An Agnostic and Practically Useful Estimator of the Stochastic Discount Factor

By

Kuntara Pukthuanthong, Richard Roll, and Junbo Wang

Authors' Coordinates		
Pukthuanthong	Roll	Wang
Department of Finance University of Missouri Columbia, MO 65211	Caltech Pasadena, CA 91125 & Emeritus, UCLA	Department of Finance Louisiana State University Business Education Complex 2900 Baton Roughe, LA 70803
619-807-6124	626-395-3890	(781) 258-8806
pukthuanthongk@missouri.edu	rroll@caltech.edu	junbowang@lsu.edu

Abstract

We propose an estimator for the stochastic discount factor (SDF) that does not require macroeconomic proxies or preference assumptions. It depends only on observed asset returns yet is immune to the form of the multivariate return distribution. It does not depend on factor structure, thus making our SDF ideal for asset pricing test. We find that SDF is correlated with a large number of the factor/firm characteristics, which confirms Kozak, Nagel, and Santosh (2018)'s finding. Moreover, assets within the equity market are not fully integrated as SDFs constructed from different asset groups are uncorrelated.

Keywords: Stochastic discount factor, asset pricing tests, agnostic estimator, practical estimator

An Agnostic and Practically Useful Estimator of the Stochastic Discount Factor

Abstract

We propose an estimator for the stochastic discount factor (SDF) that does not require macroeconomic proxies or preference assumptions. It depends only on observed asset returns yet is immune to the form of the multivariate return distribution. It does not depend on factor structure, thus making our SDF ideal for asset pricing test. We find that SDF is correlated with a large number of the factor/firm characteristics, which confirms Kozak, Nagel, and Santosh (2018)'s finding. Moreover, assets within the equity market are not fully integrated as SDFs constructed from different asset groups are uncorrelated.

Keywords: Stochastic discount factor, asset pricing tests, agnostic estimator, practical estimator

We propose an estimator of the Stochastic Discount Factor (SDF) using only the information of asset returns. The new method is inspired by recognizing that the SDF appears in a particular mathematical object, an integral equation. The solution to this integral equation makes our proposed estimator novel in several respects. First, it does not depend on macroeconomic proxies or preferences, unlike some estimators in the previous literature. It is constructed from observed asset returns only, which is why we call it a "practical" and "agnostic" estimator. Second, our estimator does not depend on the distributional properties of asset returns when the numbers of stocks and the sample periods are large. Third, the estimator is immune to the factor structure. It applies to single- and multi-factor data. Finally, our estimator is immune to the grouping of assets asymptotically. Suppose N assets share a common SDF. If N assets and sample period converge to infinity, the SDF estimator will be statistically indistinguishable when derived from N assets as a whole or from any subsets of size N/2, N/3, etc. That is, if there is an inherent true factor structure that prices all assets, the SDF does not depend on the number of stocks.

Although our estimator is quite easy to compute and is asymptotically consistent, its finite sample property is more relevant for the empirical application. In the finite sample, we show that the estimator does contain a bias. Moreover, by presenting our estimator in the form of ridge regression, we find a tradeoff between bias and estimation error of our SDF. We conduct simulations to gauge the size of the finite sample bias and its dependents. We also propose an estimator to reduce the estimation errors and conduct a battery of simulations to uncover the accuracy of the estimator. We use our agnostic SDF on the applications unaffected by the bias but rely more on the estimator's accuracy.

To examine the finite sample properties, we perform a horserace between our estimator and other estimators described below. Given our estimator does not depend on factor structure, a comparable candidate is the one that imposes factor structure while having other properties similar to ours. The most comparable measure is the one proposed by Kim and Korajczyk (2021, henceforth KK), who, modify it by adding the assumption that returns and the SDF itself are all generated by the same linear factor model to construct the SDF. Although this additional structure reduces estimation error and corrects the bias when the sample size is large, the assumption that all assets and the SDF itself follow an identical factor process is obviously not the only possibility or even the most realistic one. Some assets might not conform to a factor structure at all, and the number of factors need not be the same for each and every asset.[1] Even if it does, the basic SDF must prevail for all assets within an integrated market since, in this case, the SDF is unique.[2] With the assumption of a factor structures, a similar approach can be applied using the Hansen and Jagannathan (1991, henceforth HJ)'s

[1] This would be particularly problematic when testing whether different groups of assets are in the same market, as we do later in this paper. Two groups of assets could have different factor structures. The Kim Korajczyk approach, however, cannot test this issue.

[2] See Back (2017, pp. 56-57), for the joint conditions about the existence of a unique the SDF, no arbitrage opportunities, and the law of one price.

framework by regressing one on factor returns. The HJ method with this assumption is equivalent to GMM with identity matrix.

