The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXV, NO. 3 • JUNE 2020

Taming the Factor Zoo: A Test of New Factors

GUANHAO FENG, STEFANO GIGLIO, and DACHENG XIU*

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 accounts for *model selection mistakes* that produce a bias due to omitted variables, unlike standard approaches that assume perfect variable selection. We apply our procedure to a set of factors recently discovered in the literature. While most of these new factors are shown to be redundant relative to the existing factors, a few have statistically significant explanatory power beyond the hundreds of factors proposed in the past.

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 more recently by Harvey, Liu, and Zhu (2015), McLean and Pontiff (2016), and Hou, Xue, and Zhang (2017). A fundamental task facing the asset pricing field today is to bring more discipline to the proliferation of factors. In particular, a question that remains open is: how to judge whether a new factor adds explanatory power for asset pricing, relative to the hundreds of factors the literature has so far produced?

This paper provides a framework for systematically evaluating the contribution of individual factors relative to existing factors as well as for conducting appropriate statistical inference in this high-dimensional setting. More specifically, we provide a methodology for estimating and testing the marginal importance of any factor g_t in pricing the cross section of expected returns *beyond* what can be explained by a high-dimensional set of potential factors h_t, where

*Guanhao Feng is from City University of Hong Kong College of Business. Stefano Giglio is from Yale School of Management, NBER, and CEPR. Dacheng Xiu is from University of Chicago Booth School of Business. 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 authors have read *The Journal of Finance* disclosure policy and have no conflicts of interest to disclose.

Correspondence: Stefano Giglio, Department of Finance, Yale School of Management, NBER, and CEPR, 165 Whitney Avenue, New Haven, CT 06520; e-mail: stefano.giglio@yale.edu.

DOI: 10.1111/jofi.12883

g_t and h_t can be tradable or nontradable factors. We assume that the true asset pricing model is approximately low-dimensional. However, in addition to relevant asset pricing factors, g_t and h_t include both redundant factors that add no explanatory power to the model, as well as useless ones that have no explanatory power at all. We select the relevant factors from h_t and conduct proper inference on the contribution of g_t above and beyond those factors. Our methodology can be thought of as a conservative test for new factors, which benchmarks them against a large-dimensional set of existing factors.

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 tells us not only whether g_t is useful for pricing the cross section, but also how shocks to g_t affect marginal utility, which has a direct economic interpretation.

When h_t consists of potentially hundreds of factors, 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., the least absolute shrinkage and selection operator [LASSO]) can be useful in selecting the correct variables under certain conditions and thereby reduce 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 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, Chernozhukov, and Hansen (2014b)) with two-pass regressions such as Fama-MacBeth to evaluate the contribution of a factor to explaining asset prices 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 the factors that are useful *either* in explaining the cross section of expected returns *or* 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 thus is an ideal data set to conduct our empirical analysis. 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 three-factor model (t-statistic = 2.06) in our sample. Thus, 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 instance, it has a correlation of 0.63 with the Carhart momentum factor). Moreover, if one evaluates it against a model that includes momentum (like the Fama-French four-factor model), the alpha becomes small and statistically insignificant (t-statistic $= -0.87$). This example highlights the importance of the benchmark in evaluating new factors. Most papers in the literature that aim to produce new factors choose the benchmark model somewhat arbitrarily, subject to potential data-mining bias. Our procedure systematically constructs the best low-dimensional benchmark to evaluate new factors using the entire factor zoo.

We perform several empirical exercises that illustrate the use of our procedure in the data. First, we start by evaluating the marginal contribution of factors proposed over the last five years (2012 to 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, Xue, and Zhang (2015), and the intermediary-based factors of He, Kelly, and Manela (2017). Note that our test is conservative; it requires that a new factor g_t contributes to 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 several of the newly proposed factors (e.g., profitability) indeed have significant marginal explanatory power for expected returns.

Second, we conduct a recursive exercise in which factors are tested as they are introduced against previously proposed factors. This exercise shows that our procedure would deem most factors as redundant or spurious, finding significance for a small number of factors. Over time, our procedure would screen 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 explore an alternative application of our procedure in which some factors are determined ex ante to be part of the benchmark h_t, and the remaining factors are individually tested and added recursively (similar in spirit to forward stepwise selection), expanding the set of "preselected" factors in the benchmark at each iteration until no remaining factors contribute to the expected return variation.

Finally, 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, such as Elastic Net and principal component analysis (PCA), as well as using the stepwise procedure to select the benchmark. We also show that the results are robust to alternative portfolio constructions. Most importantly, we explore robustness with respect to the tuning parameters. Like all machine learning methods, our procedure involves the choice of tuning parameters (in particular, two tuning parameters, one for each selection step). In our main analysis, we choose them by cross-validation (CV). We show that our empirical findings are robust to varying the tuning parameters in the neighborhood of the values chosen by the CV procedure.

The double-selection (DS) estimation procedure we propose, which combines cross-sectional asset pricing regressions with the DS LASSO of Belloni, Chernozhukov, and Hansen (2014b) (originally designed 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 occurs 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 the factors in h_t. This model selection step corresponds closely to the approach used in the current literature to address the proliferation of asset pricing factors (e.g., Kozak, Nagel, and Santosh (2018)): 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, interpretation of the selected model in the literature has relied on the so-called "oracle property" of LASSO and other model selection methods, namely, an asymptotic property that guarantees that under certain assumptions, as the sample size goes to infinity, the procedures will eventually recover the true model.

In this paper, however, we take a step forward by recognizing 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. The failure of the oracle property in finite samples is a problem for addressing the question of interest in this paper: that is, whether a new factor g_t improves the explanatory power 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, but whose covariance with returns is 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, e.g., Leeb and Pötscher (2005)) and 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. To the set of controls selected by the first-stage LASSO, the second step adds 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 be sure to include even factors with small in-sample SDF loadings if omitting them may 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 possible that some variables selected from the second stage are redundant or even useless but their inclusion leads to only 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 for a standard two-pass regression using g_t and the selected control factors from h_t. This postselection estimation step is also useful to remove bias arising from regularization in any LASSO procedure; see, for example, Friedman, Hastie, and Tibshirani (2009). We then conduct asymptotic inference on the SDF loading of g_t using a central-limit result that we derive in this paper. In simulation, we show that our estimator performs well in finite samples, and outperforms alternative estimators substantially.

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, that is, should enter investors' marginal utility. However, theoretical models are often very stylized, and their ability to explain the cross section is limited. This suggests that, in reality, investors may care about 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 factors not explicitly incorporated in the estimation, then the estimate for the loading on g_t will be biased. Our methodology, which estimates the loading of g_t while taking a stand on the "omitted factors" by choosing them from the large set h_t, can be viewed as a way to address this omitted factor concern. In this sense, our approach is related to Giglio and Xiu (2016), who 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* (i.e., the compensation investors require for holding 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, but distinct, economic interpretations. Accordingly, they have different theoretical properties, and thus 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 ability of factors to explain asset prices (see Cochrane (2009)), and thus SDF loadings are the appropriate concept to focus on in disciplining the zoo of factors.

Our paper builds on several strands of the asset pricing and econometrics literatures. In addition to a large literature devoted to identifying asset pricing factors[1] and a vast econometrics literature that estimates factor models,[2] our paper is most closely related to recent literature on the high dimensionality of cross-sectional asset pricing models. Green, Hand, and Zhang (2017) test 94 firm characteristics using Fama-Macbeth regressions and find that 8 to 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, Liu, and Zhu (2015) adopt a multiple testing framework to reevaluate past research and suggest a new benchmark for current and future factor fishing. Building on this multiple-testing framework, Harvey and Liu (2016) provide a bootstrap technique to model selection. Recently, Freyberger, Neuhierl, and Weber (2020) propose a group LASSO procedure to select characteristics and to estimate how they affect expected returns nonparametrically. Kozak, Nagel, and Santosh (2018) 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 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, for example, in the GRS test of Gibbons, Ross, and Shanken (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) address.[3] Our procedure also helps alleviate the concern of data-snooping, which is another form of multiple testing (see, e.g., Lo and MacKinlay (1990), Harvey, Liu, and Zhu (2015)), because we suggest imposing discipline to the selection of controls (as opposed to the conventional practice of selecting arbitrary controls, which leaves the researcher much more freedom).

[1] Some of the factors proposed in the literature are based on economic theory (e.g., Breeden (1979), Chen, Roll, and Ross (1986), Jagannathan and Wang (1996), Lettau and Ludvigson (2001), Pástor and Stambaugh (2003), Yogo (2006), Adrian, Etula, and Muir (2014), He, Kelly, and Manela (2017)); others are constructed from firm characteristics, for example, Fama and French (1993, 2015), Carhart (1997), and Hou, Xue, and Zhang (2015).

[2] See, among the many papers, Jensen, Black, and Scholes (1972), Fama and MacBeth (1973), Ferson and Harvey (1991), Shanken (1992), Jagannathan and Wang (1996), Welch (2008), and Lewellen, Nagel, and Shanken (2010). These papers, along with the majority of the literature, rely on large T and fixed n asymptotic analysis for statistical inference and only consider models in which all factors are specified and observable. Recent literature, including Bai and Zhou (2015), Gagliardini, Ossola, and Scaillet (2016), Gagliardini, Ossola, and Scaillet (2019), Connor, Hagmann, and Linton (2012), Giglio and Xiu (2016), and Raponi, Robotti, and Zaffaroni (2017), relies on alternative asymptotic designs for better small-sample performance and robustness to model misspecification.

[3] The two methodologies could potentially be combined to produce more conservative inference that also addresses 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 question for future research. Relatedly, Giglio, Liao, and Xiu (2018) tackle the multiple testing of alphas in a linear asset pricing model.

