional ones obtained from our factor zoo, that cover additional characteristics. In particular, we try to include all sets of 3×2 bivariate-sorted portfolios from continuous factors in our factor zoo. These are the same sorting portfolios that are used to construct the long-short factors. For each firm characteristic, the bivariate-sorted 3×2 portfolios are constructed by intersecting its three groups with those formed on size (market equity). Notice that the number of stocks in each 3×2 group can be unbalanced in the bivariate intersection. We only include the resulting portfolios if each of the 6 groups contains a sufficient number of stocks (at least 10). This procedure gives us 119 sets of 3×2 bivariate-sorted portfolios, yielding 714 portfolios.⁹

As a robustness check, we alternatively use in our analysis the set of 202 portfolios used in [Giglio and Xiu \(2016\)](#): 25 portfolios sorted by size and book-to-market ratio, 17 industry portfolios, 25 portfolios sorted by operating profitability and investment, 25 portfolios sorted by size and variance, 35 portfolios sorted by size and net issuance, 25 portfolios sorted by size and accruals, 25 portfolios sorted by size and momentum, and 25 portfolios sorted by size and beta.

For a second robustness check, we use 1,825 5×5 bivariate-sorted portfolios instead of the 750 3×2 portfolios. We start from a standard set of 175 portfolios: 25 portfolios sorted by size and book-to-market ratio, 25 portfolios sorted by size and beta, 25 portfolios sorted by size and operating

⁸See the description of all portfolio construction on Kenneth French’s website: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

⁹There are 16 factors for which bivariate-sorted portfolios are not available. 8 of 16 are dummy or categorical characteristics, including New equity issue(28), Dividend initiation (29), Dividend omission (30), Number of earnings increases(45), Financial statements score (47), Financial statement Performance (90), Sin Stocks (122), and Convertible Debt Indicator (150). The remaining 8 of 16 have certain portfolios with less than 10 firms or have missing values: Industry-Adjusted Size (51), Dollar trading volume (53), Illiquidity (61), R&D increase (68), Corporate investment (69), Change in Short-term Investments (87), Return on net operating assets (116), and Return on assets (127).

profitability, 25 portfolios sorted by size and investment, 25 portfolios sorted by size and short-term reversal on prior (1-1) return, 25 portfolios sorted by size and momentum on prior (2-12) return, and 25 portfolios sorted by size and long-term reversal on prior (13-60) return. We then add 1,650 additional ones. The sorting procedure is same as that for the 3×2 portfolios, except that the stock universe is divided into five groups for each characteristic.

3.2 Evaluating New Factors

In this section we apply our methodology to factors that have been proposed in the last five years (2012-2016), drawing the benchmark model against which to evaluate them from the set of 135 factors that were proposed before then.¹⁰ By placing ourselves in the position of researchers evaluating “new” factors (as of 2012), we exemplify in this section how our procedure can be applied going forward as more factors are proposed. Note that we have no ex-ante reason to expect the results to go in either direction. On the one hand, given that the set of potential control factors is already extremely large, one might think that new factors are unlikely to contribute much to pricing the cross section of returns. On the other hand, we expect new research to potentially uncover better factors over time, yielding factors that improve over the existing ones.

3.2.1 The First LASSO

We start with the first step of our procedure: the cross-sectional LASSO, closely related to the dimension-reduction methods that recent papers in asset pricing have been using to tackle the factor zoo (e.g., [Kozak et al. \(2017\)](#)): the objective of this first LASSO is to select a parsimonious model that explains the cross-section of risk premia.

The advantage of applying model-selection methods like LASSO to a large set of factors is that they estimate a low-dimensional representation of the *entire* SDF. Here, we present and discuss the model selected from h_t by LASSO, since it is the first step in our procedure; but we also show empirically its fragility in selecting the model.

When we apply it in our context, LASSO indeed selects a relatively small model of the SDF, with four factors: SMB (21), Net external finance (99), Change in shares outstanding (109), and Profit margin (117). As discussed above, this first LASSO step corresponds closely to the way model-selection methods have been applied in the asset pricing literature to estimate a low-dimensional model for the SDF.

The main drawback of statistical model-selection methods is that in any finite sample they are likely to make mistakes in selecting the factors, thus yielding the wrong model. It is useful to quantify

¹⁰The most recent factors in our library were introduced in 2016.

the issue in our context, by showing empirically that LASSO is not able to robustly pin down the identity of the factors in the model.

To evaluate the robustness of the LASSO selection, we explore how it depends on the LASSO tuning parameter. Recall that, like other dimension-reduction methods, the LASSO estimator depends on a tuning parameter – the penalty parameter τ_0 . This parameter is not pinned down by theory, and must be selected by the researcher to trade off the fit and sparsity of the model. Different choices of τ_0 result in different models selected by the estimator; the estimator is robust if the conclusions (in this case: which factors get selected) do not change substantially as τ_0 varies.

A key question in this robustness exercise is to determine what is a reasonable range of values for τ_0 to consider. Of course, the estimator cannot be expected to be robust to the entire possible range of τ_0 , since setting $\tau_0 = 0$ always selects all factors, and $\tau_0 = \infty$ selects no factors at all. We propose here a procedure to select an ex-ante reasonable range of values τ_0 to evaluate the robustness of LASSO.

The starting point for our procedure is that in standard applications of machine learning, tuning parameters are typically chosen by simulating the performance of the algorithm in the data, and choosing values for the parameter for which the estimator performs the best in those simulations. We use 10-fold cross-validation (CV) to pin down the two tuning parameters of the two LASSO steps in our estimator. But these simulations are not deterministic: for example, in the case of 10-fold CV, we divide the whole sample period into 10 disjoint and random subsamples. This means that different sets of simulations will generally yield different values of the tuning parameters.

We therefore run the tuning-parameter-selection procedure multiple times, and explore the robustness of the results across different sets of simulations. In the case of the first-stage LASSO, we run 200 different 10-fold cross-validation exercises (by using 200 different randomization seeds). For each seed, the CV will choose a different value of the tuning parameter τ_0 . We then look at robustness of the selected model using these 200 different values for τ_0 . Therefore the range of possible values for τ_0 to consider in studying the robustness of LASSO is determined by the possible (random) outcomes of the CV selection. This will effectively exclude values of τ_0 that are unlikely to be optimal using the CV criterion.

Figure 1 shows, for each factor (identified by its ID), in what fraction of the 200 LASSO-selected models each factor appears. The figure shows striking variability in the model selection step. Only SMB among 135 factors is actually selected more than 70% of the time. Instead, most of the factors are selected in 1% to 20% of the cases, but not in the others.

If LASSO were able to perfectly select the true model, we should have found a small number of factors (say, 3 to 5) to be selected 100% of the time, and the remaining factors to be selected 0% of

the time. Instead, Figure 1 shows that LASSO clearly has difficulty in pinning down *which* factors are the correct ones. This exercise cautions against using simple LASSO to decide whether a factor should be included in the SDF or not.

3.2.2 The Second LASSO

To make proper inference on the marginal contribution of new factors g_t , our procedure adds a second LASSO step aimed at identifying the factors most likely to cause an omitted variable bias. Whereas the first LASSO only depends on h_t , this second LASSO depends on both g_t and h_t . This means that for every factor proposed after 2012, there will be a different set of factors selected in the second step. For reasons of space, we do not report all the factors for each g_t here.

That said, it is useful to compare the average number of factors selected at the two stages. As reported above, the first LASSO selects in our sample a very parsimonious model, with 4 factors. The second-stage LASSO, instead, tends to select between 20 and 80 control factors. The striking difference is due to the difference in the objective function for the two LASSO steps. The first step aims to explain the cross-section of expected returns; for this purpose, the CV exercise selects a very parsimonious model (i.e. a high τ_0 , indicating that a few factors go a long way in explaining the cross-section of returns). Instead, the second LASSO has the objective of selecting factors that have a high potential for omitted variable bias. Given that many factors in the control set h_t are highly correlated, this LASSO will retain many of those.

The number of factors selected by the first-stage LASSO can be interpreted as a measure of the dimensionality of the underlying asset pricing model, at least as long as the “oracle property” holds. There is, nonetheless, no theoretical relation between the number of factors selected in the second stage and the number of true asset pricing factors in the model. Any factor that could potentially bias the estimate of λ_g should be retained by the second LASSO, even redundant factors.

The fact that more factors are selected in the second stage is also consistent with the substantial randomness we observe in the first stage selection. Many factors are close cousins. Including a subset of them is more than enough, yet which subset to include depends on the subsamples. For this reason, we expect substantial uncertainty in the first stage selection, as well as a large omitted variable bias if only the first stage variables were used as controls.

3.2.3 The Double-selection (DS) Estimator

We now present our results about the marginal contribution of each factor g_t using the DS methodology. Table 1 reports the results for the factors proposed in the last five years, among which we find Quality-Minus-Junk (QMJ), Betting-Against-Beta (BAB), two investment factors, that is, CMA from Fama and French (2015) (thereafter, FF) and IA from Hou et al. (2015) (thereafter, HXZ),

two profitability factors, that is, RMW from FF and ROE from HXZ, the nontradable intermediary capital factor from [He et al. \(2016\)](#), and several factors constructed on accounting measures.

The table contains five columns of results, each reporting the point estimate of the SDF loading and the corresponding t-statistic. More specifically, the point estimate corresponds to the estimated slope of the cross-sectional regression of returns on (univariate) betas for each factor, using different methodologies to select the control factors: it represents the estimated average excess return in basis points per month of a portfolio with unit *univariate* beta with respect to that factor. This number, which we refer to as λ_s , is equal to the SDF loading λ_g but scaled to correspond to a unit beta exposure for ease of interpretation. A positive estimate for the SDF loading indicates that high values of the factor capture states of low marginal utility (good states of the world). We adjust the sign of each factor a priori, based on the economic theory, intuition, or story in the original paper that proposes this factor, so that a positive estimate should be viewed as being consistent with the economic implication. The t-statistic in each column corresponds to the test of the hypothesis that the slope is equal to zero, constructed using different methodologies across columns.