We show that the KK method is similar and sometimes inferior to the HJ method in a finite sample. If the factor model is correctly specified, the KK method and HJ method can achieve almost no bias, with very low estimation error (characterized by high R^2). In contrast, the agnostic method contains biases and has relatively low R^2. However, in practice, which factors price the assets is still questionable. If the test assets do not all have the same factor structure and the factor structure is not correctly specified, the KK method could lead to a 30% to 100% bias with R^2 as small as 1%. Moreover, a much simpler method (the HJ approach) using factors as the test assets can achieve a similar or even lower bias and higher R^2. Applying our approach, the bias can be much smaller, and the R^2 can be much larger. This is because our approach does not depend on a presumption of a factor structure. When the SDF is estimated by more factors than the actual number of assets, the biases in KK and HJ are small. However, the estimation errors can be very large, with low R^2s. Our proposed method can generate much higher R^2s.

We apply our agnostic SDF in asset pricing tests. Ideally, we should directly test "E(mR)=1", but the direct test would result in biases, as described in KK as well as in our simulations. The test result, in this case, is hard to interpret. Therefore, we present three applications to test asset pricing models that are not affected by the bias. These applications are discussed below.

The asset pricing literature has proposed many characteristics/factors. A natural question is whether they all affect asset prices. Researchers have attempted to answer this question, but the results have been in conflict. Neoclassical theory suggests that asset returns are driven by a small number of characteristics, but the new insight seems to suggest that the asset returns are affected by a much larger number of characteristics (Kozak, Nagel, and Santosh (2019); Bryzgalova, Pelger, and Zhu (2020)). Following Cochrane (2009), if a factor could price assets, SDF should be correlated with that factor. In fact, Cochrane (2009) shows that SDF should be a linear combination of pricing factors. Based on this theory, if we want to examine whether the long and short portfolios of factors can price assets, we can test whether they are correlated with SDF.

Classical methods assume that the SDF is a linear function of a small number of known characteristics (a predetermined factor structure) and use a large sample size (T) to identify the SDF. However, suppose the true asset returns are driven by many characteristics and T is smaller or close to the number of characteristics. In that case, the classical methods for constructing SDFs might not work because the SDF is overfitted or unidentified. In addition, a predetermined factors structure cannot fully identify the SDF if not all factors are included in the structure. The agnostic SDF is a perfect method for this test because it is constructed only by asset returns, without a presumption of characteristics. We can have an impartial test on which characteristics should affect asset prices. Our first application is to create a factor zoo and examine which factors are correlated with the agnostic SDF. Recent papers show that the returns are priced by factors constructed from non-linear combination characteristics (Kirby (2019); Bryzgalova, Pelger, and Zhu (2020)); thus, the first application examines the effect of non-linear characteristics where the testing assets are mainly portfolios. The number of non-linear characteristics can be very large; thereby, we apply the regularization techniques such as Lasso and Elastic Net method to select the most critical non-linear characteristics. Consistent with the recent literature, we find that the equity market can be affected by a large number of non-linear characteristics, even with the regularization.

Notably, although the benefit of using portfolios is that they have returns over the full sample, Jegadeesh et al. (2019) and Giglio, Xiu, and Zhang (2021) argue that portfolios constructed by stocks can mask individual stocks' relevant risk- or return-related characteristics. Therefore, in the second application, we examine which linear characteristics are related to individual stock returns. Kelly et al. (2019), Fama and French (2018), and Bessembinder et al. (2019) all emphasize the importance of linear characteristics in individual stock pricing. We can carry out a similar analysis using our agnostic SDF. Specifically, if a linear combination of characteristics can proxy for the beta for the risk factors (Kelly et al. (2019)), a time-varying covariance between SDF and returns should also be a linear combination of characteristics. We regress the covariance on the characteristics from Green, Hand, and Zhang (2017) and test whether the characteristics are significantly related to this covariance. Similar to the first application, we find that a large number of characteristics can affect the covariance.

Third, we study whether equity markets are integrated. If equity markets are perfectly integrated, the same SDF should price all equity assets. Thus, if the SDFs constructed from different asset groupings are distinct (indicating that the different assets are priced by different factors), the market is not integrated. Our agnostic method does not have any presumption about factor structures, which makes it appropriate to study market integration. We test whether different assets can be explained by the same SDF by regressing the SDF estimated from one set of assets on the SDF estimated from another set. In theory, if the slope is one between one estimated SDF and another estimated SDF, two asset markets would indicate perfect integration. Nevertheless, even with perfect integration, perfectly correlated estimated SDFs from two sets of assets is unlikely because of bias and estimation error in SDF. On the other hand, if two SDFs estimated from two sets of assets are uncorrelated, it is unlikely that the markets for these two sets of assets are integrated. We find that the SDFs constructed by different equity assets can be uncorrelated; therefore, there seems to be a segmented market for equity. Moreover, the difference in SDFs should be smaller as sampling errors are mitigated. Thus, we also apply our estimation-error-reduced method to create SDFs. We find that the SDFs constructed by different equity assets are still uncorrelated.