Of course, 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 on existing factors (e.g., Barillas and Shanken (2018) and Fama and French (2018)). Our methodology differs from existing procedures in several ways. First, we do not use an arbitrary set of factors from h_t (e.g., the three Fama-French factors) as control, but rather we select from h_t the control model that best explains the cross section of returns. Second, we not only test whether the factor of interest g_t is useful in explaining asset prices, but also estimate its role in driving marginal utility (its coefficient in the SDF). This is important to be able to interpret the results in economic terms and to relate them to the models that motivated the choice of g_t. Third, our procedure handles both traded and nontraded 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 time series.

Our paper is also related to a large statistical and machine-learning literature on variable selection and regularization using LASSO and postselection inference. For details on theoretical properties of LASSO, see Bickel, Ritov, and Tsybakov (2009), Meinshausen and Yu (2009), Tibshirani (2011), Wainwright (2009), Zhang and Huang (2008), and Belloni and Chernozhukov (2013). For details on the postselection-inference method, see, for example, Belloni et al. (2012) and Belloni, Chernozhukov, and Hansen (2014b), and review articles by Belloni, Chernozhukov, and Hansen (2014a) and Chernozhukov et al. (2015). Our asymptotic results are new to the 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 selects covariances between factors and returns, which are contaminated by estimation errors, rather than the immediately observable factors.

The rest of the paper is organized as follows. In Section I, we discuss the model, present our methodology, and develop relevant statistical inference. In Section II, we show several empirical applications of the procedure, and explore robustness of the results. In Section III, we conclude. An Internet Appendix[4] contains technical details and Monte Carlo simulations.

I. Methodology

A. Model Setup

We start from a linear specification for the SDF,

$$m_t := \gamma_0^{-1} - \gamma_0^{-1}\lambda_v^\intercal v_t := \gamma_0^{-1}\left(1 - \lambda_g^\intercal g_t - \lambda_h^\intercal h_t\right), \tag{1}$$

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,

[4] The Internet Appendix is available on *The Journal of Finance* website.

both g_t and h_t are demeaned. Therefore, they are factor innovations satisfying $\mathrm{E}(g_t) = 0$ and $\mathrm{E}(h_t) = 0$. 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 objective of 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 0. Thus, this framework potentially includes both 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) and completely useless factors (factors that have zero SDF loadings and whose covariances with returns are uncorrelated with the covariances of returns with the SDF). Part of the procedure that we propose reduces the dimensionality of h_t, in an effort to eliminate the useless and redundant factors, obtaining a low-dimensional benchmark model.

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

$$\mathrm{E}(r_t) = \iota_n \gamma_0 + C_v \lambda_v = \iota_n \gamma_0 + C_g \lambda_g + C_h \lambda_h, \tag{2}$$

where ι_n is an $n \times 1$ vector of 1s, $C_a = \mathrm{Cov}(r_t, a_t)$, for $a = g, h$, or v. Furthermore, we assume that the dynamics of r_t follow a standard linear factor model,

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

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

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,

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

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 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 exposures and risk premia. Accordingly, in the empirical analysis, we 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, Ossola, and Scaillet (2016) and Kelly, Pruitt, and Su (2019) 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 explicitly write the projection of g_t on h_t as

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

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

$$C_g = \iota_n \xi^\intercal + C_h \chi^\intercal + C_e, \tag{6}$$

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

B. Challenges with Standard Two-Pass Methods in High-Dimensional Settings

Using two-pass regressions to estimate empirical asset pricing models dates back to Jensen, Black, and Scholes (1972) and Fama and MacBeth (1973). This approach is widely used in practice due in part to its simplicity. The procedure involves two steps, namely, an asset-by-asset time-series regression that yields estimates of the individual factor loadings βs, and a cross-sectional regression of expected returns on the estimated factor loadings that yields estimates of the 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, the method above should work smoothly, as pointed out by Cochrane (2009).

However, the empirical asset pricing literature has proposed hundreds of factors, which may include useless and redundant factors in addition to useful factors. All of the useful factors should be used as controls in estimating λ_g and testing its significance. But the number of potential factors p identified in the literature has increased to the same scale as, if not greater than, n or T. As a result, 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.

Standard methodologies therefore do not work well if at all in a high-dimensional setting due to the curse of dimensionality, which implies that dimension-reduction and regularization techniques are needed for valid inference. Existing literature employs 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. Selecting an incorrect model is problematic as it can lead to omitted variable bias when useful factors are not included or to efficiency loss when many useless or redundant factors are included.

C. Sparsity

The high-dimensionality issue is not unique to asset pricing. To address it, we need to impose a low-dimensional structure on the model. In this paper, as 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 that those factors alone are relevant for the estimation of λ_g. More specifically, sparsity in our setting means there are only s nonzero 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? It turns out that 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, and 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 nonzero 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 principal components as factors. Last but not least, as advocated in Friedman, Hastie, and Tibshirani (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 involve only a very small number of factors in absolute terms, say three or five. More nonzero coefficients can be identified given better conditions (e.g., larger sample size).

D. 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 for prediction purposes. With respect to parameter estimation, however, a well-documented bias is associated with the nonzero 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 find has more desirable statistical properties. The Post-LASSO estimator runs LASSO as a model selector and then refits the least-squares problem without penalty, using only those variables that have nonzero 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 nonzero 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 to conduct *inference* because, fundamentally, LASSO and other machine learning methods aim for better *prediction*. LASSO is designed to minimize out-of-sample prediction error. Certain variables, even if they are part of the true model, may be not worth including for prediction purposes because their contribution to prediction is too small relative to the cost of inclusion. Indeed, in any finite sample, we can never be sure that LASSO or Post-LASSO will select the correct model, just like we cannot claim that 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 the Internet Appendix.

The omitted variable bias due to model selection mistakes is exacerbated if the 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 therefore need to ensure that these factors are included in the set of controls *even if LASSO would suggest excluding them*. Note that this issue is not unique to high-dimensional problems (see, e.g., Leeb and Pötscher (2005)), but it is arguably more severe in such a scenario because model selection is inevitable.

E. Two-Pass Regression with Double-Selection LASSO

To guard against omitted variable bias due to selection mistakes, we adopt a DS strategy in the same spirit as Belloni, Chernozhukov, and Hansen (2014b) for estimating the treatment effects. 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 that are potentially missed in the first step but that, if omitted, would induce a large omitted variable bias. Factors excluded from both stages of the DS procedure must have small SDF loadings and covariances that correlate only mildly in the cross section with the covariance between the 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,[5]

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

where $\widehat{C}_h=\widehat{\mathrm{Cov}}(r_t,h_t)=T^{-1}\bar{R}\bar{H}^{\intercal}$.[6] This step selects among the factors in h_t those that best explain the cross section of expected returns. Denote by $\{\widehat{I}_1\}$ as the set of indices corresponding to the selected factors in this step.

(1.b) For each factor j in g_t (with $j=1,\cdots,d$), run a cross-sectional LASSO regression of $\widehat{C}_{g,\cdot,j}$ (the covariance between returns and the jth factor of g_t) on $\widehat{C}_h$ (the covariance between returns and all factors h_t),[7]

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

This step identifies the factors whose exposures are highly correlated with the exposures to g_t in the cross section. This is the crucial second step in the DS algorithm, which searches for factors that may be missed in 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 by $\{\widehat{I}_{2,j}\}$ the set of indices corresponding to the selected factors in the jth regression, where $\widehat{I}_2=\bigcup_{j=1}^{d}\widehat{I}_{2,j}$.

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

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

[7] For any matrix A, we use $A_{i,\cdot}$ and $A_{\cdot,j}$ to denote the ith row and jth column of A, respectively.

(2) Post-selection estimation
Run an OLS cross-sectional regression using covariances between the selected factors from *both* steps and returns,

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

We refer to this procedure as the DS approach, as opposed to the single-selection (SS) approach that involves only (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 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. (2018) for treatment-effect estimation, or by subset selection, partial least squares, and PCA regressions (or with LASSO selection on top of PCs similar to Kozak, Nagel, and Santosh (2018)). Chernozhukov et al. (2018) refer to this general procedure double machine learning. We advocate use of 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, Hastie, and Tibshirani (2009) (a so-called "greedy" algorithm). Their algorithm evaluates the contribution of each factor relative to a preselected best model through model comparison, and builds up the best model sequentially. Just as LASSO cannot deliver the true model with certainty, this algorithm cannot do so either because it commits to certain variables too early, which prevents the algorithm from finding the best overall solution later. Specifically, if one of the factors in the preselected model is redundant relative to the factor under consideration (i.e., the latter factor is in the DGP and the former factor is a noisy version of it), the latter factor could 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 reason, 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. We use this procedure as one of our robustness checks later.

Our LASSO regression contains nonnegative regularization parameters, for example, τ_j ($j = 0, 1, \ldots, d$), to control the level of the 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 at a time, so that this procedure involves two regularization parameters τ_0 and τ_1. To determine these parameters, we adopt the widely used CV procedure; see Friedman, Hastie, and Tibshirani (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) that assigns weights to λ_h that are 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.

F. Statistical Inference

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

THEOREM 1: *Under Assumptions 1 to 6 in Internet Appendix B, if* $s^2T^{1/2}(n^{-1} + T^{-1})\log(n \vee p \vee T) = o(1)$, *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

$$\Pi = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathrm{E}\big((1 - \lambda^{\intercal} v_t)(1 - \lambda^{\intercal} v_s)\Sigma_z^{-1} z_t z_s^{\intercal} \Sigma_z^{-1}\big), \quad \Sigma_z = \mathrm{Var}(z_t).$$

Note that 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 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 7 in the Internet Appendix holds. If* $qs^{3/2}(T^{-1/2} + n^{-1/2})\|V\|_{\mathrm{MAX}}\|Z\|_{\mathrm{MAX}} = o_p(1)$,[8] *we have*

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

[8] We use $\|A\|_{\mathrm{MAX}}$ to denote the $\mathbb{L}_\infty$-norm of A in the vector space.