The first column reports our main result – the estimates of SDF loadings for the factors introduced since 2012, with corresponding t-statistics, obtained with our DS procedure. Most of the new factors appear statistically insignificant – our test therefore deems them redundant or useless relative to the factors introduced up to 2011. However, we still find a few important factors useful in explaining the cross section, as their estimates are significantly different from zero: in particular, profitability is strongly significant (this is true both of the version of HXZ and that of FF). HXZ’s investment factor is also significant, as are the intermediary investment and QMJ (interestingly, [Gagliardini et al. \(2017\)](#) also find empirical evidence in favor of the recently introduced factors, like investment and profitability, using a different econometric strategy.) All other factors appear statistically insignificant. These results show that our DS method can discriminate between useful and redundant factors even when the set of controls contains hundreds of factors.

The second set of results reports the estimates that one would obtain using the naive SS methodology – that is, simply using one cross-sectional LASSO to select the factors to use as controls, without the second selection step that is useful to avoid the omitted variable bias due to mistakes in model selection. The results are quite different from the DS approach, with only one factor, the convertible debt factor, appearing significant (with a negative sign); none of the other factors that appear significant with the DS method do so when using SS. Given our discussion in the previous sections, it should not be surprising that results obtained using the SS method differ from those obtained using the DS method: our theoretical results and simulations show that the SS method is biased in finite samples. This table shows that these biases play a major role empirically.

The third column shows instead what the estimates for the various factors would be if one

simply used the Fama-French three factors (Market, SMB, HML) as controls, rather than selecting the controls optimally among the myriad of potential factors. The results differ noticeably from the benchmark with double selection. 9 out of 15 factors are significant against the Fama-French three factor model. Of course, if the *true* SDF was known ex ante, selecting all and only the true factors as controls would lead to the most efficient estimate for λ_g . In practice, however, it is unlikely that we can pin down the entire SDF with certainty. The aim of our DS procedure is precisely to select the controls statistically – avoiding arbitrary choices of control factors – while at the same time minimizing the potential omitted variable bias.

The fourth column shows one more alternative way to compute SDF loadings: using standard OLS estimation including in the cross-sectional regression *all* the hundreds of potential controls. This panel therefore shows what happens if no selection is applied at all on the factors. As discussed in the previous sections, this approach is unbiased but inefficient. We expect therefore (and confirm in the table) that the results appear much more noisy and the estimates less significant than when operating variable selection through our DS method. This result highlights the importance of machine learning methods when sorting through the myriad of existing factors.

The last column of the table shows the average excess return of the factors, that is, their risk premia. This number represents the compensation investors obtain from bearing exposure to that factor, holding exposures to all other risk factors constant. As discussed, for example, in [Cochrane \(2009\)](#), the risk premium of a factor does *not* correspond to its ability to price other assets. Using the risk premium to assess the importance of a factor in a pricing model would be misleading. For example, consider two factors that are both equally exposed to the same underlying risk, plus some noise. Both factors will command an identical risk premium. Yet those factors are not *both* useful to price other assets—regardless of their level of statistical significance. The most promising way to reduce the proliferation of factors is not to look at their risk premium (no matter how significant it is), but to evaluate whether they add any pricing information to the existing factors. Our paper proposes a way to make this feasible even in a context of high dimensionality, when the set of potential control factors is large. We come back to this point in the next section.¹¹

To sum up, Table 1 shows that which factors are chosen as controls, and which econometric procedure is used for estimation, make a large difference for the conclusions about the SDF loadings and the usefulness of factors. Both the theoretical analysis and the simulations provided in this paper suggest that the DS method allows researchers to make full use of the information in the existing zoo of factors without introducing biases while accounting for efficiency losses.

¹¹It is interesting to note that about half of these factors do not have a significant risk premium, while they typically did in the original publications. This is partly due to the different sample period used here, and partly because we use a unified sorting methodology in this paper, rather than the heterogeneous methods used in the original papers. This result is consistent with the findings of [Hou et al. \(2017\)](#).

3.3 Evaluating Factors Recursively

One of the motivations for using our methodology is that it can help distinguish useful from useless and redundant factors as they are introduced in the literature. Over time, this should help limit the proliferation of factors, and retain only those new factors that actually contain novel information to price the cross section.

To illustrate this point, in each year starting in 1994 we consider the factors introduced during that year, and use our DS procedure to test whether they are useful or redundant relative to factors existing up to then. Note that the exercise is fully recursive, using only information available up to time t when evaluating a factor introduced at time t , both in choosing the set of potential controls h_t and in constructing the test portfolios (which are therefore sorted on characteristics introduced in the literature up to time t).

Table 2 reports the results. In the table, the factors introduced since 1994 are identified by their ID; the table underlines the ones that appear to be statistically significant according to our test, relative to the factors introduced before them. The table also reports the number of test assets used in each year and the number of control factors in h_t .

The results show that had our DS test been applied year by year starting in 1994, only 17 factors would have been considered useful, and a large majority would have been identified as redundant or useless.

It is useful to think about this exercise in light of the recent literature (e.g., [McLean and Pontiff \(2016\)](#), [Harvey et al. \(2015\)](#)) that has highlighted and tried to address the existence of a multitude of seemingly significant anomalies. The literature has proposed a variety of approaches, including adopting a stricter requirement for significance (such as using a threshold for the t-statistic of 3). Although the overarching theme is to tame the factor zoo, the perspectives are rather different. The aforementioned papers emphasize the bias of data-snooping or raise the concern of multiple testing, whereas our focus is on omitted controls. All these problems could contribute to the proliferation of factors.

Our approach differs from the proposals in the existing literature in four substantial ways. First, and most important, we explicitly address the problem of omitted variable bias due to potential model selection mistakes when making inference about factors' contribution to asset prices. Second, our method directly takes into account the correlation among factors, rather than considering factors individually and using Bonferroni-type bounds to assess their joint significance. We provide a statistical test of a factor's contribution with desirable asymptotic properties, as demonstrated in the previous sections, and do not rely on simulation or bootstrap methods whose statistical properties in this context are unknown. Third, our method is specifically designed to handle hundreds of factors as

controls, exploiting model-selection econometric advances to reduce the dimensionality of the factor set. Fourth, the criterion we employ for selecting factors is based on the SDF loading, not the risk premium of the factors (see a more detailed discussion on their differences in Section 3.2), as it is the right quantity to evaluate the contribution of a factor to explaining asset prices.

The various approaches that have been proposed in the literature so far address complementary issues to be overcome on the path to disciplining the zoo of factor. We leave for future research refinements of these methods that can potentially combine insights from our work and other recent papers.

Finally, it is useful to remark that this recursive exercise is simply meant as an illustration of possible applications of our method. We do not deal here with some of the potential issues that arise in ordering factors by their discovery date, such as the fact that the publication year might not capture precisely when researchers and investors first learn about the factor.¹² However, our methodology is quite general, and does not require h_t and g_t to be ordered temporally at all. For example, h_t might contain all factors obtained from equity markets, and g_t could contain factors from option markets, in which case our test could be interpreted as evaluating whether option-based factors help explain the cross section beyond what is explained by equity factors. We leave these other applications to future research.

3.4 Robustness

In this section we explore the robustness of our estimator, and discuss some extensions of our setup. The most important robustness test – presented first – is with respect to the tuning parameters, especially since we have shown in Section 3.2.1 that the first step of our procedure (the LASSO model selection) is *not* very robust to these changes.

3.4.1 Robustness to the Choice of Tuning Parameters

We explore in this section how robust our conclusions are to changes in the tuning parameters. Recall that each dimension-reduction step via LASSO depends on one tuning parameter. Our DS procedure uses LASSO in two separate steps, so two tuning parameters are needed. In this section, we focus on our benchmark estimates in Table 1, and check the robustness of our inference about the marginal contribution of the factors proposed after 2012.

Just as in Section 3.2.1, we need to decide on a reasonable range of values for the two tuning

¹²One alternative ordering that we have explored, and that is included in the internet appendix, uses the last year in each paper’s sample rather than the publication year as an alternative – though still imperfect – measure of the year in which the factor was discovered. Results are similar – 9 out of 12 factors significant in Table A1 are also significant in Table 2 – though of course the results are by construction not invariant to the ordering of factors.

parameters. We follow the procedure described before: we choose 200 different seeds for the CV simulations; for every set of simulations, we obtain one estimate for the two tuning parameters. We then look at how each λ_g 's t-statistic varies across choices of the tuning parameters. As before, this procedure ensures that we only consider values for the tuning parameters that are reasonable, in the specific sense that they are optimal given one set of CV simulations. Therefore, we exclude from the robustness analysis values of the parameters that do not maximize the cross-validation criterion for any of the 200 simulations.

We report the results of this robustness analysis in Figure 2 using heatmaps. Each panel corresponds to a different factor g_t . Different colors correspond to different levels of the t-statistics. The two axes correspond to values for the two tuning parameters (in logs).

Each panel reports 200 black dots, corresponding to a choice of tuning parameters in one of the CV simulation sets. The red cross in each graph is the average of these 200 tuning parameters, and that is the level we use to generate the baseline results (Table 1). The figure shows that inference for some factors is more robust than for others. The factors that appear significant in the baseline appear generally robust, in the sense that the vast majority of choices for the tuning parameters yield statistically significant results. Some of them (investment and profitability) appear very robust. Others, like the intermediary investment factor, do not appear very robust, in the sense that for a nontrivial subset of the tuning parameters considered, its significance vanishes. Other factors appear strong and robust, though not statistically significant at standard levels in our main results (for example, the Betting Against Beta factor). Finally, most other factors (like Growth in Advertising Expense and Fama and French's CMA) appear insignificant in the baseline, and robustly so across the range of tuning parameters. These results confirm the main conclusions of our baseline analysis, showing that a few of the recent factors appear to contribute significantly to explaining the cross-section, and most of the remaining ones are redundant or useless; at the same time, they provide a more nuanced view of the contribution of some of the factors.