The Appendix shows that the three applications described above do not depend on the SDF biases. Before presenting a formal derivation of the estimator, we embark on a brief literature review that places it in the context of previous work. This is followed by sections containing the derivation, remarks on the sampling distribution, simulations, and finally, empirical applications. A summary concludes.

I. Previous SDF Literature

The Stochastic Discount Factor (the SDF) has become a dominant paradigm in recent asset pricing literature. For example, Ferson (1995) shows how the main asset pricing results (mean/variance efficiency, multi-beta models) are special cases of the basic SDF relation. Cochrane (2005) begins with the SDF relation in chapter 1 and expands it into almost all other known models of assets. Bossaerts (2002) relies on the SDF paradigm in criticizing empirical work on asset pricing. The SDF foundation is established in the first chapter of Singleton (2006) and exploited to study asset price dynamics. Campbell (2014) ordains the SDF as "The Framework of Contemporary Finance," (p. 3) in his essay explaining the 2013 Nobel Prizes awarded to Fama, Hansen, and Shiller. Excellent reviews are provided by Ferson (1995) and Cochrane and Culp (2003).

The empirical success of the SDF approach is less apparent. In many (but not all) previous empirical applications, the SDF is proxied by a construct that depends typically on aggregate consumption, but occasionally on some other macroeconomic quantity, combined with a risk aversion parameter. For example, Cochrane (1996) employs aggregate consumption changes along with power utility (and a particular level risk aversion) to measure the SDF. Despite giving this specification every empirical benefit of the doubt, Cochrane (2005, p. 45) admits that it still "…does not do well." A similar imperfect fit between consumption changes, over various horizons, and both equities and bonds, is reported by Singleton (1990).

Lettau and Ludvigson (2001) add macro variables such as labor income and find that the deviation in wealth from its shared trend with consumption and labor income has strong predictive power for excess stock returns at business cycle frequencies, thereby suggesting that risk premia vary countercyclically. Chapman (1997) adds technology shocks and several conditioning variables, transforming them with orthogonal polynomials, which serve to eliminate the small firm effect but still produce "statistically and economically large pricing errors", (p. 1406.) Da and Yun (2017) employ electricity generation as a proxy for aggregate consumption.[3] Adrian, Crump and Moench (2013) employ an exponential function of a grouping of state variables, which are themselves principal components of Treasury bond returns.

In research published just prior to the hegemony of the SDF paradigm, Long (1990) shows that a "Numeraire" portfolio has many similar properties. Long's Numeraire portfolio η has strictly positive gross returns $(1+R_\eta)$ and exists only if there is no arbitrage within a list of assets from which it is composed. In this case, the expected value of the ratio $(1+R_j)/(1+R_\eta)$ is unity for all assets j on the list, which implies that $1/(1+R_\eta)$ is essentially the same as the modern the SDF. Long notes that the Numeraire portfolio is also the growth optimum portfolio. The latter is examined by Roll (1973) who provides an empirical test of whether the expected ratio above is the same for all assets. (He does not find evidence against it.)

[3] See also the variety of specifications discussed by Cochrane and Hansen (1992) in Section III, "Other Candidate Discount Factors."

Recognizing that aggregate consumption changes are too "smooth" to be well connected with asset prices (Mehra and Prescott [1985)]) and that consumption is likely measured with significant error (Rosenberg and Engle [2002]), recent literature avoids aggregate consumption data. In addition to Rosenberg and Engle, such an approach is taken by Aït-Sahalia and Lo (1998, 2000), and Chen and Ludvigson (2009). However, as pointed out by Araujo, Issler, and Fernandes (2005) and Araujo and Issler (2011), the above scholars still find it necessary to impose what might be considered rather ad hoc restrictions on preferences.

It is widely understood that the SDF is (by definition) whatever process generates the equality in Equation (1) below. Hansen and Jagannathan (1991) avoid the specification of preferences and are still able to develop their famous bound on the mean and volatility of the SDF, given that the SDF is unique. A sample of more recent literatures gives a better sense of how widespread this understanding is. Dew-Becker and Giglio (2016) adopt a frequency domain approach and a very general specification for the SDF process (their equation 1). Then they discuss "consumption-based models" with several different types of preferences including Power Utility (Section 2.1), Habit formation (2.2), and Epstein-Zin (2.3.) Sections 3 and 4 of their paper provide more general approaches but they use consumption growth as an observable factor and/or Epstein-Zin preferences.