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

$$\widehat{\Pi} = \frac{1}{T}\sum_{t=1}^{T}(1-\widehat{\lambda}^\intercal v_t)^2 \widehat{\Sigma}_z^{-1}\widehat{z}_t\widehat{z}_t^\intercal \widehat{\Sigma}_z^{-1}$$

$$+\frac{1}{T}\sum_{k=1}^{q}\sum_{t=k+1}^{T}\left(1-\frac{k}{q+1}\right)\left((1-\widehat{\lambda}^\intercal v_t)(1-\widehat{\lambda}^\intercal v_{t-k})\widehat{\Sigma}_z^{-1}\left(\widehat{z}_t\widehat{z}_{t-k}^\intercal + \widehat{z}_{t-k}\widehat{z}_t^\intercal\right)\widehat{\Sigma}_z^{-1}\right),$$

$$\widehat{\Sigma}_z = \frac{1}{T}\sum_{t=1}^{T}\widehat{z}_t\widehat{z}_t^\intercal, \quad \widehat{z}_t = g_t - \widetilde{\eta}_{\widetilde{I}}h_t, \quad \widetilde{\eta}_{\widetilde{I}} = \arg\min_{\eta}\left\{\|G-\eta H\|^2 : \eta_{\cdot,j}=0, \quad j\notin \widetilde{I}\right\},$$

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

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

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, Chernozhukov, and Hansen (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.

In Internet Appendix I, we provide an extensive set of simulations that demonstrate the finite-sample performance of our estimator.

II. Empirical Analysis

A. Data

A.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. These factors come from multiple sources. First, we download all workhorse factors in the U.S. equity market from Ken French's data library. We then add several published factors directly from the authors' websites, including liquidity from Pástor and Stambaugh (2003), the q-factors from Hou, Xue, and Zhang (2015), and the intermediary asset pricing factors from He, Kelly, and Manela (2017). 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, Xue, and Zhang (2017) and Green, Hand, and Zhang (2017).

To construct these factors, we include only stocks of 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, and then build and rebalance a long-short value-weighted portfolio (top 30% − bottom 30% or 1 − 0 dummy characteristic) each June for a 12-month holding period. Both Fama and French (2008) and Hou, Xue, and Zhang (2017) discuss the importance of using NYSE breakpoints and value-weighted portfolios. Microcaps, that is, 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 the Appendix, we report a complete list of the 150 factors and various descriptive statistics (publication year, ending year of the sample used in the paper, monthly average return, and annualized Sharpe ratio), as well as the corresponding references.

A.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 but 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 and 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 choices 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; see Litzenberger and Ramaswamy (1979). In line with the suggestion of Lewellen, Nagel, and Shanken (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.[9]

We add to these 36 portfolios 714 additional portfolios obtained from our factor zoo that cover additional characteristics. In particular, we try to include

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

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 include the resulting portfolios only if each of the six groups contains a sufficient number of stocks (at least 10). This procedure results in 119 sets of 3 × 2 bivariate-sorted portfolios, yielding 714 portfolios.[10]

As a robustness check, we alternatively use the set of 202 portfolios employed by 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 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 portfolios. The sorting procedure is the same as that for the 3×2 portfolios, except that the stock universe is divided into five groups for each characteristic.

B. Evaluating New Factors

In this section, we apply our methodology to factors proposed over the last five years (2012 to 2016), drawing the benchmark model against which to evaluate them from the set of 135 factors proposed before this recent period.[11] By placing ourselves in the position of researchers evaluating "new" factors (as of 2012), we demonstrate how our procedure can be applied going forward as more factors are proposed. In this exercise, all factors proposed in the 2012 to 2016 period are evaluated against the same benchmark, namely, the factors available up to 2012, so here we do not explicitly take into account the relative timing of introduction for the factors published after 2012 (we tackle in this the

[10] There are 16 factors for which bivariate-sorted portfolios are not available. Eight 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), research and development (R&D) increase (68), corporate investment (69), change in short-term investments (87), return on net operating assets (116), and return on assets (127).

[11] The most recent factors in our library were introduced in 2016.

next section). 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, further research may uncover better factors over time, yielding factors that improve over the existing ones.

B.1. The First LASSO

We start with the first step of our procedure: the cross-sectional LASSO, which is closely related to the dimension-reduction methods used in recent asset pricing research to tackle the factor zoo (e.g., Kozak, Nagel, and Santosh (2018)). The objective of this first LASSO is to select a parsimonious model that explains the cross section of expected returns.

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. We also show empirically its fragility in selecting the model.

When we apply it in our context, LASSO does indeed select 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 and thus yield the wrong model. It is useful to quantify 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 hence 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, the factors that get selected) do not change substantially as τ_0 varies.

A key question to address 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 the observation that in standard applications of machine learning, tuning parameters are typically chosen by simulating the performance of the algorithm in the data and then choosing the values for the parameters for which the estimator performs the best in those simulations. We use 10-fold 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 full 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 to 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 CV exercises using 200 different randomization seeds. For each seed, the CV chooses a different value of the tuning parameter τ_0. We then evaluate the robustness of the selected model using these 200 different values for τ_0. The range of possible values for τ_0 to consider in studying the robustness of LASSO is therefore determined by the possible (random) outcomes of the CV selection. This procedure effectively excludes values of τ_0 that are unlikely to be optimal using the CV criterion.

For each factor (identified by its ID), Figure 1 shows the fraction of the 200 LASSO-selected models in which the factor appears. The figure shows striking variability in the model selection step. Across 135 factors, only SMB is selected more than 70% of the time. Most of the factors are selected in only 1% to 20% of the cases.

If LASSO were able to perfectly select the true model, we should find that a small number of factors (say, three to five) are selected 100% of the time, while the remaining factors are 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 thus cautions against using simple LASSO to decide whether a factor should be included in the SDF. Broadly speaking, no machine learning methods can robustly determine the true model.

B.2. The Second LASSO

To make proper inference on the marginal contribution of new factors g_t, our procedure adds a second LASSO step that aims to identify the factors most likely to cause omitted variable bias. While the first LASSO depends only on h_t, this second LASSO depends on both g_t and h_t. This means that for each factor proposed after 2012, a different set of factors will be selected in the second step. For brevity, we do not report all of 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 a very parsimonious model, with four factors. The second-stage LASSO, in contrast, tends to select between 20 and 80 control factors. The striking difference is due to the difference in objective function across 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

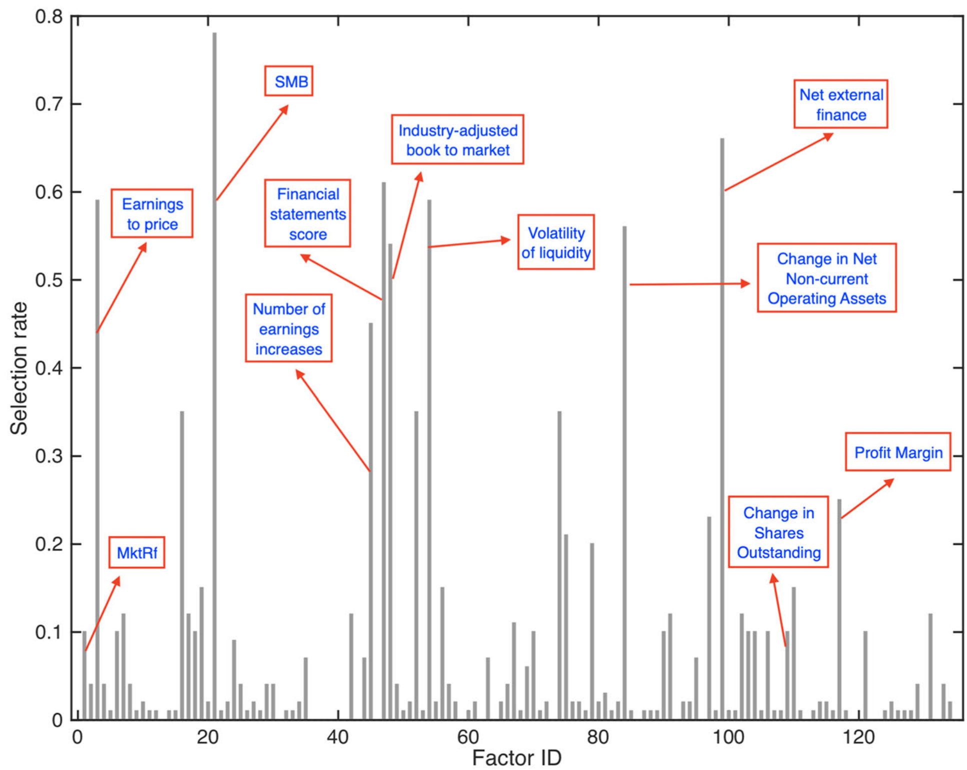

Figure 1. Subsamples: factor first selection rate. The figure shows the control factor selection rates for the tests of Table I (i.e., the factors selected by the first LASSO step of the double-selection procedure by cross-validation), across 200 random seeds shown in the heat maps in Figure 2 (corresponding to the 200 dots). In particular, for each factor identified by the factor ID (on the *X*-axis), the figure shows in what fraction of the 200 random seeds each factor is selected by cross-validation. (Color figure can be viewed at wileyonlinelibrary.com)

LASSO seeks to select the 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 factors.

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. Nonetheless, there is 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 large omitted variable bias if only the first-stage variables were used as controls.

B.3. The Double-Selection Estimator

We now present our results on the marginal contribution of each factor g_t using the DS methodology. Table I reports the results for the factors proposed over the last five years, among which we find Quality-Minus-Junk (QMJ), Betting-Against-Beta (BAB), two investment factors, (i.e., CMA from Fama and French (2015) (hereafter, FF) and IA from Hou, Xue, and Zhang (2015) (hereafter, HXZ)), two profitability factors, (i.e., RMW from FF and ROE from HXZ), the nontradable intermediary capital factor from He, Kelly, and Manela (2017), 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 or intuition in the paper that proposes the given factor, so that a positive estimate should be viewed as consistent with the economic implication. The t-statistic in each column corresponds to the test of the hypothesis that the slope is equal to 0, constructed using different methodologies across columns.