Figure 3 shows the size of the selected model (the union of the factors selected at both steps of our DS procedure) as a function of the two tuning parameters. The figure shows that our 200 tuning parameters span a large subset of the parameter space: they induce the two-step selection procedure to select models as small as 0 - 5 factors and as large as 120 factors. The range of tuning parameters we consider therefore represents a statistically and economically meaningful set of possible choices.

Overall, Figures 2 and 3 are useful to refine the conclusions of our statistical analysis in Table 1, highlighting the most robust factors. We therefore recommend the use of heat maps like these to evaluate the robustness of the significant discoveries by model selection procedures like ours.

3.4.2 Robustness to Test Assets and Regularization Method

In this section we further explore the robustness of our results, with respect to the test assets used for the estimation and the machine learning methodology used to select the control factors. As before, we focus our robustness tests on the evaluation of recent factors (Table 1).

Column (1) of Table 3 reports our baseline results for convenience (as in the first column of Table 1). Column (2) shows that the results are similar when sorting the test assets in 5×5 instead of 3×2 portfolios. In Column (3), we show consistent results when using a smaller number of test assets, the 202 portfolios used in Giglio and Xiu (2016) and described in section 3.1.2.

Columns (4) and (5) show that our results also hold when using different dimension-reduction procedures. Which method is preferred in each context depends on the underlying model assumptions, and given the assumptions we make, LASSO would be the most suitable model-selection method. However, Elastic Net is a reasonable alternative to explore in this context: it combines a penalty from LASSO with that of the Ridge regression. The model selected by the Elastic Net is naturally larger, but, as column (4) in the table shows, the results are consistent with our benchmark based on pure LASSO. An alternative, following Kozak et al. (2017), is to first construct PCA of the factors, and then use LASSO on those. The results are reported in column (5). The results are statistically weaker but broadly in line with those of the benchmark specification.¹³

Overall, while across the different robustness tests the significance of some factors varies, the main conclusions of Table 1 appear quite robust to these changes in specification. That is, several of the factors introduced recently have significant additional pricing power relative to all factors introduced in the literature before 2012.

4 Conclusion

In this paper we propose a regularized two-pass cross-sectional regression approach to establish the contribution to asset pricing of a factor g_t relative to a set of control factors h_t , where the potential control set can have high dimensionality and include useless or redundant factors. Our procedure uses recent model-selection econometric techniques (specifically the double-selection procedure of Belloni et al. (2014b)) to systematically select the best control model out of the large set of factors, while explicitly taking into account model selection mistakes.

We apply this methodology to a large set of factors that the literature has proposed in the last

¹³We should remark that the standard errors and the test we built are derived exactly for the case of LASSO. In light of Chernozhukov et al. (2016), we expect the same formulae to work for other machine learning methods, such as LASSO on PCs, despite the lack of theory (they do perform well in simulations). Nonetheless, it is interesting to see that the conclusions are broadly similar.

30 years. We uncover several interesting empirical findings. First, several newly proposed factors (especially different versions of profitability) are useful in explaining asset prices, even after accounting for the large set of existing factors proposed up to 2012. Second, the SDF loadings' estimates for several factors (and the evaluation of the usefulness of those factors) are robust to changes in the tuning parameters, despite the fact that the models selected vary substantially when the tuning parameters are changed. This demonstrates how the two-step procedure is able to produce correct inference overcoming the model selection mistakes that necessarily arise when applying statistical selection methods. Third, applying our test recursively over time would have deemed only a small number of factors proposed in the literature significant. Lastly, we demonstrate how our results differ starkly from the conclusions one would obtain simply by using the risk premia of the factors or the standard Fama-French three factor model as control (as opposed to the model selection procedure we advocate).

Taken together, our results are quite encouraging about the continuing progress of asset pricing research, and suggest that studying the marginal contribution of new factors relative to the vast set of existing ones is a conservative and productive way to screen new factors and, going forward, bring discipline to the “zoo of factors.”

References

- Abarbanell, J. S. and Bushee, B. J. (1998). Abnormal returns to a fundamental analysis strategy. *Accounting Review*, pages 19–45.
- Adrian, T., Etula, E., and Muir, T. (2014). Financial intermediaries and the cross-section of asset returns. *The Journal of Finance*, 69(6):2557–2596.
- Ali, A., Hwang, L.-S., and Trombley, M. A. (2003). Arbitrage risk and the book-to-market anomaly. *Journal of Financial Economics*, 69(2):355–373.
- Almeida, H. and Campello, M. (2007). Financial constraints, asset tangibility, and corporate investment. *Review of Financial Studies*, 20(5):1429–1460.
- Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets*, 5(1):31–56.
- Amihud, Y. and Mendelson, H. (1989). The effects of beta, bid-ask spread, residual risk, and size on stock returns. *The Journal of Finance*, 44(2):479–486.
- Anderson, C. W. and Garcia-Feijoo, L. (2006). Empirical evidence on capital investment, growth options, and security returns. *The Journal of Finance*, 61(1):171–194.
- Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299.
- Asness, C. and Frazzini, A. (2013). The devil in hml’s details. *The Journal of Portfolio Management*, 39(4):49–68.
- Asness, C. S., Frazzini, A., and Pedersen, L. H. (2014). Low-risk investing without industry bets. *Financial Analysts Journal*, 70(4):24–41.
- Asness, C. S., Porter, R. B., and Stevens, R. L. (2000). Predicting stock returns using industry-relative firm characteristics. Technical report, AQR Capital Investment.
- Bai, J. and Zhou, G. (2015). Fama-macbeth two-pass regressions: Improving risk premia estimates. *Finance Research Letters*, 15:31–40.
- Balakrishnan, K., Bartov, E., and Faurel, L. (2010). Post loss/profit announcement drift. *Journal of Accounting and Economics*, 50(1):20–41.
- Bali, T. G., Cakici, N., and Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics*, 99(2):427–446.
- Bandyopadhyay, S. P., Huang, A. G., and Wirjanto, T. S. (2010). The accrual volatility anomaly. Technical report, working paper, University of Waterloo.
- Barbee Jr, W. C., Mukherji, S., and Raines, G. A. (1996). Do sales-price and debt-equity explain stock returns better than book-market and firm size? *Financial Analysts Journal*, pages 56–60.
- Barillas, F. and Shanken, J. (2018). Comparing asset pricing models. *The Journal of Finance*, 73(2):715–754.
- Barth, M. E., Elliott, J. A., and Finn, M. W. (1999). Market rewards associated with patterns of increasing earnings. *Journal of Accounting Research*, 37(2):387–413.
- Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *The Journal of Finance*, 32(3):663–682.
- Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429.

- Belloni, A. and Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19(2):521–547.
- Belloni, A., Chernozhukov, V., and Hansen, C. (2014a). High-dimensional methods and inference on structural and treatment effects. *The Journal of Economic Perspectives*, 28(2):29–50.
- Belloni, A., Chernozhukov, V., and Hansen, C. (2014b). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650.
- Belo, F. and Lin, X. (2011). The inventory growth spread. *The Review of Financial Studies*, 25(1):278–313.
- Belo, F., Lin, X., and Bazdresch, S. (2014). Labor hiring, investment, and stock return predictability in the cross section. *Journal of Political Economy*, 122(1):129–177.
- Bhandari, L. C. (1988). Debt/equity ratio and expected common stock returns: Empirical evidence. *The Journal of Finance*, 43(2):507–528.
- Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, 37(4):1705–1732.
- Bondt, W. F. and Thaler, R. (1985). Does the stock market overreact? *The Journal of Finance*, 40(3):793–805.
- Boudoukh, J., Michaely, R., Richardson, M., and Roberts, M. R. (2007). On the importance of measuring payout yield: Implications for empirical asset pricing. *The Journal of Finance*, 62(2):877–915.
- Bradshaw, M. T., Richardson, S. A., and Sloan, R. G. (2006). The relation between corporate financing activities, analysts’ forecasts and stock returns. *Journal of Accounting and Economics*, 42(1-2):53–85.
- Brandt, M. W., Kishore, R., Santa-Clara, P., and Venkatachalam, M. (2008). Earnings announcements are full of surprises. Technical report, Duke University.
- Breedon, D. T. (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics*, 7(3):265–296.
- Brown, D. P. and Rowe, B. (2007). The productivity premium in equity returns. Technical report, University of Wisconsin-Madison.
- Bryzgalova, S. (2015). Spurious factors in linear asset pricing models. Technical report, Stanford University.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82.
- Chan, L. K., Lakonishok, J., and Sougiannis, T. (2001). The stock market valuation of research and development expenditures. *The Journal of Finance*, 56(6):2431–2456.
- Chandrashekar, S. and Rao, R. K. (2009). The productivity of corporate cash holdings and the cross-section of expected stock returns. Technical report, University of Texas at Austin.
- Chen, L. and Zhang, L. (2010). A better three-factor model that explains more anomalies. *Journal of Finance*, 65(2):563–595.
- Chen, N.-F., Roll, R., and Ross, S. A. (1986). Economic forces and the stock market. *Journal of Business*, pages 383–403.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., and Newey, W. K. (2016). Double machine learning for treatment and causal parameters. Technical report, MIT.
- Chernozhukov, V., Hansen, C., and Spindler, M. (2015). Valid post-selection and post-regularization inference: An elementary, general approach. *Annual Review of Economics*, 7(1):649–688.
- Chordia, T., Subrahmanyam, A., and Anshuman, V. R. (2001). Trading activity and expected stock returns. *Journal of Financial Economics*, 59(1):3–32.
- Cochrane, J. H. (2009). *Asset Pricing:(Revised Edition)*. Princeton university press.

- Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance*, 66(4):1047–1108.
- Connor, G., Hagmann, M., and Linton, O. (2012). Efficient semiparametric estimation of the fama–french model and extensions. *Econometrica*, 80(2):713–754.
- Daniel, K. and Titman, S. (2006). Market reactions to tangible and intangible information. *The Journal of Finance*, 61(4):1605–1643.
- Datar, V. T., Naik, N. Y., and Radcliffe, R. (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets*, 1(2):203–219.
- Desai, H., Rajgopal, S., and Venkatachalam, M. (2004). Value-glamour and accruals mispricing: One anomaly or two? *The Accounting Review*, 79(2):355–385.
- Dichev, I. D. (1998). Is the risk of bankruptcy a systematic risk? *the Journal of Finance*, 53(3):1131–1147.
- Eberhart, A. C., Maxwell, W. F., and Siddique, A. R. (2004). An examination of long-term abnormal stock returns and operating performance following r&d increases. *The Journal of Finance*, 59(2):623–650.
- Eisfeldt, A. L. and Papanikolaou, D. (2013). Organization capital and the cross-section of expected returns. *The Journal of Finance*, 68(4):1365–1406.
- Fairfield, P. M., Whisenant, S., and Yohn, T. L. (2003). The differential persistence of accruals and cash flows for future operating income versus future profitability. *Review of Accounting Studies*, 8(2-3):221–243.
- Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56.
- Fama, E. F. and French, K. R. (2008). Dissecting anomalies. *The Journal of Finance*, 63(4):1653–1678.
- Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.
- Fama, E. F. and French, K. R. (2016). Choosing factors. Technical report, University of Chicago.
- Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636.
- Ferson, W. E. and Harvey, C. R. (1991). The variation of economic risk premiums. *Journal of Political Economy*, 99(2):385–415.
- Francis, J., LaFond, R., Olsson, P. M., and Schipper, K. (2004). Costs of equity and earnings attributes. *The accounting review*, 79(4):967–1010.
- Frazzini, A. and Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, 111(1):1–25.
- Freyberger, J., Neuhierl, A., and Weber, M. (2017). Dissecting characteristics nonparametrically. Technical report, University of Wisconsin-Madison.
- Friedman, J., Hastie, T., and Tibshirani, R. (2009). *The Elements of Statistical Learning - Data Mining, Inference, and Prediction, Second Edition*. New York, NY: Springer-Verlag New York.
- Gagliardini, P., Ossola, E., and Scaillet, O. (2016). Time-varying risk premium in large cross-sectional equity data sets. *Econometrica*, 84(3):985–1046.
- Gagliardini, P., Ossola, E., and Scaillet, O. (2017). A diagnostic criterion for approximate factor structure. Technical report, Swiss Finance Institute.
- Gettleman, E. and Marks, J. M. (2006). Acceleration strategies. Technical report, Bentley University.
- Gibbons, M., Ross, S., and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, 57:1121–1152.
- Giglio, S., Liao, Y., and Xiu, D. (2018). Thousands of alpha tests. Technical report, University of Chicago.

- Giglio, S. W. and Xiu, D. (2016). Asset pricing with omitted factors. Technical report, University of Chicago.
- Green, J., Hand, J. R., and Zhang, F. (2016). The characteristics that provide independent information about average u.s. monthly stock returns. Technical report, Penn State University.
- Hafzalla, N., Lundholm, R., and Matthew Van Winkle, E. (2011). Percent accruals. *The Accounting Review*, 86(1):209–236.
- Harvey, C. R. and Liu, Y. (2016). Lucky factors. Technical report, Duke University.
- Harvey, C. R., Liu, Y., and Zhu, H. (2015). ... and the cross-section of expected returns. *Review of Financial Studies*, 29(1):5–68.
- Haugen, R. A. and Baker, N. L. (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, 41(3):401–439.
- He, Z., Kelly, B., and Manela, A. (2016). Intermediary asset pricing: New evidence from many asset classes. Technical report, National Bureau of Economic Research.
- He, Z., Kelly, B., and Manela, A. (2017). Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics*, 126(1):1–35.
- Heston, S. L. and Sadka, R. (2008). Seasonality in the cross-section of stock returns. *Journal of Financial Economics*, 87(2):418–445.
- Hirshleifer, D., Hou, K., Teoh, S. H., and Zhang, Y. (2004). Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics*, 38:297–331.
- Holthausen, R. W. and Larcker, D. F. (1992). The prediction of stock returns using financial statement information. *Journal of Accounting and Economics*, 15(2-3):373–411.
- Hong, H. and Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, 93(1):15–36.
- Hou, K. and Moskowitz, T. J. (2005). Market frictions, price delay, and the cross-section of expected returns. *Review of Financial Studies*, 18(3):981–1020.
- Hou, K. and Robinson, D. T. (2006). Industry concentration and average stock returns. *The Journal of Finance*, 61(4):1927–1956.
- Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705.
- Hou, K., Xue, C., and Zhang, L. (2017). Replicating anomalies. Technical report, National Bureau of Economic Research.
- Huang, A. G. (2009). The cross section of cashflow volatility and expected stock returns. *Journal of Empirical Finance*, 16(3):409–429.
- Jagannathan, R. and Wang, Z. (1996). The conditional capm and the cross-section of expected returns. *The Journal of finance*, 51(1):3–53.
- Jegadeesh, N. and Livnat, J. (2006). Revenue surprises and stock returns. *Journal of Accounting and Economics*, 41(1-2):147–171.
- Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48(1):65–91.
- Jensen, M. C., Black, F., and Scholes, M. S. (1972). The capital asset pricing model: Some empirical tests. In *Studies in the theory of capital markets*. New York: Praeger.

- Jiang, G., Lee, C. M., and Zhang, Y. (2005). Information uncertainty and expected returns. *Review of Accounting Studies*, 10(2-3):185–221.
- Kama, I. (2009). On the market reaction to revenue and earnings surprises. *Journal of Business Finance & Accounting*, 36(1-2):31–50.
- Kelly, B., Pruitt, S., and Su, Y. (2017). Characteristics are risk exposures. Technical report, University of Chicago.
- Kozak, S., Nagel, S., and Santosh, S. (2017). Shrinking the cross section. Technical report, University of Michigan.
- Lakonishok, J., Shleifer, A., and Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *The Journal of Finance*, 49(5):1541–1578.
- Lamont, O., Polk, C., and Saaá-Requejo, J. (2001). Financial constraints and stock returns. *The review of financial studies*, 14(2):529–554.
- Leeb, H. and Pötscher, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory*, 21(01):21–59.
- Lerman, A., Livnat, J., and Mendenhall, R. R. (2008). The high-volume return premium and post-earnings announcement drift. Technical report, Yale University.
- Lettau, M. and Ludvigson, S. (2001). Resurrecting the (c) capm: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy*, 109(6):1238–1287.
- Lev, B. and Nissim, D. (2004). Taxable income, future earnings, and equity values. *The Accounting Review*, 79(4):1039–1074.
- Lewellen, J., Nagel, S., and Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial economics*, 96(2):175–194.
- Litzenberger, R. H. and Ramaswamy, K. (1979). The effects of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics*, 7:163–195.
- Liu, W. (2006). A liquidity-augmented capital asset pricing model. *Journal of Financial Economics*, 82(3):631–671.
- Lo, A. W. and MacKinlay, A. C. (1990). Data-snooping biases in tests of financial asset pricing models. *Review of financial studies*, 3(3):431–467.
- Lou, D. (2014). Attracting investor attention through advertising. *The Review of Financial Studies*, 27(6):1797–1829.
- Loughran, T. and Ritter, J. R. (1995). The new issues puzzle. *The Journal of finance*, 50(1):23–51.
- Loughran, T. and Wellman, J. W. (2011). New evidence on the relation between the enterprise multiple and average stock returns. *Journal of Financial and Quantitative Analysis*, 46(6):1629–1650.
- Lyandres, E., Sun, L., and Zhang, L. (2008). The new issues puzzle: Testing the investment-based explanation. *The Review of Financial Studies*, 21(6):2825–2855.
- McLean, R. D. and Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, 71(1):5–32.
- Meinshausen, N. and Yu, B. (2009). Lasso-type recovery of sparse representations for high-dimensional data. *The Annals of Statistics*, 37(1):246–270.
- Michaely, R., Thaler, R. H., and Womack, K. L. (1995). Price reactions to dividend initiations and omissions: Overreaction or drift? *The Journal of Finance*, 50(2):573–608.