Alvarez and Jermann (2005) use this agnostic representation to study how volatile permanent shocks to the SDF process must be; nonetheless, their objective is to derive a bound on the permanent component of the SDF. To estimate the SDF, they suggest a representative agent with recursive preferences and a particular consumption process (see p. 2004.) Martin (2017) places minimal restrictions on the features of the SDF such that option prices allow us to recover forward-looking expected returns. He makes a significant contribution by deducing certain properties of the SDF, particularly that it must be highly volatile and negatively associated with returns. To determine bounds or likely volatilities and correlations are valuable insights because they represent properties that the true SDF should possess and consequently suggest tests for any the SDF estimator. Alvarez and Jermann (2005) and Martin (2017) offer such insights but do not constitute an estimator per se.

Ross (2015) proposes to recover the full forward-looking distribution of the SDF in an agnostic manner using derivative prices. He motivates his result economically in terms of the marginal rate of substitution between current and future consumption. This is completely compatible with our own approach, which is to propose an empirical estimator for the SDF that depends only on returns. Ross's objective is different from ours. He succeeds in recovering the objective probability distribution while we propose an estimator of the SDF that depends only on asset returns and can be easily implemented. In addition, we propose to estimate the past realizations of the SDF during a given historical period, not an estimate of the future the SDF as in Ross's paper. Our estimator does not require option prices, preferences, state prices, macroeconomic proxies, or anything other than a sample of asset returns.

Campbell (1993) surmounts the annoyance of preference specifications with various approximations of nonlinear multiperiod consumption and portfolio-choices. He develops a formula for risk premia that can be tested without using consumption data and suggests a new way to use imperfect data about both market returns and consumption.

Araujo, Issler and Fernandes (2005, hereafter AIF) get around these difficulties by noting that the SDF should be the only serial correlation common feature of the data in the sense of Engle and Kozicki (1993). Then, by exploiting a log transform of returns, they derive a measure of the SDF that does not depend on a macroeconomic variable (notably including the problematic aggregate consumption) and also avoids the imposition of preferences.

Araujo and Issler (2011, hereafter AI) take a similar tack, noting via a logarithmic series expansion that the natural logarithm of the SDF is the only common factor in the log of all returns. Thus, the log the SDF can be eliminated by a simple difference in returns. Essentially, the log the SDF represents the (single) common APT factor in the sense of Ross (1976).

In both AIF and AI, the SDF measure is a function of average arithmetic and geometric asset returns. AIF compute their measure empirically and report its temporal evolution along with various statistical properties. They also compare it to the time series of riskless returns. AI find that relatively low risk aversion parameters are consistent with their estimated SDFs. They are able to price some stocks successfully, but not stocks with low capitalization levels.

Both AIF and AI essentially assume that the SDF is unique, rejecting that proposition only indirectly in the case of AI with low cap stocks. Our the SDF estimator does not depend on a factor model or a logarithmic approximation, or any other structural condition. Also, it applies regardless of the multivariate distribution of returns, whatever its form, provided that certain lower order moments exist.[4] Kozak et al. (2018) apply several characteristics and suggest different methodologies to obtain a robust the SDF from various traded factors, which achieves the bound by Hansen and Jagannathan (1991).

II. An Agnostic Estimator for the SDF

This section first shows (in Section II.A) how SDFs can be approximated by a transformation of returns, without any additional information about preferences, consumption or other macro-economic data. The following Section (II.B) proves that the same SDF estimator arises naturally from minimizing a particular sum of average surprises. This development allows us to infer some useful properties of the SDF estimator. Section II.C contains a proof that the estimator is asymptotically consistent and Section II.D shows an alternative proof of consistency.

II.A. Estimating the SDF from Returns Alone

[4] We explain the required moments below.

Let $p_{i,t}$ denote the cash value of asset i at time t. When markets are complete, the SDF paradigm implies the existence of a unique m_t, the SDF, such that

$$E_{t-1}(\tilde{m}_t \tilde{p}_{i,t}) = p_{i,t-1} \qquad \forall i,t.^{5} \tag{1}$$

where $\tilde{m}$ and $\tilde{p}$ mean random SDF and price (m and p are realized SDF and price). Denoting a gross return between t-1 and t by $R_{i,t} \equiv p_{i,t}/p_{i,t-1}$, Equation (1) is the same as

$$E_{t-1}(\tilde{m}_t \tilde{R}_{i,t}) = 1 \qquad \forall i,t.^{6} \tag{2}$$

In the SDF literature on consumption-based asset pricing, Eq. (2) was originally derived and the SDF is considered a function of preferences and consumption, often as held by a representative investor. From a mathematical perspective, however, Eq. (2) is a special case of an Integral or Volterra equation and, indeed, it is the simplest of these, called a "Fredholm equation of the first type."[7] As such, m_t is simply an unknown mathematical function, not necessarily related to anything economic including the probability distribution of R. This is why it should be obvious from the outset that estimators of the SDF could, but need not, depend on the multivariate form of the distribution of returns.[8]

Integral equations often do not have analytic solutions, so mathematicians and physicists solve them numerically, typically by a "quadrature rule" whereby a system of equations with an equal number of unknowns provides a set of discrete values for the unknown function, which in our application would be some set of observations m_t for, say, t=1,…,T. We are proposing an analogous approach, discretizing as usual but with an over-identified system whose solution is rendered unique by a statistical restriction on the error of estimation.