The first column reports our main results—the estimates of SDF loadings for the factors introduced since 2012, with corresponding t-statistics, obtained using 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 for both of the version of HXZ and the version of FF). HXZ's investment factor is also significant, as are intermediary investment and QMJ (interestingly, Gagliardini, Ossola, and Scaillet (2019) also find empirical evidence in favor of the recently introduced factors, such as investment and profitability, using a different econometric strategy.) All other factors appear to be 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 column reports the estimates that would obtain using the naive SS methodology, that is, using just one cross-sectional LASSO to select the factors to use as controls, without the second selection step that is useful to

Table I

Testing for Factors Introduced in the 2012 to 2016 Period

The table reports results of tests for the contribution of factors introduced in the 2012 to 2016 period, 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. The sample period is from July 1976 to December 2017. In columns (1) to (4), for ease of interpretation, we show the estimate of the SDF loading scaled to correspond to a unit beta exposure, λ_s, and the corresponding 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, which controls only 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. Significance levels: *(p < 0.10), **(p < 0.05), ***(p < 0.01).

		(1) DS		(2) SS		(3) FF3		(4) No Selection		(5) Avg. Ret.	
id	Factor Description	λ_s (bp)	tstat (DS)	λ_s (bp)	tstat (SS)	λ_s (bp)	tstat (OLS)	λ_s (bp)	tstat (OLS)	avg.ret. (bp)	tstat
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^{*}

avoid omitted variable bias due to mistakes in model selection. The results are quite different from those of 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 under the DS method do so when using SS. Given that our theoretical results and simulations show that the SS method is biased in finite samples, it should not be surprising that empirical results obtained using the SS method differ from those obtained using the DS method. The results in Table I thus show that omitted variable bias plays a major role empirically.

The third column shows instead what the estimates for the various factors would be if one instead used the Fama-French three factors (Market, SMB, HML) as controls, rather than selecting the controls optimally from among the various potential factors. The results differ noticeably from the benchmark with DS, with 9 out of 15 factors significant against the Fama-French three-factor model. Of course, if the *true* SDFs were known ex ante, selecting all of and only the true factors as controls would lead to the most efficient estimate for λ_g. In practice, however, we are not likely to be able to pin down the entire SDF with certainty. The aim of our DS procedure is to select the controls statistically—avoiding arbitrary choices of control factors—while at the same time minimizing potential omitted variable bias.

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

The last column of the table reports the average excess returns of the factors, that is, their risk premia. These results capture the compensation investors obtain from bearing exposure to a given factor, holding exposures to all other risk factors constant.[12] As discussed in, for example, 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

[12] We note that about half of these factors do not have a significant risk premium, while they typically did in the original publications. This is due in part to the different sample period used here, and in part to the fact that 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, Xue, and Zhang (2017).

(no matter how significant it is), but rather 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.

To summarize, Table I shows that the factors chosen as controls, and the econometric procedure used for estimation, have a large impact on conclusions drawn 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 contained in the existing factor zoo without introducing biases while at the same time accounting for efficiency losses.

C. Evaluating Factors Recursively

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

To illustrate this point, for each year starting in 1994, we use our DS procedure to test whether the factors introduced during the given year are useful or redundant relative to factors existing up to that year. Note that the exercise is fully recursive, using only the information available up to time t when evaluating a factor introduced in 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 II reports the results. In the table, the factors introduced since 1994 are identified by their ID. Those factors that appear to be statistically significant according to our test, relative to the factors introduced before them, are underlined. 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 each year starting in 1994, only 17 factors would have been considered useful, with a large majority identified as redundant or useless.

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

Our approach differs from those of existing literature in four substantial ways. First, and most important, we explicitly address the problem of omitted

Table II

Testing Factors Recursively by Year of Publication

The table reports results of a recursive factor-testing exercise over the period 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 the given year. Column (2) reports the number of controls available in each year t, that is, the number of potential controls in h_t based on factors published up to the given year. Column (3) shows for each year the IDs of the factors that were published with data up to the given year. We then test whether each new factor helps to explain asset prices relative to the factors published in previous years, using only the data up to the publication year t. We underline ID in column (3) each 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.

	(1)	(2)	(3)											
Year	# 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	$\underline{34}$											
1998	174	34	35	36	37	$\underline{38}$	39	40	$\underline{41}$	42	43	$\underline{44}$		
1999	228	44	45	46										
2000	234	46	47	48	49	$\underline{50}$	$\underline{51}$							
2001	252	51	52	$\underline{53}$	54	55	56	57	58					
2002	294	58	59	60	61									
2003	312	61	62	63	$\underline{64}$	65	$\underline{66}$							
2004	336	66	67	68	69	70	71	$\underline{72}$	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	$\underline{95}$	96	97	98	$\underline{99}$	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	$\underline{123}$	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	$\underline{140}$	141	142	143	144							
2015	738	144	$\underline{145}$	146	$\underline{147}$	$\underline{148}$								
2016	750	148	149	150										

variable bias due to potential model selection mistakes when making inference about factors' contribution to asset prices. Second, our method directly accounts for the correlations 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 econometric advances in model selection to reduce the dimensionality of the factor set. Fourth, the criterion that 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 II.B), as it is the right quantity to evaluate the contribution of a factor to explaining asset prices.

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

Finally, it is worth noting that this recursive exercise is simply meant to illustrate possible applications of our method. We do not address here 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.[13] However, our methodology is quite general, and does not require h_t and g_t to be ordered temporally. 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 present an alternative recursive application in the next section, and leave other applications to future research.

D. A Forward Stepwise Procedure

Rather than testing each factor against those that have already been introduced in the literature, here we propose an alternative recursive exercise in which factors in the benchmark are chosen by a stepwise procedure based on DS.[14]

Key to this exercise is the fact that when we choose the factors from h_t to form the benchmark to evaluate g_t, we can choose some factors that will not be penalized by LASSO and therefore that will be guaranteed to be selected in both stages. We start with a small set of "preselected" factors from h_t (in what follows, we start from the Fama-French four-factor model, which includes momentum). We then run our DS estimator on all other factors g_t, one at a time, evaluating each against a benchmark that includes the preselected factors plus any additional ones selected by LASSO. We select the factor g_t with the highest t-statistic, and impose the requirement that in all future iterations, this factor always be preselected from h_t. We then iterate this procedure recursively, so that the set of "preselected" factors expands by exactly one factor at each iteration.

[13] One 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. The results are similar—9 out of 12 factors significant in Table IA.I are also significant in Table II—though by construction, the results are not invariant to the ordering of factors.

[14] We are grateful to an anonymous referee for this suggestion.

This recursive exercise ends when no more factors are deemed to make a marginal contribution to the existing set. In our data, the additional factors in the last iteration are: 148, 88, 51, 62, 74, 61, 49, 122, 6, 55, 72, 53, 119, 140, 44, 147, 65, 32, 31, 87, 123, and 5 in order of their selection (the identities of the factors are reported in the Appendix). There is an interesting overlap between the factors that appear significant in this case (where the relative timing of introduction is ignored) and those selected in the historical exercise in the previous section: about half of the significant factors among those tested in both exercises are the same (the recursive exercise only tests factors introduced after 1994). This suggests that several factors (e.g., Betting Against Beta, HXZ investment and profitability) make an important contribution relative to not only the factors introduced beforehand, but also all remaining factors. Of course, the fact that other factors do not overlap should not be surprising—it is just another indication that the choice of benchmark plays a major role in determining which factors are significant and which are not. Overall, both exercises lead to a substantial reduction in the total number of factors in the zoo.

While the historical exercise in Section II.C mimics the discovery process over time, the stepwise procedure in this section illustrates the conclusions reached by researchers with different *priors* on the correct benchmark model. At each iteration, a subset of the factors h_t is guaranteed to be selected: this captures the case in which the researcher has a strong prior view that those factors should be in the benchmark, whereas other factors will enter the benchmark only if they are successful in explaining the cross section of returns (i.e., they are selected by LASSO). As new evidence comes in about factors that appear to be useful, researchers update the set of "preselected" factors.

We conclude with a caveat. The recursive procedures described both in this section and in the previous one do not guarantee to perfectly identify the full asset pricing model. As we emphasize in this paper, the oracle property of LASSO and other model selection methods fails in finite sample, and hence the list of selected factors will be prone to model selection error. Instead, these exercises should be viewed as mimicking possible situations in which researchers interested in testing *one* additional factor g_t at a time may want to apply our DS methodology.

E. Robustness

In this section, we explore the robustness of our estimator, and we discuss some extensions of our setup. The most important robustness test, which we present first, is with respect to the tuning parameters, especially since we show in Section II.B.1 that the first step of our procedure (LASSO model selection) is *not* very robust to these changes.

E.1. Robustness to the Choice of Tuning Parameters

In this section, we explore 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 employ our benchmark estimates in Table I and check the robustness of inference about the marginal contribution of the factors proposed after 2012.

Just as in Section II.B.1, we need to determine a reasonable range of values for the two tuning parameters. We follow the procedure described before: we first choose 200 different seeds for the CV simulations, where for every set of simulations, we obtain one estimate for the two tuning parameters, and 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 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 CV criterion for any of the 200 simulations.

We display 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 dots, each of which corresponds to a choice of tuning parameters in one of the CV simulation sets. The cross in each graph is the average of these 200 tuning parameters. This is the level we use to generate the baseline results (Table I). The figure shows that inference for some factors is more robust than for others. The factors that appear significant in the baseline generally appear robust in the sense that the vast majority of choices for the tuning parameters yield statistically significant results. Some of them (e.g., investment and profitability) appear highly robust. Others, such as the intermediary investment factor, do not appear very robust, in the sense that for a nontrivial subset of the tuning parameters considered, their significance vanishes. Other factors appear strong and robust, though not statistically significant at conventional levels in our main results (e.g., the Betting Against Beta factor). Finally, most other factors (e.g., 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, notably, that a few of the recent factors appear to contribute significantly to explaining the cross section, while most of the remaining factors are redundant or useless, but they provide a more nuanced view of the contribution of some of the factors.