- Miller, M. H. and Scholes, M. S. (1982). Dividends and taxes: Some empirical evidence. *Journal of Political Economy*, 90(6):1118–1141.
- Mohanram, P. S. (2005). Separating winners from losers among lowbook-to-market stocks using financial statement analysis. *Review of Accounting Studies*, 10(2-3):133–170.
- Moskowitz, T. J. and Grinblatt, M. (1999). Do industries explain momentum? *The Journal of Finance*, 54(4):1249–1290.
- Novy-Marx, R. (2011). Logical implications of gasb’s methodology for valuing pension liabilities. Technical report, National Bureau of Economic Research.
- Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, 108(1):1–28.
- Ortiz-Molina, H. and Phillips, G. M. (2014). Real asset illiquidity and the cost of capital. *Journal of Financial and Quantitative Analysis*, 49(1):1–32.
- Ou, J. A. and Penman, S. H. (1989). Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics*, 11(4):295–329.
- Palazzo, B. (2012). Cash holdings, risk, and expected returns. *Journal of Financial Economics*, 104(1):162–185.
- Pástor, L. and Stambaugh, R. F. (2003a). Liquidity risk and expected stock returns. *Journal of Political Economy*, 111(3):642–685.
- Pástor, L. and Stambaugh, R. F. (2003b). Liquidity risk and expected stock returns. *Journal of Political economy*, 111(3):642–685.
- Penman, S. H., Richardson, S. A., and Tuna, I. (2007). The book-to-price effect in stock returns: accounting for leverage. *Journal of Accounting Research*, 45(2):427–467.
- Piotroski, J. D. (2000). Value investing: The use of historical financial statement information to separate winners from losers. *Journal of Accounting Research*, pages 1–41.
- Pontiff, J. and Woodgate, A. (2008). Share issuance and cross-sectional returns. *The Journal of Finance*, 63(2):921–945.
- Rajgopal, S., Shevlin, T., and Venkatachalam, M. (2003). Does the stock market fully appreciate the implications of leading indicators for future earnings? evidence from order backlog. *Review of Accounting Studies*, 8(4):461–492.
- Raponi, V., Robotti, C., and Zaffaroni, P. (2017). Testing beta-pricing models using large cross-sections. Technical report, Imperial College London.
- Rendleman, R. J., Jones, C. P., and Latane, H. A. (1982). Empirical anomalies based on unexpected earnings and the importance of risk adjustments. *Journal of Financial Economics*, 10(3):269–287.
- Richardson, S. A., Sloan, R. G., Soliman, M. T., and Tuna, I. (2005). Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics*, 39(3):437–485.
- Shanken, J. (1992). On the estimation of beta-pricing models. *Review of Financial studies*, 5(1):1–33.
- Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *Accounting Review*, 71(3):289–315.
- Soliman, M. T. (2008). The use of dupont analysis by market participants. *The Accounting Review*, 83(3):823–853.

- Thomas, J. K. and Zhang, H. (2002). Inventory changes and future returns. *Review of Accounting Studies*, 7(2-3):163–187.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58(1):267–288.
- Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3):273–282.
- Titman, S., Wei, J., and Xie, F. (2004). Capital investments and stock returns. *Journal of Financial and Quantitative Analysis*, 39(4):677–700.
- Tuzel, S. (2010). Corporate real estate holdings and the cross-section of stock returns. *Review of Financial Studies*, 23(6):2268–2302.
- Valta, P. (2016). Strategic default, debt structure, and stock returns. *Journal of Financial and Quantitative Analysis*, 51(01):197–229.
- Wainwright, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using l_1 -constrained quadratic programming (lasso). *IEEE Transactions on Information Theory*, 55(5):2183–2202.
- Welch, I. (2008). The link between fama-french time-series tests and fama-macbeth cross-sectional tests. Technical report, UCLA.
- Whited, T. M. and Wu, G. (2006). Financial constraints risk. *The Review of Financial Studies*, 19(2):531–559.
- Xing, Y. (2008). Interpreting the value effect through the q-theory: An empirical investigation. *The Review of Financial Studies*, 21(4):1767–1795.
- Yogo, M. (2006). A consumption-based explanation of the cross-section of expected stock returns. *The Journal of Finance*, 61(2):539–580.
- Zhang, C.-H. and Huang, J. (2008). The sparsity and bias of the lasso selection in high-dimensional linear regression. *Ann. Statist.*, 36(4):1567–1594.

Table 1: Testing for Factors Introduced in 2012-2016

id	Factor Description	(1)		(2)		(3)		(4)		(5)	
		DS		SS		FF3		No Selection		Avg. Ret.	
		λ_s	tstat	λ_s	tstat	λ_s	tstat	λ_s	tstat	avg.ret.	tstat
		(bp)	(DS)	(bp)	(SS)	(bp)	(OLS)	(bp)	(OLS)	(bp)	
136	Cash holdings	-34	-0.42	15	0.17	10	0.54	-18	-0.16	13	0.98
137	HML Devil	54	1.04	-13	-0.25	-100	-2.46**	68	0.84	23	1.46
138	Gross profitability	20	0.48	3	0.06	23	2.00**	13	0.26	15	1.45
139	Organizational Capital	28	0.92	-1	-0.03	20	1.91*	16	0.41	21	2.05**
140	Betting Against Beta	35	1.45	38	1.50	36	2.25**	49	1.49	91	5.98***
141	Quality Minus Junk	73	2.03**	4	0.11	39	3.10***	50	1.04	43	3.87***
142	Employee growth	43	1.36	-4	-0.12	-12	-0.89	18	0.37	8	0.83
143	Growth in advertising	-12	-1.18	0	0.03	12	1.32	-2	-0.13	7	0.84
144	Book Asset Liquidity	40	1.07	5	0.12	20	1.59	20	0.42	9	0.79
145	RMW	160	4.45***	15	0.41	20	1.80*	74	1.48	34	3.21***
146	CMA	38	1.10	0	0.01	3	0.28	7	0.14	26	3.02***
147	HXZ IA	51	2.11**	5	0.21	21	1.94*	40	1.08	34	4.17***
148	HXZ ROE	77	3.37***	23	0.83	33	2.92***	104	2.87***	57	4.99***
149	Intermediary Risk Factor	112	2.21**	60	1.19	4	0.08	22	0.32		
150	Convertible debt	-15	-1.36	-39	-3.22***	26	3.32***	17	1.01	11	1.70*

Note. The table reports tests for the contribution of factors introduced in 2012-2016 relative to the set of 135 potential control factors introduced up to 2011. The test assets include 750 3×2 bivariate sorted portfolios published up to 2016. Sample period is from July 1976 to December 2017. For columns (1) - (4), we show the estimate of the SDF loading scaled to correspond to a unit beta exposure for ease of interpretation, λ_s , and the t-statistic. The first column uses the double-selection (DS) method, our benchmark. The tuning parameters chosen are the average of selections by 10-fold cross-validation using 200 random seeds. The second column uses the single-selection (SS) method that only controls for the first stage model. The third column uses the Fama-French three factors as controls. The fourth column uses all factors as controls, without using dimension-reduction techniques, with simple OLS. The last column reports the risk premium of each tradable factor.

Table 2: Testing Factors Recursively by Year of Publication

Year	(1)	(2)	(3)												
	# Assets	# Controls	New factors (IDs)												
1994	138	25	26	27											
1995	150	27	28	29	30										
1996	150	30	31	32	33										
1997	168	33	<u>34</u>												
1998	174	34	35	36	37	<u>38</u>	39	40	<u>41</u>	42	43	<u>44</u>			
1999	228	44	45	46											
2000	234	46	47	48	49	<u>50</u>	<u>51</u>								
2001	252	51	52	<u>53</u>	54	55	56	57	58						
2002	294	58	59	60	61										
2003	312	61	62	63	<u>64</u>	65	<u>66</u>								
2004	336	66	67	68	69	70	71	<u>72</u>	73	74					
2005	372	74	75	76	77	78	79	80	81	82	83	84	85	86	
			87	88	89	90									
2006	456	90	91	92	93	94	<u>95</u>	96	97	98	<u>99</u>	100	101	102	
2007	516	102	103	104	105	106	107	108							
2008	552	108	109	110	111	112	113	114	115	116	117	118	119	120	
2009	618	120	121	122	<u>123</u>	124									
2010	636	124	125	126	127	128	129								
2011	666	129	130	131	132	133	134	135							
2012	702	135	136												
2013	708	136	137	138	139										
2014	720	139	<u>140</u>	141	142	143	144								
2015	738	144	<u>145</u>	146	<u>147</u>	<u>148</u>									
2016	750	148	149	150											

Note. The table reports the results of a recursive factor-testing exercise, from 1994 to 2016. We test the factors using data available up to the publication year of each paper. For each year t , column (1) reports the number of test assets available for the test at that point in time, sorted on characteristics published up to then. Column (2) reports the number of controls available in each year t , i.e. the number of potential controls in h_t based on factors published up to then. Column (3) shows for each year the IDs of the factors that were published with data up to that year. We then test whether each new factor contributes to explaining asset prices relative to the factors published in previous years, using only the data up to the publication year t . We underline the IDs in column (3) every time the factor appears significant and robust based on our double-selection test. The tuning parameters chosen are the average of selections by 10-fold cross-validation using 200 random seeds.

Table 3: Robustness for Factors Introduced in 2012-2016

id	Factor Description	(1)		(2)		(3)		(4)		(5)	
		Bivariate 3×2		Bivariate 5×5		202 Portfolios		Elastic Net		PCA	
		λ_s (bp)	tstat (DS)	λ_s (bp)	tstat (DS)	λ_s (bp)	tstat (DS)	λ_s (bp)	tstat (DS)	λ_s (bp)	tstat (DS)
136	Cash holdings	-34	-0.42	34	0.40	131	0.89	-13	-0.14	-65	-0.62
137	HML Devil	54	1.04	15	0.29	56	0.57	62	1.23	-27	-0.51
138	Gross profitability	20	0.48	28	0.66	88	1.42	-11	-0.26	16	0.35
139	Organizational Capital	28	0.92	23	0.75	6	0.16	12	0.38	21	0.57
140	Betting Against Beta	35	1.45	43	1.94*	31	1.03	28	1.12	59	2.56***
141	Quality Minus Junk	73	2.03**	58	1.67	123	2.45**	74	2.13**	71	1.89*
142	Employee growth	43	1.36	12	0.34	54	1.34	51	1.49	-4	-0.09
143	Growth in advertising	-12	-1.18	6	0.57	17	1.30	9	0.74	-6	-0.57
144	Book Asset Liquidity	40	1.07	-24	-0.61	37	0.77	26	0.68	24	0.63
145	RMW	160	4.45***	104	3.13***	112	1.98**	125	3.43***	88	2.11**
146	CMA	38	1.10	19	0.59	33	0.52	32	0.85	18	0.44
147	HXZ IA	51	2.11**	44	1.87*	-45	-1.42	69	2.77***	36	1.31
148	HXZ ROE	77	3.37***	72	2.62***	116	2.22***	103	3.85***	41	1.46
149	Intermediary Risk Factor	112	2.21**	38	0.73	-16	-0.33	-16	-0.33	103	1.92*
150	Convertible debt	-15	-1.36	-6	-0.56	68	5.13***	-12	-1.08	-9	-0.88

Note. The table reports robustness tests for the estimates of SDF loadings for factors introduced in 2012-2016 relative to the set of 135 factors introduced up to 2011. The first column shows the same results as in the first column of Table 1 for convenience. The second column shows the results using bivariate-sorted 5×5 portfolios, and the third column uses 202 downloaded portfolios. In the fourth column, we use Elastic Net selection for control factors using the double-selection method. In the last column, we use the principal components of factors as controls using the double-selection method. The tuning parameters chosen are the average of selections by 10-fold cross-validation using 200 random seeds.