We begin by noting that the expectation in Eq. (2), must correspond to a realization at time t; i.e.,

$$m_t R_{i,t} = E_{t-1}(\tilde{m}_t \tilde{R}_{i,t}) + \varepsilon_{i,t} \tag{3}$$

where $\varepsilon_{i,t}$ denotes the (complete) surprise in the mR product for asset i in period t. For each time period t, the realization in (3) is determined by whatever state occurs among the many encapsulated in the expectation

---

[5] For a representative agent, m is the discounted future marginal utility of consumption divided by the current marginal utility of consumption. The tilde denotes a random variable as of period t-1.

[6] Eq. (2) is the only moment condition required by the SDF theory. However, the basic the SDF relation applies similarly to multiple periods; e.g., $E_t(\tilde{m}_{t+\tau} \tilde{R}_{i,t+\tau}) = 1$ for $\tau>1$ where the gross return and m span τ periods. This could provide some interesting features involving a term structure of SDFs but we do not explore that possibility in this paper.

[7] We are grateful to Francis Longstaff for pointing out this isomorphism. See also Polyanin and Manzhirov (1998). It is implied in McCulloch (2003) who shows that the SDF (or "pricing kernel") has finite payoffs even when returns follow stable laws whose second moments are infinite.

[8] In terms of the asymptotic distribution of the estimator, we are not specific about it because there is no single distribution that would apply in all cases. Our estimator is a function of observed returns; thus, it would have their distribution. If returns had a Gaussian, for instance, so would the estimator. Returns might also have a more complex distribution such as being driven by a factor structure. Perhaps returns have fat tails, etc.

(2). The surprise is complete if expectations are rational; i.e., if agents can freely change their expectation in response to new information.

We hasten to add that the surprise is complete only in a time series sense, for period t, but the surprises are very likely to be correlated across assets at t. There are two reasons: first, returns themselves are usually cross-sectionally correlated, $\mathrm{Cov}(R_{i,t}, R_{j,t}) \neq 0$, for $j \neq i$ and second, the products in (3) have a common element, the SDF itself, so $Cov(\tilde{m}_t\tilde{R}_{it}, \tilde{m}_t\tilde{R}_{jt}) \neq 0$.[9]

Since there is a state realization for each time t, over T time periods, we have, from Eqs. (2) and (3),

$$\frac{1}{T}\sum_{t=1}^{T} m_t R_{i,t} = \frac{1}{T}\sum_{t=1}^{T}\left[E_{t-1}(\tilde{m}_t\tilde{R}_{i,t})\right] + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{i,t} = 1 + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{i,t} \cong 1 \; . \tag{4}$$

where the approximation indicates that the average surprise is not exactly zero in a finite sample, though it should vanish for each asset separately as $T \to \infty$. The summation on the left of Eq. (4) is essentially the sample analog of the expectation on the left of Eq. (2), which implies that its approximation will improve as T increases.

The approximation error in Eq. (4) equals the time series sample mean of the surprises in the SDF-gross return product, a mean for asset i which we hereafter denote

$$\bar{\varepsilon}_i \equiv \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{i,t} \; .$$

Rational expectations rules out any serial dependence in the surprises,

$$\mathrm{Cov}(\varepsilon_{i,t}, \varepsilon_{i,t-j}) = 0, \; j \neq 0$$

but the surprises could be heteroscedastic. Hence,

$$\mathrm{Var}(\bar{\varepsilon}_i) = \frac{1}{T^2}\sum_{t=1}^{T}\mathrm{Var}_t(\varepsilon_{i,t}) = \frac{1}{T}\bar{\sigma}_i^2$$

where $\bar{\sigma}_i^2$ denotes the mean variance of surprises for asset i over the particular sample period, t=1,…T. Unless the mean variance is growing without bound, the approximation error should disappear as T grows larger.

As mentioned earlier, the individual surprises in (3) are likely to be cross-sectionally correlated at time t, so the average surprised are likely to be as well.