Figure 3 displays 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 zero to five factors and as large as 120 factors. The range of tuning parameters that 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 I, highlighting the most robust factors. We therefore

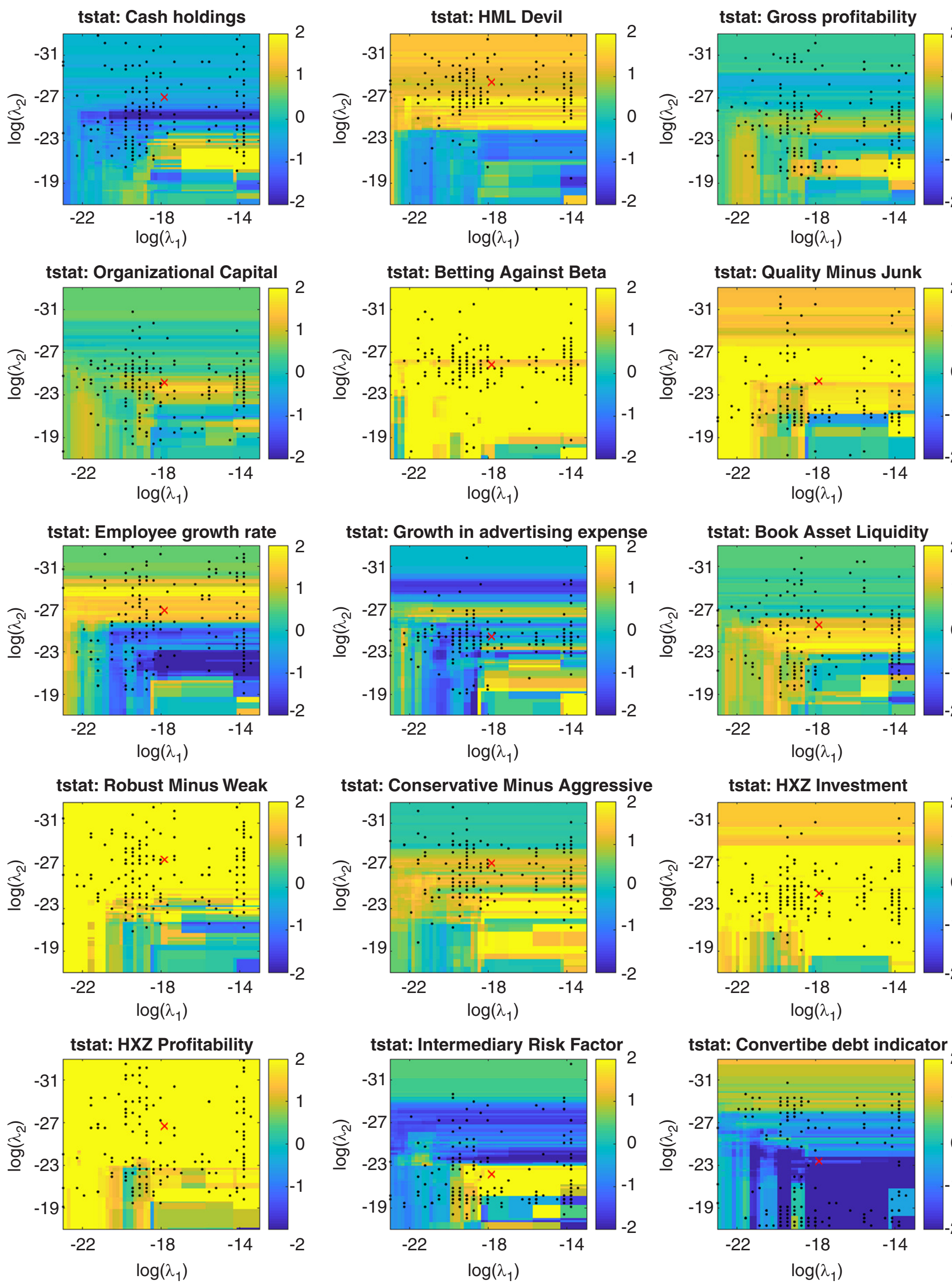

Figure 2. Factors introduced in the 2012 to 2016 period: robustness to tuning parameters (*t*-statistics). The figure provides heat maps for double-selection tests of factors introduced in the 2012 to 2016 period, as in the first column of Table I, using a wide range of tuning parameters, for the first LASSO stage on the X-axis and for the second LASSO stage on the Y-axis. The heat maps display t-statistics for each factor in different models. Dots are the result of 200 time-series cross-validation estimations of the tuning parameter. The red "×" is the average of the 200 dots, which corresponds to the model used in Table I. (Color figure can be viewed at wileyonlinelibrary.com)

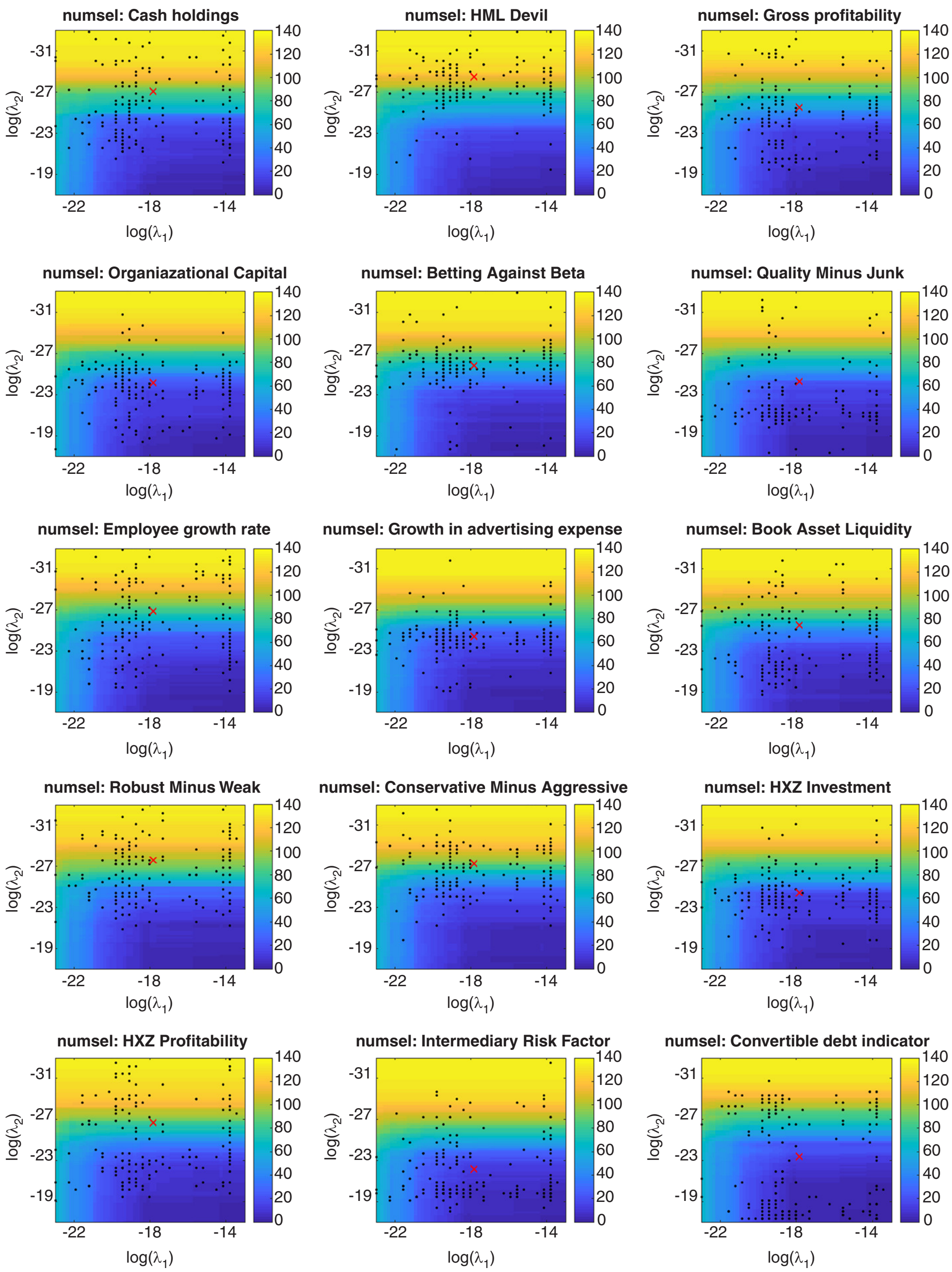

Figure 3. Factors introduced in the 2012 to 2016 period: robustness to tuning parameters (# selected controls). The figure provides heat maps for double-selection tests of factors introduced in the 2012 to 2016 period, as in the first column of Table I, using a wide range of tuning parameters, for the first LASSO stage on the X-axis and for the second LASSO stage on the Y-axis. The heat maps display numbers of controls selected for each factor. Dots are the result of 200 time-series cross-validation estimations of the tuning parameter. The red "×" is the average of the 200 dots, which corresponds to the model used in Table I. (Color figure can be viewed at wileyonlinelibrary.com)

recommend the use of heat maps like these to evaluate the robustness of significant discoveries by model selection procedures such as those here.

E.2. Robustness to Test Assets and Regularization Method

In this section, we 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 I).

Column (1) of Table III replicates our baseline results for convenience (as in the first column of Table I). Column (2) shows that the results are similar when sorting the test assets into 5×5 instead of 3×2 portfolios. In column (3), our results continue to hold when using a smaller number of test assets, namely, the 202 portfolios used in Giglio and Xiu (2016) and described in Section II.A.2.