Table 4: Factor Zoo

ID	Description	Year.pub	Year.end	Avg.Ret.	Annual S.R.	Reference
1	Excess Market Return	1972	1965	0.64%	50.6%	Jensen et al. (1972)
2	Market Beta	1973	1968	-0.08%	-5.4%	Fama and MacBeth (1973)
3	Earnings to price	1977	1971	0.28%	29.7%	Basu (1977)
4	Dividend to price	1979	1977	0.01%	0.6%	Litzenberger and Ramaswamy (1979)
5	Unexpected quarterly earnings	1982	1980	0.12%	26.3%	Rendleman et al. (1982)
6	Share price	1982	1978	0.02%	2.2%	Miller and Scholes (1982)
7	Long-Term Reversal	1985	1982	0.34%	36.3%	Bondt and Thaler (1985)
8	Leverage	1988	1981	0.21%	24.3%	Bhandari (1988)
9	Cash flow to debt	1989	1984	-0.09%	-17.0%	Ou and Penman (1989)
10	Current ratio	1989	1984	0.06%	7.7%	Ou and Penman (1989)
11	% change in current ratio	1989	1984	0.00%	0.5%	Ou and Penman (1989)
12	% change in quick ratio	1989	1984	-0.04%	-11.9%	Ou and Penman (1989)
13	% change sales-to-inventory	1989	1984	0.17%	46.2%	Ou and Penman (1989)
14	Quick ratio	1989	1984	-0.02%	-2.9%	Ou and Penman (1989)
15	Sales to cash	1989	1984	0.01%	1.5%	Ou and Penman (1989)
16	Sales to inventory	1989	1984	0.09%	16.1%	Ou and Penman (1989)
17	Sales to receivables	1989	1984	0.14%	22.8%	Ou and Penman (1989)
18	Bid-ask spread	1989	1979	-0.04%	-3.3%	Amihud and Mendelson (1989)
19	Depreciation / PP&E	1992	1988	0.11%	12.1%	Holthausen and Larcker (1992)
20	% change in depreciation	1992	1988	0.08%	23.1%	Holthausen and Larcker (1992)
21	Small Minus Big	1993	1991	0.21%	24.5%	Fama and French (1993)
22	High Minus Low	1993	1991	0.28%	34.3%	Fama and French (1993)
23	Short-Term Reversal	1993	1989	0.15%	21.7%	Jegadeesh and Titman (1993)
24	6-month momentum	1993	1989	0.21%	27.8%	Jegadeesh and Titman (1993)
25	36-month momentum	1993	1989	0.09%	13.4%	Jegadeesh and Titman (1993)
26	Sales growth	1994	1990	0.04%	5.8%	Lakonishok et al. (1994)
27	Cash flow-to-price	1994	1990	0.31%	32.5%	Lakonishok et al. (1994)
28	New equity issue	1995	1990	0.10%	8.7%	Loughran and Ritter (1995)
29	Dividend initiation	1995	1988	-0.03%	-3.4%	Michaely et al. (1995)
30	Dividend omission	1995	1988	-0.18%	-18.0%	Michaely et al. (1995)

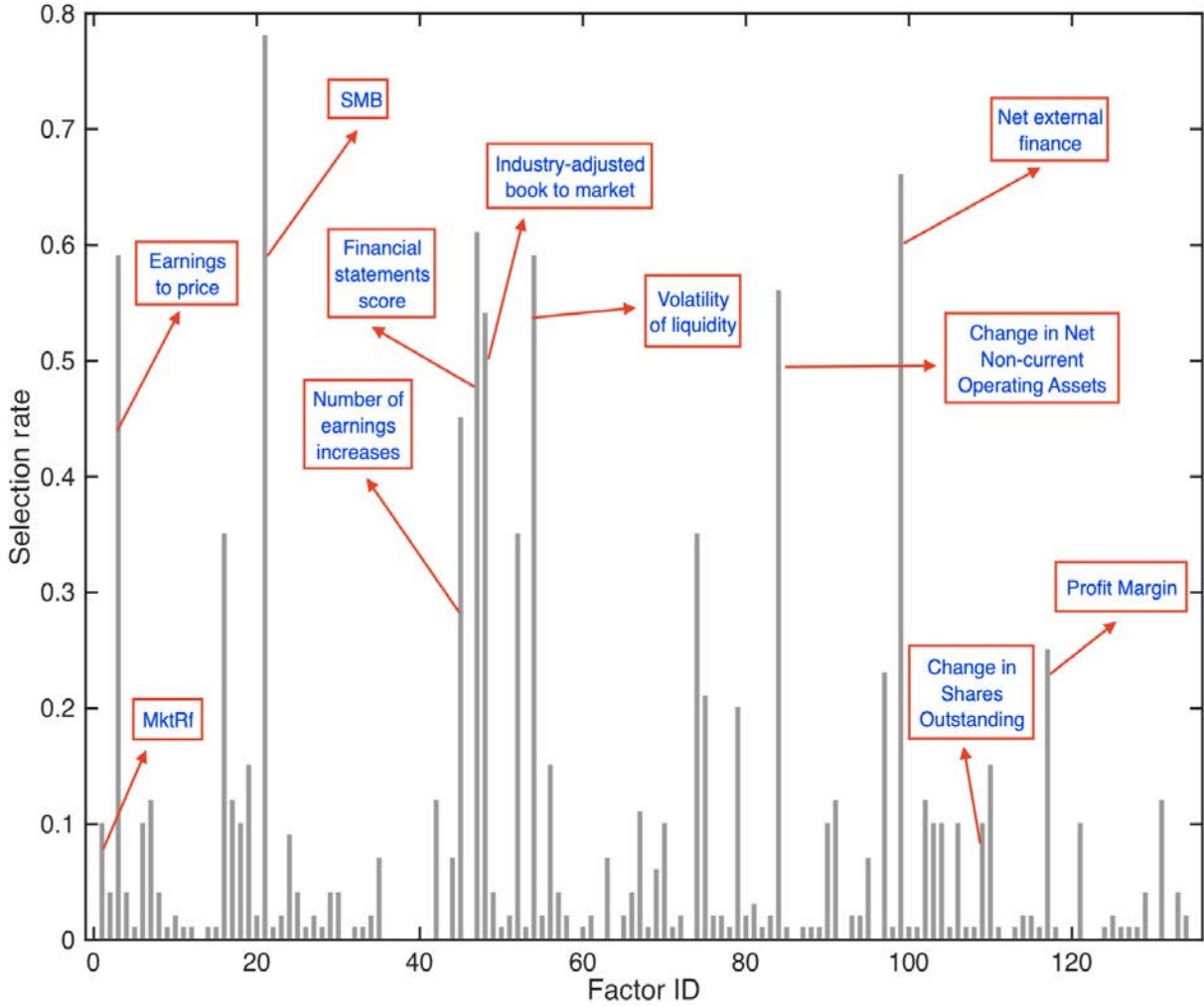
Note. The factor zoo contains 150 tradable factors for monthly data from July 1976 to December 2017. In addition to these publicly available factors, we follow [Fama and French \(1993\)](#) to construct value-weighted portfolios as factors using firm characteristics collected in [Green et al. \(2016\)](#) and [Hou et al. \(2017\)](#). In the table, we have listed the factor publication year, the end year of the test sample in the original paper, the monthly average return, the annualized Sharpe ratios, and the paper references.