Now consider a sample of N assets with <u>simultaneous</u> observations over T periods, with N > T. The ensemble of gross returns for the N assets can be expressed as a matrix **R** (hereafter boldface denotes a matrix or vector). There are N columns in **R** and the i^{th} column is $[R_{i,1}:\ldots:R_{i,T}]'$. We also need a column vector **m** ≡ $[m_1:\ldots:m_T]'$ to hold T realized values of the SDF and a N-element column unit vector **1** ≡ [1:…:1]'. The entire SDF ensemble of realizations for all assets and periods can then be written compactly as

[9] We are grateful to a previous referee for pointing out this likely cross-sectional correlation.

$$\mathbf{R'm}/\mathrm{T} \cong \mathbf{1}. \tag{5}$$

Pre-multiply Eq. (5) by **R**, to obtain

$$(\mathbf{RR'})\mathbf{m}/\mathrm{T} \cong \mathbf{R1}.$$

Since we have chosen N > T, the cross-sectional time-product matrix **RR'** is non-singular unless there are two or more periods with linearly dependent cross-sectional vectors of returns.[10] Hence, we can usually solve for a time-varying vector of estimated stochastic discount factors as

$$\mathbf{m}/\mathrm{T} \cong (\mathbf{RR'})^{-1}\mathbf{R1}. \tag{6}$$

N.B.: It is very important to emphasize that our solution (6) absolutely requires the number of assets to exceed the number of time periods; i.e., N > T. The condition does not imply anything egregious such as the existence of an arbitrage because we are simply estimating T sample realizations of m, not the entire state space of m in each time period t.[11]*In Section III, we will extend our solution to the scenarios when N is not necessarily larger than T.*

Hansen and Jagannathan (1991, p. 233) derive an expression that appears similar to Eq. (6), but the resemblance is superficial.[12] Their expression involves a covariance matrix of payoffs (or returns). Our **RR**' is not a covariance matrix. They note that their solution involves the first and second moments of the future payoffs and prices. If **RR**' above were diagonal, Eq. (6) would also involve first and second moments but in this case the (sample) moments would be the cross-sectional mean return in each period divided by the cross-sectional mean of the individual squared returns in that period. We examine the HJ method in simulation for a special case. When there is a factor model that can price assets, the HJ method is equivalent to estimate coefficients of a linear combination of factors in a GMM setting. If we assume that the SDF is linear on all asset returns, as long as the number of assets is smaller than the sample periods, the HJ method can be applied.

II.B. The Minimum Sum of Squared Average Surprises

The exact form of Eq. (5), (i.e., with no approximation), is

$$\mathbf{R}'\mathbf{m}/\mathrm{T} = \mathbf{1} + \bar{\boldsymbol{\varepsilon}} \tag{7}$$

[10] That is, unless the return of every individual asset in a given period is a linear function of the returns on that asset in another or several other periods, (not that the returns are linearly dependent relative to each other in a given period), the cross-sectional time-product matrix **RR'** is non-singular.

[11] A large cross-section of stocks is also advocated by Pelger and Lattau (2017) and used for a different purpose, estimating risk factors.

[12] The Hansen/Jagannathan approach is implemented for performance measurement by Chen and Knez (1996) and is further refined by He, Ng, and Zhang (1999.)

where $\bar{\boldsymbol{\varepsilon}}$ is the column N vector that contains the average surprises for each asset. A least squares estimator for **m** is available by minimizing the sum of squared average surprises with respect to **m**; i.e.,

$$\min_{\mathbf{m}}(\bar{\boldsymbol{\varepsilon}}'\bar{\boldsymbol{\varepsilon}}) = (\mathbf{R'm}/\mathrm{T}-\mathbf{1})'(\mathbf{R'm}/\mathrm{T}-\mathbf{1}).$$

The first-order condition is

$$\frac{\partial}{\mathbf{m}}(\mathbf{m'RR'm}/\mathrm{T}^2 - 2\mathbf{m'R1}/\mathrm{T}) = 2\mathbf{RR'm}/\mathrm{T}^2 - 2\mathbf{R1}/\mathrm{T} = \mathbf{0}$$

and the minimal values is achieved for the $\hat{\mathbf{m}}$ that satisfies

$$\hat{\mathbf{m}}/\mathrm{T} = (\mathbf{RR'})^{-1}\mathbf{R1} \tag{8}$$

which shows that $\hat{\mathbf{m}}$ is the approximation Eq. (6) in Section II.A. The second order condition is strictly positive because **RR'** is positive definite (by assumption); hence $\hat{\mathbf{m}}$ provides the minimum sum of squares for the average the SDF surprises.

One may legitimately question why the estimator in Eqs. (6) or (8) should involve a cross-sectional sum of returns (**R1**) in each period. Obviously, it is different from the estimator of Hansen and Jagannathan (1991) who derive an SDF that is equal to a mean/variance tangency portfolio plus an error term. Our estimator is dictated by the mathematical fact that the basic SDF Eq. (2) has a 1.0 on the right side for every asset in every period while the tangency portfolio can change over time.