Columns (4) and (5) show that our results also hold when using different dimension-reduction procedures. Which method is preferred in a given 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, which combines a penalty from LASSO with that of the Ridge regression, is a reasonable alternative to explore in this context. The model selected by the Elastic Net is naturally larger, but, as column (4) shows, the results are consistent with our benchmark based on pure LASSO. An alternative, following Kozak, Nagel, and Santosh (2020), is to first construct PCA of the factors, and then use LASSO on the principal components. The results using this approach, reported in column (5), are statistically weaker but broadly in line with those of the benchmark specification.[15]

Finally, following our discussion in Section I.E, column (6) chooses the benchmark models for each factor using the forward stepwise regression approach suggested by Harvey and Liu (2016), instead of LASSO in both (1.a) and (1.b) of our DS procedure. This approach starts with a prior model (e.g., Fama-French four-factor model) and then continuously adds factors, one at a time, to the first- and second-stage models, until no more factors lead to improvement according to the bayesian information criterion (BIC). Again, the results are similar to those of our baseline specification.

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

[15] We should note that the standard errors and the test we build are derived for the case of LASSO. In light of Chernozhukov et al. (2018), we expect the same formulas to work for other machine learning methods, such as LASSO on principal components, despite the lack of theory (they do perform well in simulations). Nonetheless, it is interesting to see that the conclusions are broadly similar.

Table III

Robustness for Factors Introduced in the 2012 to 2016 Period

The table reports robustness tests for the estimates of SDF loadings for factors introduced in the 2012 to 2016 period, relative to the set of 135 factors introduced up to 2011. The first column reports the results from the first column of Table I 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 fifth 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. In the last column, we use the forward stepwise regression in place of LASSO in the double-selection procedure. Significance levels: $^{*}(p < 0.10)$, $^{**}(p < 0.05)$, $^{***}(p < 0.01)$.

		(1) Bivariate 3 × 2		(2) Bivariate 5 × 5		(3) 202 Portfolios		(4) Elastic Net		(5) PCA		(6) Stepwise	
id	Factor Description	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)	λ_s (bp)	tstat (DS)
136	Cash holdings	−34	−0.42	34	0.40	131	0.89	−13	−0.14	−65	−0.62	−73	−0.87
137	HML Devil	54	1.04	15	0.29	56	0.57	62	1.23	−27	−0.51	49	1.01
138	Gross profitability	20	0.48	28	0.66	88	1.42	−11	−0.26	16	0.35	16	0.47
139	Organizational Capital	28	0.92	23	0.75	6	0.16	12	0.38	21	0.57	0	0.01
140	Betting Against Beta	35	1.45	43	1.94^{*}	31	1.03	28	1.12	59	2.56^{***}	62	2.57^{***}
141	Quality Minus Junk	73	2.03^{**}	58	1.67	123	2.45^{**}	74	2.13^{**}	71	1.89^{*}	40	1.16
142	Employee growth	43	1.36	12	0.34	54	1.34	51	1.49	−4	−0.09	33	0.98
143	Growth in advertising	−12	−1.18	6	0.57	17	1.30	9	0.74	−6	−0.57	3	0.27
144	Book Asset Liquidity	40	1.07	−24	−0.61	37	0.77	26	0.68	24	0.63	33	1.00
145	RMW	160	4.45^{***}	104	3.13^{***}	112	1.98^{**}	125	3.43^{***}	88	2.11^{**}	96	2.71^{***}
146	CMA	38	1.10	19	0.59	33	0.52	32	0.85	18	0.44	23	0.67
147	HXZ IA	51	2.11^{**}	44	1.87^{*}	−45	−1.42	69	2.77^{***}	36	1.31	49	1.92^{*}
148	HXZ ROE	77	3.37^{***}	72	2.62^{***}	116	2.22^{***}	103	3.85^{***}	41	1.46	101	3.87^{***}
149	Intermediary Risk Factor	112	2.21^{**}	38	0.73	−16	−0.33	−16	−0.33	103	1.92^{*}	−10	−0.17
150	Convertible debt	−15	−1.36	−6	−0.56	68	5.13^{***}	−12	−1.08	−9	−0.88	0	−0.02

III. Conclusion

In this paper, we propose a regularized two-pass cross-sectional regression approach to establish the asset pricing contribution 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 DS procedure of Belloni, Chernozhukov, and Hansen (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 our methodology to a large set of factors proposed in the literature over the last 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 factors proposed prior 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 finding illustrates how the two-step procedure is able to produce correct inference by overcoming the model selection mistakes that necessarily arise when applying statistical selection methods. Third, we show that 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 a control (as opposed to the model selection procedure we advocate).

Taken together, our results are quite encouraging about the ongoing 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 bring discipline to the "zoo of factors."

Initial submission: September 6, 2017; Accepted: June 20, 2019
Editors: Stefan Nagel, Philip Bond, Amit Seru, and Wei Xiong

Appendix: Factor Zoo

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, Hand, and Zhang (2017) and Hou, Xue, and Zhang (2017). For each factor, the table lists the factor publication year, the end year of the sample in the original paper, the monthly average return, the annualized Sharpe ratio, and the corresponding references.

ID	Description	Year.pub	Year.end	Avg.Ret.	Ann.SR	Reference
1	Excess Market Return	1972	1965	0.64%	50.6%	Jensen, Black, and Scholes (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, Jones, and Latane (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)

(Continued)

Appendix: Factor Zoo— *Continued*

ID	Description	Year.pub	Year.end	Avg.Ret.	Ann.SR	Reference
25	36-month momentum	1993	1989	0.09%	13.4%	Jegadeesh and Titman (1993)
26	Sales growth	1994	1990	0.04%	5.8%	Lakonishok, Shleifer, and Vishny (1994)
27	Cash flow-to-price	1994	1990	0.31%	32.5%	Lakonishok, Shleifer, and Vishny (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, Thaler, and Womack (1995)
30	Dividend omission	1995	1988	−0.18%	−18.0%	Michaely, Thaler, and Womack (1995)
31	Working capital accruals	1996	1991	0.22%	46.0%	Sloan (1996)
32	Sales to price	1996	1991	0.35%	41.8%	Barbee Jr, Mukherji, and Raines (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, Naik, and Radcliffe (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, Elliott, and Finn (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, Porter, and Stevens (2000)
49	Industry-adjusted cash flow to price ratio	2000	1998	0.26%	52.1%	Asness, Porter, and Stevens (2000)
50	Industry-adjusted change in employees	2000	1998	−0.01%	−1.5%	Asness, Porter, and Stevens (2000)
51	Industry-adjusted size	2000	1998	0.36%	36.3%	Asness, Porter, and Stevens (2000)
52	Dollar trading volume	2001	1995	0.38%	35.8%	Chordia, Subrahmanyam, and Anshuman (2001)
53	Volatility of liquidity (dollar trading volume)	2001	1995	0.20%	38.8%	Chordia, Subrahmanyam, and Anshuman (2001)
54	Volatility of liquidity (share turnover)	2001	1995	0.02%	2.1%	Chordia, Subrahmanyam, and Anshuman (2001)

(*Continued*)

Appendix: Factor Zoo— *Continued*

ID	Description	Year.pub	Year.end	Avg.Ret.	Ann.SR	Reference
55	Advertising Expense-to-market	2001	1995	−0.13%	−15.6%	Chan, Lakonishok, and Sougiannis (2001)
56	R&D Expense-to-market	2001	1995	0.34%	36.2%	Chan, Lakonishok, and Sougiannis (2001)
57	R&D-to-sales	2001	1995	0.06%	5.5%	Chan, Lakonishok, and Sougiannis (2001)
58	Kaplan-Zingales Index	2001	1997	0.22%	25.3%	Lamont, Polk, and Saaá-Requejo (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 (2003)
63	Idiosyncratic return volatility	2003	1997	0.07%	5.1%	Ali, Hwang, and Trombley (2003)
64	Growth in long term net operating assets	2003	1993	0.22%	51.8%	Fairfield, Whisenant, and Yohn (2003)
65	Order backlog	2003	1999	0.05%	5.7%	Rajgopal, Shevlin, and Venkatachalam (2003)
66	Changes in Long-term Net Operating Assets	2003	1993	0.24%	56.0%	Fairfield, Whisenant, and Yohn (2003)
67	Cash flow to price ratio	2004	1997	0.27%	31.7%	Desai, Rajgopal, and Venkatachalam (2004)
68	R&D increase	2004	2001	0.06%	11.1%	Eberhart, Maxwell, and Siddique (2004)
69	Corporate investment	2004	1995	0.13%	36.4%	Titman, Wei, and Xie (2004)
70	Earnings volatility	2004	2001	0.10%	10.7%	Francis et al. (2004)
71	Abnormal Corporate Investment	2004	1995	0.13%	31.2%	Titman, Wei, and Xie (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, Lee, and Zhang (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 Noncash Working Capital	2005	2001	0.11%	25.2%	Richardson et al. (2005)
82	Change in Noncurrent Operating Assets	2005	2001	0.21%	44.5%	Richardson et al. (2005)
83	Change in Noncurrent Operating Liabilities	2005	2001	0.04%	9.6%	Richardson et al. (2005)
84	Change in Net Noncurrent Operating Assets	2005	2001	0.23%	35.4%	Richardson et al. (2005)

(*Continued*)

Appendix: Factor Zoo— *Continued*

ID	Description	Year.pub	Year.end	Avg.Ret.	Ann.SR	Reference
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, Richardson, and Sloan (2006)
98	Net debt finance	2006	2000	0.17%	48.3%	Bradshaw, Richardson, and Sloan (2006)
99	Net external finance	2006	2000	0.22%	38.6%	Bradshaw, Richardson, and Sloan (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, Richardson, and Tuna (2007)
108	Enterprise book-to-price	2007	2001	0.14%	14.7%	Penman, Richardson, and Tuna (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, Livnat, and Mendenhall (2008)
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, Sun, and Zhang (2008)
114	Investment Growth	2008	2003	0.17%	39.5%	Xing (2008)
115	Composite Debt Issuance	2008	2005	0.08%	21.6%	Lyandres, Sun, and Zhang (2008)

(Continued)

Appendix: Factor Zoo— *Continued*

ID	Description	Year.pub	Year.end	Avg.Ret.	Ann.SR	Reference
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, Huang, and Wirjanto (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, Bartov, and Faurel (2010)
128	Accrual volatility	2010	2008	0.19%	26.6%	Bandyopadhyay, Huang, and Wirjanto (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, Cakici, and Whitelaw (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, Frazzini, and Pedersen (2019)
142	Employee growth rate	2014	2010	0.08%	12.9%	Belo, Lin, and Bazdresch (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, Xue, and Zhang (2015)
148	HXZ Profitability	2015	2012	0.57%	77.5%	Hou, Xue, and Zhang (2015)
149	Intermediary Investment	2016	2012			He, Kelly, and Manela (2017)
150	Convertible debt indicator	2016	2012	0.11%	26.4%	Valta (2016)

REFERENCES

Abarbanell, Jeffery S., and Brian J. Bushee, 1998, Abnormal returns to a fundamental analysis strategy, *The Accounting Review* 73, 19–45.