ID	Description	Year.pub	Year.end	Avg.Ret.	Annual S.R.	Reference
31	Working capital accruals	1996	1991	0.22%	46.0%	Sloan (1996)
32	Sales to price	1996	1991	0.35%	41.8%	Barbee Jr et al. (1996)
33	Capital turnover	1996	1993	-0.11%	-16.6%	Haugen and Baker (1996)
34	Momentum	1997	1993	0.63%	50.2%	Carhart (1997)
35	Share turnover	1998	1991	-0.02%	-2.1%	Datar et al. (1998)
36	% change in gross margin - % change in sales	1998	1988	-0.05%	-12.4%	Abarbanell and Bushee (1998)
37	% change in sales - % change in inventory	1998	1988	0.14%	42.1%	Abarbanell and Bushee (1998)
38	% change in sales - % change in A/R	1998	1988	0.14%	43.5%	Abarbanell and Bushee (1998)
39	% change in sales - % change in SG&A	1998	1988	0.09%	19.6%	Abarbanell and Bushee (1998)
40	Effective Tax Rate	1998	1988	-0.04%	-9.1%	Abarbanell and Bushee (1998)
41	Labor Force Efficiency	1998	1988	-0.03%	-8.5%	Abarbanell and Bushee (1998)
42	Ohlson's O-score	1998	1995	0.05%	9.3%	Dichev (1998)
43	Altman's Z-score	1998	1995	0.20%	22.1%	Dichev (1998)
44	Industry adjusted % change in capital expenditures	1998	1988	0.10%	20.5%	Abarbanell and Bushee (1998)
45	Number of earnings increases	1999	1992	0.01%	2.8%	Barth et al. (1999)
46	Industry momentum	1999	1995	0.01%	1.4%	Moskowitz and Grinblatt (1999)
47	Financial statements score	2000	1996	0.08%	18.4%	Piotroski (2000)
48	Industry-adjusted book to market	2000	1998	0.22%	38.0%	Asness et al. (2000)
49	Industry-adjusted cash flow to price ratio	2000	1998	0.26%	52.1%	Asness et al. (2000)
50	Industry-adjusted change in employees	2000	1998	-0.01%	-1.5%	Asness et al. (2000)
51	Industry-adjusted size	2000	1998	0.36%	36.3%	Asness et al. (2000)
52	Dollar trading volume	2001	1995	0.38%	35.8%	Chordia et al. (2001)
53	Volatility of liquidity (dollar trading volume)	2001	1995	0.20%	38.8%	Chordia et al. (2001)
54	Volatility of liquidity (share turnover)	2001	1995	0.02%	2.1%	Chordia et al. (2001)
55	Advertising Expense-to-market	2001	1995	-0.13%	-15.6%	Chan et al. (2001)
56	R&D Expense-to-market	2001	1995	0.34%	36.2%	Chan et al. (2001)
57	R&D-to-sales	2001	1995	0.06%	5.5%	Chan et al. (2001)
58	Kaplan-Zingales Index	2001	1997	0.22%	25.3%	Lamont et al. (2001)
59	Change in inventory	2002	1997	0.18%	40.7%	Thomas and Zhang (2002)
60	Change in tax expense	2002	1997	0.09%	18.0%	Thomas and Zhang (2002)
61	Illiquidity	2002	1997	0.34%	28.6%	Amihud (2002)
62	Liquidity	2003	2000	0.38%	38.6%	Pástor and Stambaugh (2003b)
63	Idiosyncratic return volatility	2003	1997	0.07%	5.1%	Ali et al. (2003)
64	Growth in long term net operating assets	2003	1993	0.22%	51.8%	Fairfield et al. (2003)
65	Order backlog	2003	1999	0.05%	5.7%	Rajgopal et al. (2003)
66	Changes in Long-term Net Operating Assets	2003	1993	0.24%	56.0%	Fairfield et al. (2003)
67	Cash flow to price ratio	2004	1997	0.27%	31.7%	Desai et al. (2004)
68	R&D increase	2004	2001	0.06%	11.1%	Eberhart et al. (2004)
69	Corporate investment	2004	1995	0.13%	36.4%	Titman et al. (2004)
70	Earnings volatility	2004	2001	0.10%	10.7%	Francis et al. (2004)

ID	Description	Year.pub	Year.end	Avg.Ret.	Annual S.R.	Reference
71	Abnormal Corporate Investment	2004	1995	0.13%	31.2%	Titman et al. (2004)
72	Net Operating Assets	2004	2002	0.31%	66.6%	Hirshleifer et al. (2004)
73	Changes in Net Operating Assets	2004	2002	0.14%	41.6%	Hirshleifer et al. (2004)
74	Tax income to book income	2004	2000	0.14%	28.3%	Lev and Nissim (2004)
75	Price delay	2005	2001	0.07%	16.8%	Hou and Moskowitz (2005)
76	# Years since first Compustat coverage	2005	2001	0.01%	1.1%	Jiang et al. (2005)
77	Growth in common shareholder equity	2005	2001	0.15%	27.6%	Richardson et al. (2005)
78	Growth in long-term debt	2005	2001	0.06%	13.3%	Richardson et al. (2005)
79	Change in Current Operating Assets	2005	2001	0.19%	34.6%	Richardson et al. (2005)
80	Change in Current Operating Liabilities	2005	2001	0.03%	6.3%	Richardson et al. (2005)
81	Changes in Net Non-cash Working Capital	2005	2001	0.11%	25.2%	Richardson et al. (2005)
82	Change in Non-current Operating Assets	2005	2001	0.21%	44.5%	Richardson et al. (2005)
83	Change in Non-current Operating Liabilities	2005	2001	0.04%	9.6%	Richardson et al. (2005)
84	Change in Net Non-current Operating Assets	2005	2001	0.23%	35.4%	Richardson et al. (2005)
85	Change in Net Financial Assets	2005	2001	0.23%	59.0%	Richardson et al. (2005)
86	Total accruals	2005	2001	0.19%	44.8%	Richardson et al. (2005)
87	Change in Short- term Investments	2005	2001	-0.03%	-8.3%	Richardson et al. (2005)
88	Change in Financial Liabilities	2005	2001	0.18%	56.1%	Richardson et al. (2005)
89	Change in Book Equity	2005	2001	0.17%	30.0%	Richardson et al. (2005)
90	Financial statements performance	2005	2001	0.17%	37.1%	Mohanram (2005)
91	Change in 6-month momentum	2006	2006	0.21%	29.8%	Gettleman and Marks (2006)
92	Growth in capital expenditures	2006	1999	0.14%	30.4%	Anderson and Garcia-Feijoo (2006)
93	Return volatility	2006	2000	-0.02%	-1.7%	Ang et al. (2006)
94	Zero trading days	2006	2003	-0.05%	-4.4%	Liu (2006)
95	Three-year Investment Growth	2006	1999	0.11%	23.6%	Anderson and Garcia-Feijoo (2006)
96	Composite Equity Issuance	2006	2003	-0.01%	-2.2%	Daniel and Titman (2006)
97	Net equity finance	2006	2000	0.08%	9.7%	Bradshaw et al. (2006)
98	Net debt finance	2006	2000	0.17%	48.3%	Bradshaw et al. (2006)
99	Net external finance	2006	2000	0.22%	38.6%	Bradshaw et al. (2006)
100	Revenue Surprises	2006	2003	0.05%	9.0%	Jegadeesh and Livnat (2006)
101	Industry Concentration	2006	2001	0.03%	3.8%	Hou and Robinson (2006)
102	Whited-Wu Index	2006	2001	-0.02%	-2.6%	Whited and Wu (2006)
103	Return on invested capital	2007	2005	0.18%	29.3%	Brown and Rowe (2007)
104	Debt capacity/firm tangibility	2007	2000	0.05%	7.1%	Almeida and Campello (2007)
105	Payout yield	2007	2003	0.16%	17.5%	Boudoukh et al. (2007)
106	Net payout yield	2007	2003	0.16%	17.2%	Boudoukh et al. (2007)
107	Net debt-to-price	2007	1950	0.02%	2.5%	Penman et al. (2007)
108	Enterprise book-to-price	2007	2001	0.14%	14.7%	Penman et al. (2007)
109	Change in shares outstanding	2008	1969	0.24%	36.1%	Pontiff and Woodgate (2008)
110	Abnormal earnings announcement volume	2008	2006	-0.08%	-17.0%	Lerman et al. (2008)