Another possible question might occur to some readers in that the estimator seems to use information across all observed time periods even though for any given period t within T, the SDF is a random variable. But the answer is simply that we are estimating the best fit to the entire <u>vector</u> $\hat{\mathbf{m}}$ whose elements have already occurred as realizations of the random variable at each t. There is an associated surprise in each t as well and the estimator simply minimizes the sum of squared average surprises. This puzzlement is similar to the one we would be faced if we run a market model regression, (t=1,...,T) to estimate the intercept and slope. Another issue is the look-ahead bias. In estimating the SDF at time t (when t<T), returns at t+1 until T are used. The SDF is used for asset pricing tests. Therefore, we can always avoid the look-ahead bias by using return from time 1 until t to construct the time-varying SDF.

The least squares estimator in Eq. (8) differs from a standard regression estimator in one important respect; since the "dependent" variable here is the T element unit vector, (with every element a constant 1.0), the estimator could be biased in finite samples.

This setup is reminiscent of Britten-Jones (1999), who shows that a no-intercept regression of the unit vector (as dependent variable) on a matrix of returns (as independent variables) yields slope coefficients that are proportional to the weights of the sample mean/variance efficient tangency portfolio from the riskless rate. Britten-Jones notes, as we do also, that a dependent variable of all 1's induces a connection between the regression residuals and the explanatory variables which raises finite sample concerns.

The big difference between Britten-Jones' regression and our superficially similar Equation (8) is that his is time series and ours is cross-sectional. As a consequence, his coefficients are proportional to cross-sectional portfolio weights while our coefficients constitute a time series vector of estimated the SDF realizations.[13]

To elucidate this issue, solve Eq. (7) for $\mathbf{1}$ and substitute the result in Eq. (8), which simplifies to,

$$\hat{\mathbf{m}} - \mathbf{m} = -\mathrm{T}(\mathbf{RR}')^{-1}\mathbf{R}\bar{\boldsymbol{\varepsilon}} \;. \tag{9}$$

The expected value of this expression is the bias. Expanding $\mathbf{R}\bar{\boldsymbol{\varepsilon}}$ term by term,

$$\mathbf{R}\bar{\boldsymbol{\varepsilon}} = \frac{1}{\mathrm{T}}\mathbf{R}\sum_{s=1}^{\mathrm{T}}\boldsymbol{\varepsilon}_s$$

For time t specifically, $\mathbf{R}_t\bar{\boldsymbol{\varepsilon}} = \frac{1}{\mathrm{T}}\mathbf{R}_t\sum_{s=1}^{\mathrm{T}}\boldsymbol{\varepsilon}_s$. We observe that most of the elements are close to zero because they involve products such as $(\varepsilon_{j,t}\ R_{i,t-k})$ for $k\neq 0$. However, there are a few elements that are unlikely to disappear. For period t, $\mathbf{R}_t\boldsymbol{\varepsilon}_t$ can be written as

$$R_{1,t}\varepsilon_{1,t} + R_{2,t}\varepsilon_{2,t} + \ldots + R_{N,t}\varepsilon_{N,t} = m_t(R_{1,t}^2 + R_{2,t}^2 + \ldots + R_{N,t}^2) - \sum_{j=1}^{N} R_{j,t}$$

The above equation is not zero because $\varepsilon_{i,t}$ is the pricing error for stock i at time t, and it should be correlated with the return of the stock at the same time. When T is large, the impact of this term ($\mathbf{R}_t\boldsymbol{\varepsilon}_t$) on $\mathbf{R}_t\bar{\boldsymbol{\varepsilon}}$ is small. The bias should be small. But if T is finite, the impact can be large, so the bias could be large.

We will study the extent of the resulting bias in the next section using simulation but note already that the bias terms are atypical because the dependence between the explanatory variables (the R's) and the disturbances (the ε's) is not linear. As we can see, the bias is only in finite sample (T is finite). Kim (2017) provides a formal proof of the consistency. Our proof of consistency is presented in the next subsection.

The estimator in Eq. (8) shares some attractive features with OLS regression estimates. In particular, it can be used to define residuals, estimates of the true disturbances, as

[13] Interestingly, Britten-Jones foresaw that a similar approach might one day help estimate the SDF. He said, "Future research could attempt to analyze and identify stochastic discount factors using the statistical inference procedures developed here," (p. 657, footnote 7.) He was right and he was very close.