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

Ali, Ashiq, Lee-Seok Hwang, and Mark A. Trombley, 2003, Arbitrage risk and the book-to-market anomaly, *Journal of Financial Economics* 69, 355–373.

Almeida, Heitor, and Murillo Campello, 2007, Financial constraints, asset tangibility, and corporate investment, *Review of Financial Studies* 20, 1429–1460.

Amihud, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Amihud, Yakov, and Haim Mendelson, 1989, The effects of beta, bid-ask spread, residual risk, and size on stock returns, *Journal of Finance* 44, 479–486.

Anderson, Christopher W., and Luis Garcia-Feijoo, 2006, Empirical evidence on capital investment, growth options, and security returns, *Journal of Finance* 61, 171–194.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *Journal of Finance* 61, 259–299.

Asness, Clifford, and Andrea Frazzini, 2013, The devil in hml's details, *The Journal of Portfolio Management* 39, 49–68.

Asness, Clifford S., Andrea Frazzini, and Lasse Heje Pedersen, 2019, Quality minus junk, *Review of Accounting Studies* 24, 34–112.

Asness, Clifford S., R. Burt Porter, and Ross L. Stevens, 2000, Predicting stock returns using industry-relative firm characteristics, Technical report, AQR Capital Investment.

Bai, Jushan, and Guofu Zhou, 2015, Fama-macbeth two-pass regressions: Improving risk premia estimates, *Finance Research Letters* 15, 31–40.

Balakrishnan, Karthik, Eli Bartov, and Lucile Faurel, 2010, Post loss/profit announcement drift, *Journal of Accounting and Economics* 50, 20–41.

Bali, Turan G., Nusret Cakici, and Robert F. Whitelaw, 2011, Maxing out: Stocks as lotteries and the cross-section of expected returns, *Journal of Financial Economics* 99, 427–446.

Bandyopadhyay, Sati P., Alan G. Huang, and Tony S. Wirjanto, 2010, The accrual volatility anomaly, Technical report, Working paper, University of Waterloo.

Barbee Jr, William C., Sandip Mukherji, and Gary A. Raines, 1996, Do sales-price and debt-equity explain stock returns better than book-market and firm size? *Financial Analysts Journal* 52, 56–60.

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

Barth, Mary E., John A. Elliott, and Mark W. Finn, 1999, Market rewards associated with patterns of increasing earnings, *Journal of Accounting Research* 37, 387–413.

Basu, Sanjoy, 1977, Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis, *Journal of Finance* 32, 663–682.

Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen, 2012, Sparse models and methods for optimal instruments with an application to eminent domain, *Econometrica* 80, 2369–2429.

Belloni, Alexandre, and Victor Chernozhukov, 2013, Least squares after model selection in high-dimensional sparse models, *Bernoulli* 19, 521–547.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, 2014a, High-dimensional methods and inference on structural and treatment effects, *Journal of Economic Perspectives* 28, 29–50.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, 2014b, Inference on treatment effects after selection among high-dimensional controls, *The Review of Economic Studies* 81, 608–650.

Belo, Frederico, and Xiaoji Lin, 2011, The inventory growth spread, *Review of Financial Studies* 25, 278–313.

Belo, Frederico, Xiaoji Lin, and Santiago Bazdresch, 2014, Labor hiring, investment, and stock return predictability in the cross section, *Journal of Political Economy* 122, 129–177.

Bhandari, Laxmi Chand, 1988, Debt/equity ratio and expected common stock returns: Empirical evidence, *Journal of Finance* 43, 507–528.

Bickel, Peter J., Ya'acov Ritov, and Alexandre B. Tsybakov, 2009, Simultaneous analysis of Lasso and Dantzig selector, *The Annals of Statistics* 37, 1705–1732.

Bondt, Werner F.M., and Richard Thaler, 1985, Does the stock market overreact? *Journal of Finance* 40, 793–805.

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R Roberts, 2007, On the importance of measuring payout yield: Implications for empirical asset pricing, *Journal of Finance* 62, 877–915.

Bradshaw, Mark T., Scott A. Richardson, and Richard G. Sloan, 2006, The relation between corporate financing activities, analysts' forecasts and stock returns, *Journal of Accounting and Economics* 42, 53–85.

Brandt, Michael W., Runeet Kishore, Pedro Santa-Clara, and Mohan Venkatachalam, 2008, Earnings announcements are full of surprises, Technical report, Duke University.

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

Brown, David P., and Bradford Rowe, 2007, The productivity premium in equity returns, Technical report, University of Wisconsin-Madison.

Bryzgalova, Svetlana, 2015, Spurious factors in linear asset pricing models, Technical report, Stanford University.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chan, Louis KC, Josef Lakonishok, and Theodore Sougiannis, 2001, The stock market valuation of research and development expenditures, *Journal of Finance* 56, 2431–2456.

Chandrashekar, Satyajit, and Ramesh KS Rao, 2009, The productivity of corporate cash holdings and the cross-section of expected stock returns, Technical report, University of Texas at Austin.

Chen, Long, and Lu Zhang, 2010, A better three-factor model that explains more anomalies, *Journal of Finance* 65, 563–595.

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

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey, 2018, Double/debiased machine learning for treatment and structural parameters, *The Econometrics Journal* 21, 1–68.

Chernozhukov, Victor, Christian Hansen, and Martin Spindler, 2015, Valid post-selection and post-regularization inference: An elementary, general approach, *Annual Review of Economics* 7, 649–688.

Chordia, Tarun, Avanidhar Subrahmanyam, and V. Ravi Anshuman, 2001, Trading activity and expected stock returns, *Journal of Financial Economics* 59, 3–32.

Cochrane, John H., 2009, *Asset Pricing (Revised Edition)* (Princeton University Press, Princeton, NJ).

Cochrane, John H., 2011, Presidential address: Discount rates, *Journal of Finance* 66, 1047–1108.

Connor, Gregory, Matthias Hagmann, and Oliver Linton, 2012, Efficient semiparametric estimation of the Fama–French model and extensions, *Econometrica* 80, 713–754.

Daniel, Kent, and Sheridan Titman, 2006, Market reactions to tangible and intangible information, *Journal of Finance* 61, 1605–1643.

Datar, Vinay T., Narayan Y. Naik, and Robert Radcliffe, 1998, Liquidity and stock returns: An alternative test, *Journal of Financial Markets* 1, 203–219.

David McLean, R., and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability? *Journal of Finance* 71, 5–32.

Desai, Hemang, Shivaram Rajgopal, and Mohan Venkatachalam, 2004, Value-glamour and accruals mispricing: One anomaly or two? *The Accounting Review* 79, 355–385.

Dichev, Ilia D., 1998, Is the risk of bankruptcy a systematic risk? *Journal of Finance* 53, 1131–1147.

Eberhart, Allan C., William F. Maxwell, and Akhtar R. Siddique, 2004, An examination of long-term abnormal stock returns and operating performance following R&D increases, *Journal of Finance* 59, 623–650.

Eisfeldt, Andrea L., and Dimitris Papanikolaou, 2013, Organization capital and the cross-section of expected returns, *Journal of Finance* 68, 1365–1406.

Fairfield, Patricia M., Scott Whisenant, and Teri Lombardi Yohn, 2003, The differential persistence of accruals and cash flows for future operating income versus future profitability, *Review of Accounting Studies* 8, 221–243.

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

Fama, Eugene F., and Kenneth R. French, 2008, Dissecting anomalies, *Journal of Finance* 63, 1653–1678.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F., and Kenneth R. French, 2018, Choosing factors, *Journal of Financial Economics* 128, 234–252.

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Ferson, Wayne E., and Campbell R. Harvey, 1991, The variation of economic risk premiums, *Journal of Political Economy* 99, 385–415.

Francis, Jennifer, Ryan LaFond, Per M. Olsson, and Katherine Schipper, 2004, Costs of equity and earnings attributes, *The Accounting Review* 79, 967–1010.

Frazzini, Andrea, and Lasse H. Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1–25.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber, 2020, Dissecting characteristics nonparametrically, *Review of Financial Studies* , in press.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani, 2009, *The Elements of Statistical Learning - Data Mining, Inference, and Prediction*, second edition (Springer-Verlag, New York, NY).

Gagliardini, Patrick, Elisa Ossola, and Olivier Scaillet, 2016, Time-varying risk premium in large cross-sectional equity data sets, *Econometrica* 84, 985–1046.

Gagliardini, Patrick, Elisa Ossola, and Olivier Scaillet, 2019, A diagnostic criterion for approximate factor structure, *Journal of Econometrics* 212, 503–521.