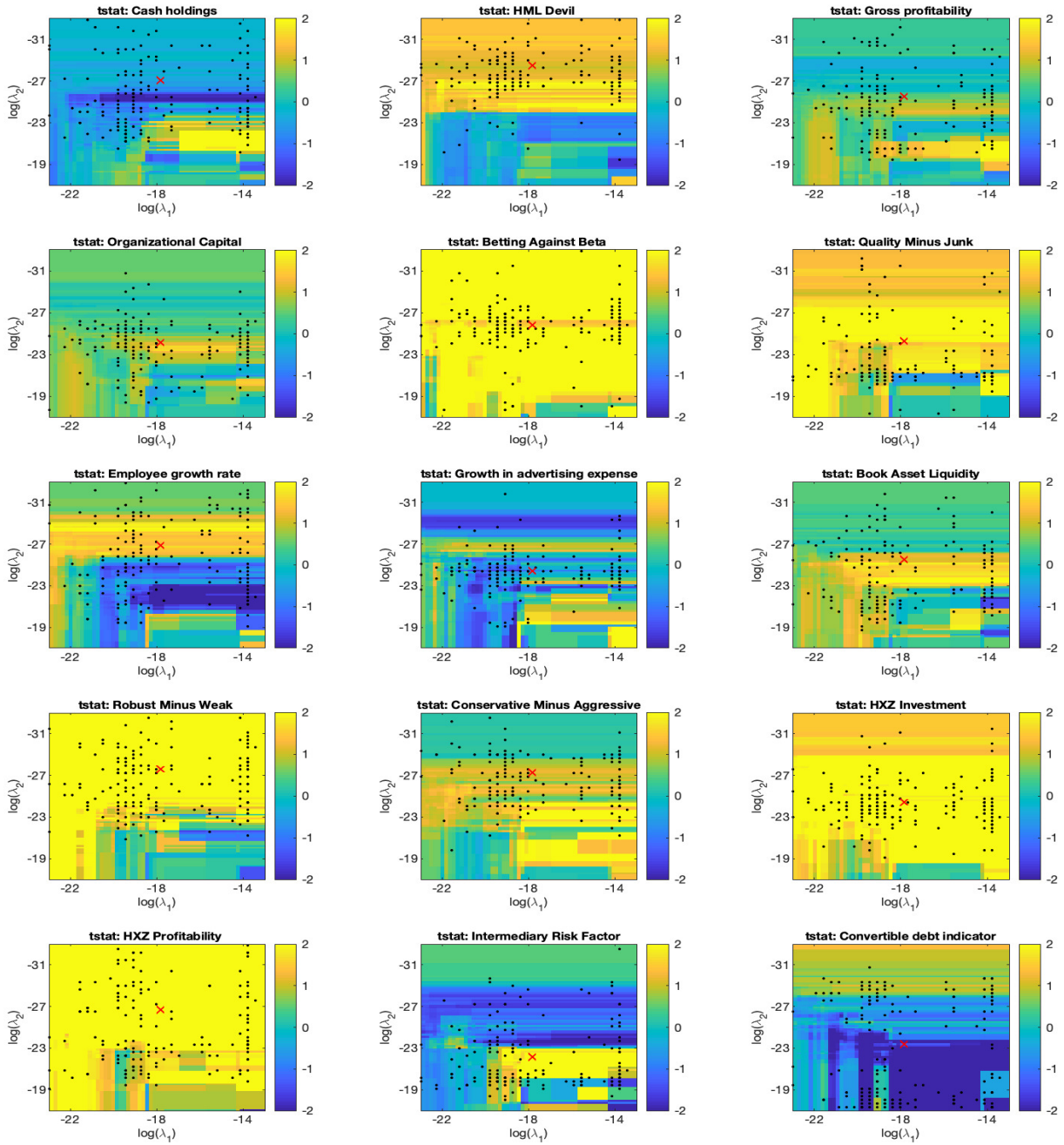
ID	Description	Year.pub	Year.end	Avg.Ret.	Annual S.R.	Reference
111	Earnings announcement return	2008	2004	0.02%	6.8%	Brandt et al. (2008)
112	Seasonality	2008	2002	0.16%	17.3%	Heston and Sadka (2008)
113	Changes in PPE and Inventory-to-assets	2008	2005	0.19%	42.0%	Lyandres et al. (2008)
114	Investment Growth	2008	2003	0.17%	39.5%	Xing (2008)
115	Composite Debt Issuance	2008	2005	0.08%	21.6%	Lyandres et al. (2008)
116	Return on net operating assets	2008	2002	0.09%	8.6%	Soliman (2008)
117	Profit margin	2008	2002	0.02%	4.4%	Soliman (2008)
118	Asset turnover	2008	2002	0.06%	6.7%	Soliman (2008)
119	Industry-adjusted change in asset turnover	2008	2002	0.14%	41.1%	Soliman (2008)
120	Industry-adjusted change in profit margin	2008	2002	-0.01%	-3.2%	Soliman (2008)
121	Cash productivity	2009	2009	0.27%	37.6%	Chandrashekar and Rao (2009)
122	Sin stocks	2009	2006	0.44%	41.6%	Hong and Kacperczyk (2009)
123	Revenue surprise	2009	2005	0.12%	19.3%	Kama (2009)
124	Cash flow volatility	2009	2008	0.20%	26.6%	Huang (2009)
125	Absolute accruals	2010	2008	-0.05%	-8.6%	Bandyopadhyay et al. (2010)
126	Capital expenditures and inventory	2010	2006	0.19%	42.8%	Chen and Zhang (2010)
127	Return on assets	2010	2005	-0.09%	-13.9%	Balakrishnan et al. (2010)
128	Accrual volatility	2010	2008	0.19%	26.6%	Bandyopadhyay et al. (2010)
129	Industry-adjusted Real Estate Ratio	2010	2005	0.11%	17.3%	Tuzel (2010)
130	Percent accruals	2011	2008	0.16%	35.0%	Hafzalla et al. (2011)
131	Maximum daily return	2011	2005	0.00%	-0.3%	Bali et al. (2011)
132	Operating Leverage	2011	2008	0.20%	32.8%	Novy-Marx (2011)
133	Inventory Growth	2011	2009	0.13%	30.1%	Belo and Lin (2011)
134	Percent Operating Accruals	2011	2008	0.15%	28.9%	Hafzalla et al. (2011)
135	Enterprise multiple	2011	2009	0.11%	17.6%	Loughran and Wellman (2011)
136	Cash holdings	2012	2009	0.13%	15.3%	Palazzo (2012)
137	HML Devil	2013	2011	0.23%	22.6%	Asness and Frazzini (2013)
138	Gross profitability	2013	2010	0.15%	22.5%	Novy-Marx (2013)
139	Organizational Capital	2013	2008	0.21%	31.9%	Eisfeldt and Papanikolaou (2013)
140	Betting Against Beta	2014	2012	0.91%	92.8%	Frazzini and Pedersen (2014)
141	Quality Minus Junk	2014	2012	0.43%	60.1%	Asness et al. (2014)
142	Employee growth rate	2014	2010	0.08%	12.9%	Belo et al. (2014)
143	Growth in advertising expense	2014	2010	0.07%	13.0%	Lou (2014)
144	Book Asset Liquidity	2014	2006	0.09%	12.3%	Ortiz-Molina and Phillips (2014)
145	Robust Minus Weak	2015	2013	0.34%	49.8%	Fama and French (2015)
146	Conservative Minus Aggressive	2015	2013	0.26%	46.8%	Fama and French (2015)
147	HXZ Investment	2015	2012	0.34%	64.7%	Hou et al. (2015)
148	HXZ Profitability	2015	2012	0.57%	77.5%	Hou et al. (2015)
149	Intermediary Investment	2016	2012			He et al. (2017)
150	Convertible debt indicator	2016	2012	0.11%	26.4%	Valta (2016)

Figure 1: Subsamples: Factor 1st Selection Rate



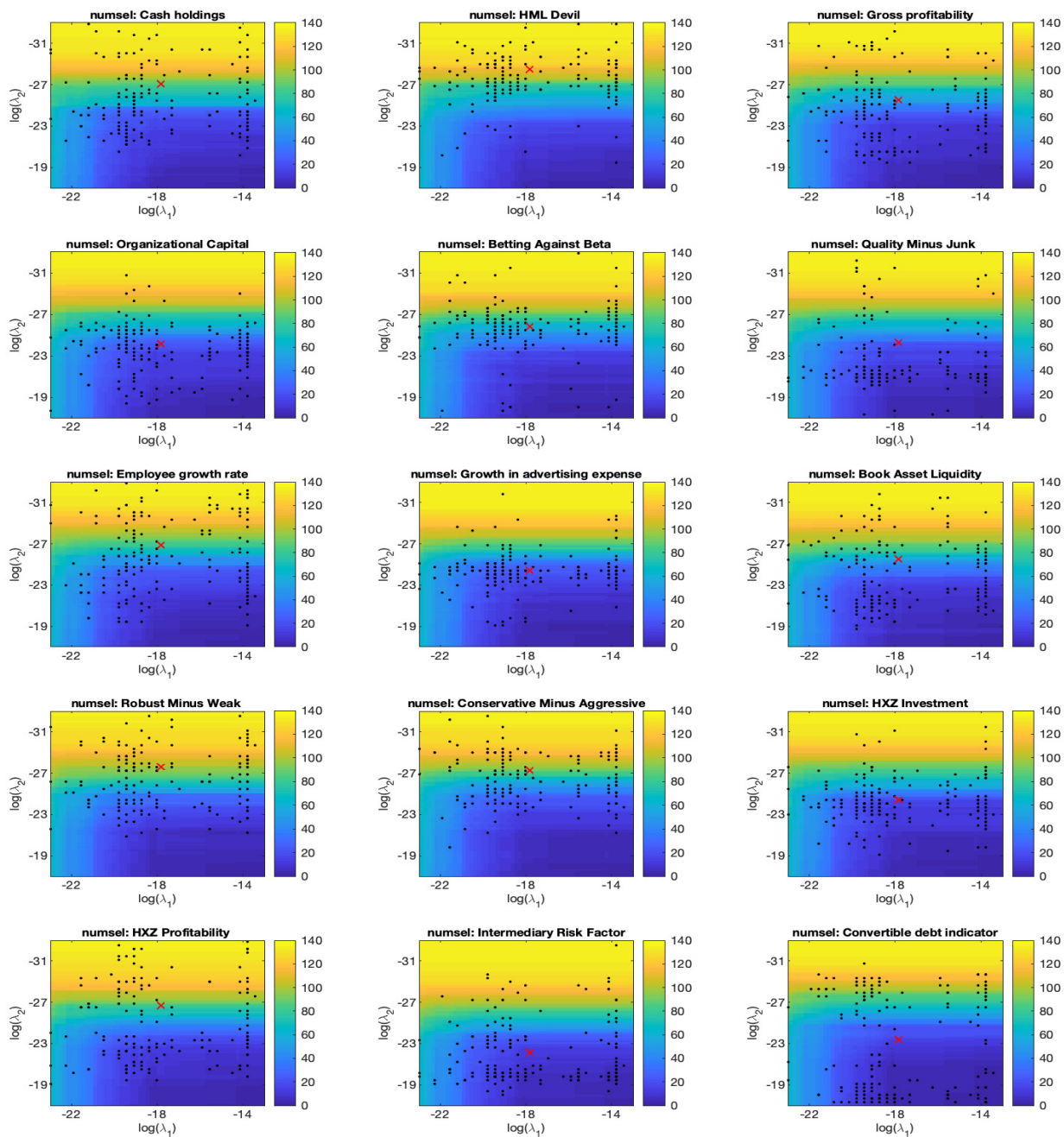
Note. The figure reports the control factor selection rates for the tests of Table 1 (i.e., the factors selected by the first LASSO step of the double-selection procedure by cross-validation), across 200 random seeds shown in the heat maps (corresponding to the 200 black dots). The figure shows, for each factor identified by the factor ID (on the X axis), in what fraction of the 200 random seeds each factor is selected by cross-validation.

Figure 2: Factors Introduced in 2012-2016: Robustness to Tuning Parameters (t-statistics)



Note. The figures provide heat maps for double-selection tests of factors introduced in 2012-2016, as in the first column of Table 1, using a wide range of tuning parameters, for the first LASSO stage on the X axis and for the second stage on the Y axis. The t-statistics for each factor in different models are shown on the heat maps. The dots are the results of 200 time-series cross-validation estimations of the tuning parameter. The red “x” is the average of the 200 black dots, which corresponds to the model used in Table 1.

Figure 3: Factors Introduced in 2012-2016: Robustness to Tuning Parameters (# selected controls)



Note. The figures provide heat maps for double-selection tests of factors introduced in 2012-2016, as in the first column of Table 1, using a wide range of tuning parameters, for the first LASSO stage on the X axis and for the second stage on the Y axis. The numbers of controls selected for each factor are shown in the heat maps. The dots are the results for 200 time-series cross-validation estimations of the tuning parameter. The red “x” is the average of the 200 black dots, which corresponds to the model used in Table 1.