Gettleman, Eric, and Joseph M. Marks, 2006, Acceleration strategies, Technical report, Bentley University.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.

Giglio, Stefano, Yuan Liao, and Dacheng Xiu, 2018, Thousands of alpha tests, Technical report, University of Chicago.

Giglio, Stefano W., and Dacheng Xiu, 2016, Asset pricing with omitted factors, Technical report, University of Chicago.

Green, Jeremiah, John R. M. Hand, and X. Frank Zhang, 2017, The characteristics that provide independent information about average U.S. monthly stock returns, *Review of Financial Studies* 30, 4389–4436.

Guoming Huang, Alan, 2009, The cross section of cashflow volatility and expected stock returns, *Journal of Empirical Finance* 16, 409–429.

Hafzalla, Nader, Russell Lundholm, and E. Matthew Van Winkle, 2011, Percent accruals, *The Accounting Review* 86, 209–236.

Harvey, Campbell R., and Yan Liu, 2016, Lucky factors, Technical report, Duke University.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2015, . . . and the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

Haugen, Robert A., and Nardin L. Baker, 1996, Commonality in the determinants of expected stock returns, *Journal of Financial Economics* 41, 401–439.

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

Hirshleifer, David, Kewei Hou, Siew Hong Teoh, and Yinglei Zhang, 2004, Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics* 38, 297–331.

Holthausen, Robert W., and David F. Larcker, 1992, The prediction of stock returns using financial statement information, *Journal of Accounting and Economics* 15, 373–411.

Hong, Harrison, and Marcin Kacperczyk, 2009, The price of sin: The effects of social norms on markets, *Journal of Financial Economics* 93, 15–36.

Hou, Kewei, and Tobias J. Moskowitz, 2005, Market frictions, price delay, and the cross-section of expected returns, *Review of Financial Studies* 18, 981–1020.

Hou, Kewei, and David T. Robinson, 2006, Industry concentration and average stock returns, *Journal of Finance* 61, 1927–1956.

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

Hou, Kewei, Chen Xue, and Lu Zhang, 2017, Replicating anomalies, Technical report, National Bureau of Economic Research.

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

Jegadeesh, Narasimhan, and Joshua Livnat, 2006, Revenue surprises and stock returns, *Journal of Accounting and Economics* 41, 147–171.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Jensen, Michael C., Fischer Black, and Myron S. Scholes, 1972, The capital asset pricing model: Some empirical tests, in M. C. Jensen, ed.: *Studies in the Theory of Capital Markets* (New York: Praeger).

Jiang, Guohua, Charles M.C. Lee, and Yi Zhang, 2005, Information uncertainty and expected returns, *Review of Accounting Studies* 10, 185–221.

Kama, Itay, 2009, On the market reaction to revenue and earnings surprises, *Journal of Business Finance & Accounting* 36, 31–50.

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

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2020, Shrinking the cross section, *Journal of Financial Economics* 135, 271–292.

Lakonishok, Josef, Andrei Shleifer, and Robert W Vishny, 1994, Contrarian investment, extrapolation, and risk, *Journal of Finance* 49, 1541–1578.

L Heston, Steven, and Ronnie Sadka, 2008, Seasonality in the cross-section of stock returns, *Journal of Financial Economics* 87, 418–445.

Lamont, Owen, Christopher Polk, and Jesús Saaá-Requejo, 2001, Financial constraints and stock returns, *Review of Financial Studies* 14, 529–554.

Leeb, Hannes, and Benedikt M. Pötscher, 2005, Model selection and inference: Facts and fiction, *Econometric Theory* 21, 21–59.

Lerman, Alina, Joshua Livnat, and Richard R. Mendenhall, 2008, The high-volume return premium and post-earnings announcement drift, Technical report, Yale University.

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

Lev, Baruch, and Doron Nissim, 2004, Taxable income, future earnings, and equity values, *The Accounting Review* 79, 1039–1074.

Lewellen, Jonathan, Stefan Nagel, and Jay Shanken, 2010, A skeptical appraisal of asset pricing tests, *Journal of Financial Economics* 96, 175–194.

Litzenberger, Robert H., and Krishna Ramaswamy, 1979, The effects of personal taxes and dividends on capital asset prices: Theory and empirical evidence, *Journal of Financial Economics* 7, 163–195.

Liu, Weimin, 2006, A liquidity-augmented capital asset pricing model, *Journal of Financial Economics* 82, 631–671.

Lo, Andrew W., and A. Craig MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies* 3, 431–467.

Lou, Dong, 2014, Attracting investor attention through advertising, *Review of Financial Studies* 27, 1797–1829.

Loughran, Tim, and Jay R. Ritter, 1995, The new issues puzzle, *Journal of Finance* 50, 23–51.

Loughran, Tim, and Jay W Wellman, 2011, New evidence on the relation between the enterprise multiple and average stock returns, *Journal of Financial and Quantitative Analysis* 46, 1629–1650.

Lyandres, Evgeny, Le Sun, and Lu Zhang, 2008, The new issues puzzle: Testing the investment-based explanation, *Review of Financial Studies* 21, 2825–2855.

Meinshausen, Nicolai, and Bin Yu, 2009, Lasso-type recovery of sparse representations for high-dimensional data, *The Annals of Statistics* 37, 246–270.

Michaely, Roni, Richard H. Thaler, and Kent L. Womack, 1995, Price reactions to dividend initiations and omissions: Overreaction or drift? *Journal of Finance* 50, 573–608.

Miller, Merton H., and Myron S. Scholes, 1982, Dividends and taxes: Some empirical evidence, *Journal of Political Economy* 90, 1118–1141.

Mohanram, Partha S., 2005, Separating winners from losers among lowbook-to-market stocks using financial statement analysis, *Review of Accounting Studies* 10, 133–170.

Moskowitz, Tobias J., and Mark Grinblatt, 1999, Do industries explain momentum? *Journal of Finance* 54, 1249–1290.

Novy-Marx, Robert, 2011, Operating leverage, *Review of Finance* 15, 103–134.

Novy-Marx, Robert, 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108, 1–28.

Ortiz-Molina, Hernán, and Gordon M. Phillips, 2014, Real asset illiquidity and the cost of capital, *Journal of Financial and Quantitative Analysis* 49, 1–32.

Ou, Jane A., and Stephen H. Penman, 1989, Financial statement analysis and the prediction of stock returns, *Journal of Accounting and Economics* 11, 295–329.

Palazzo, Berardino, 2012, Cash holdings, risk, and expected returns, *Journal of Financial Economics* 104, 162–185.

Pástor, L'uboš, and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

Penman, Stephen H., Scott A. Richardson, and Irem Tuna, 2007, The book-to-price effect in stock returns: Accounting for leverage, *Journal of Accounting Research* 45, 427–467.

Piotroski, Joseph D., 2000, Value investing: The use of historical financial statement information to separate winners from losers, *Journal of Accounting Research* 38, 1–52.

Pontiff, Jeffrey, and Artemiza Woodgate, 2008, Share issuance and cross-sectional returns, *Journal of Finance* 63, 921–945.

Rajgopal, Shivaram, Terry Shevlin, and Mohan Venkatachalam, 2003, Does the stock market fully appreciate the implications of leading indicators for future earnings? Evidence from order backlog, *Review of Accounting Studies* 8, 461–492.

Raponi, Valentina, Cesare Robotti, and Paolo Zaffaroni, 2017, Testing beta-pricing models using large cross-sections, Technical report, Imperial College London.

Rendleman, Richard J., Charles P. Jones, and Henry A. Latane, 1982, Empirical anomalies based on unexpected earnings and the importance of risk adjustments, *Journal of Financial Economics* 10, 269–287.

Richardson, Scott A., Richard G. Sloan, Mark T. Soliman, and Irem Tuna, 2005, Accrual reliability, earnings persistence and stock prices, *Journal of Accounting and Economics* 39, 437–485.

Shanken, Jay, 1992, On the estimation of beta-pricing models, *Review of Financial Studies* 5, 1–33.

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

Soliman, Mark T., 2008, The use of Dupont analysis by market participants, *The Accounting Review* 83, 823–853.

Thomas, Jacob K., and Huai Zhang, 2002, Inventory changes and future returns, *Review of Accounting Studies* 7, 163–187.

Tibshirani, Robert, 1996, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 58, 267–288.

Tibshirani, Robert, 2011, Regression shrinkage and selection via the lasso: A retrospective, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73, 273–282.

Titman, Sheridan, John Wei, and Feixue Xie, 2004, Capital investments and stock returns, *Journal of Financial and Quantitative Analysis* 39, 677–700.

Tuzel, Selale, 2010, Corporate real estate holdings and the cross-section of stock returns, *Review of Financial Studies* 23, 2268–2302.

Valta, Philip, 2016, Strategic default, debt structure, and stock returns, *Journal of Financial and Quantitative Analysis* 51, 197–229.

Wainwright, Martin J., 2009, Sharp thresholds for high-dimensional and noisy sparsity recovery using l_1-constrained quadratic programming (lasso), *IEEE Transactions on Information Theory* 55, 2183–2202.

Welch, Ivo, 2008, The link between Fama-French time-series tests and Fama-Macbeth cross-sectional tests, Technical report, UCLA.

Whited, Toni M., and Guojun Wu, 2006, Financial constraints risk, *Review of Financial Studies* 19, 531–559.

Xing, Yuhang, 2008, Interpreting the value effect through the q-theory: An empirical investigation, *Review of Financial Studies* 21, 1767–1795.

Yogo, Motohiro, 2006, A consumption-based explanation of the cross-section of expected stock returns, *Journal of Finance* 61, 539–580.

Zhang, Cun-Hui, and Jian Huang, 2008, The sparsity and bias of the lasso selection in high-dimensional linear regression, *The Annals of Statistics* 36, 1567–1594.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: Internet Appendix.
Replication code.