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{R}'\hat{\mathbf{m}}/\mathrm{T} - \mathbf{1} = \mathbf{R}'(\mathbf{RR}')^{-1}\mathbf{R1} - \mathbf{1} = -[\mathbf{I}\text{-}\mathbf{R}'(\mathbf{RR}')^{-1}\mathbf{R}]\mathbf{1} \tag{10}$$

The matrix in brackets in Eq. (10) is idempotent, so the sum of squared residuals divided by the degrees-of-freedom, N-T, is

$$\frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{\text{N-T}} = \frac{\mathbf{1}'[\mathbf{I} - \mathbf{R}'(\mathbf{RR}')^{-1}\mathbf{R}]\mathbf{1}}{\text{N-T}} = \frac{\mathrm{N}}{\text{N-T}} - \frac{\mathbf{1}'[\mathbf{R}'(\mathbf{RR}')^{-1}\mathbf{R}]\mathbf{1}}{\text{N-T}} \tag{11}$$

The mean squared residual in (11) clearly depends on both N and T and on their difference. Because the mean surprises are likely to be cross-sectionally dependent (recall the discussion on page 8 above), we do not have a simplified analytic expression for (11), but we do investigate it later with simulations.

The square root of Eq. (10) gives the standard error of the estimate,

$$\mathrm{s} \equiv \sqrt{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}/(\mathrm{N}-\mathrm{T})}\,.$$

The covariance matrix of the estimated SDFs is given by

$$\mathrm{E}[(\hat{\mathbf{m}} - \mathbf{m})(\hat{\mathbf{m}} - \mathbf{m})' \,|\, \mathbf{R}, \mathrm{T}] = (\mathbf{RR}')^{-1}\mathbf{R}\mathrm{E}(\mathbf{V}_{\boldsymbol{\Sigma\varepsilon}})\mathbf{R}'(\mathbf{RR}')^{-1} \tag{12}$$

where the (N X N) symmetric matrix $\mathbf{V}_{\boldsymbol{\Sigma\varepsilon}}$ has the following element in the j[th] row and k[th] column:

$$(\varepsilon_{j,1} + \varepsilon_{j,2} + \ldots + \varepsilon_{j,T})(\varepsilon_{k,1} + \varepsilon_{k,2} + \ldots + \varepsilon_{k,T})\,.$$

Unlike the analogous covariance matrix of disturbances in standard OLS regressions, the diagonal elements of $\mathbf{V}_{\boldsymbol{\Sigma\varepsilon}}$ are not necessarily equal to each other and the off-diagonal elements need not have zero expectation. However, we can safely assume that cross-products separated in time, such as $\varepsilon_{j,t}\varepsilon_{k,\tau}$ for $t \neq \tau$, are zero; otherwise, the ε's would not be surprises. This implies that the element in the j[th] row and k[th] column of $\mathbf{V}_{\boldsymbol{\Sigma\varepsilon}}$ reduces to $\sum_{t=1}^{T}\varepsilon_{j,t}\varepsilon_{k,t}$. Moreover, if the ε's are not correlated across assets, an arguably unlikely condition, this sum has an expected value of zero for j≠k and then E($\mathbf{V}_{\boldsymbol{\Sigma\varepsilon}}$) becomes diagonal and equal to $\mathbf{I}\boldsymbol{\sigma}^2_{\Sigma\boldsymbol{\varepsilon}}$ where I is the identity matrix and $\boldsymbol{\sigma}^2_{\Sigma\boldsymbol{\varepsilon}}$ is the N element column vector whose j[th] element is $\mathrm{Var}(\sum_{t=1}^{T}\varepsilon_{j,t})$. If the variance of the surprises were the same scalar σ^2 for all assets and time periods, perhaps an even more dubious condition, then Eq. (12) simplifies further to

$$\mathrm{E}[(\hat{\mathbf{m}} - \mathbf{m})(\hat{\mathbf{m}} - \mathbf{m})' | \mathbf{R}, \mathrm{T}] = \mathrm{T}\sigma^2(\mathbf{RR}')^{-1} \tag{13}$$

Except for the presence of T, this is the standard regression covariance matrix of the coefficients given IID disturbances. The square roots of the T diagonal elements of Eqs. (13) provide the standard errors of SDFs period-by-period.

II.C. Proof of Asymptotic Consistency

This proof, presented in Kim (2017), establishes the consistency of the agnostic estimator based on the following

Assumptions: As N,T $\rightarrow \infty$, (i) $\frac{\mathrm{N}}{\mathrm{T}} \rightarrow \infty$, (ii) there exists a positive constant c such that the minimum eigenvalue of $\frac{\mathbf{R}'\mathbf{R}}{\mathbf{N}}$ is larger than c and (iii) $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} < \infty$.

The first condition is that the size of cross-section increases faster than the size of the time-series, which is generally true for many data sets except, perhaps, with trade-to-trade observations. The second condition states that the minimum eigenvalue of $\frac{\mathbf{R}'\mathbf{R}}{\mathbf{N}}$ does not vanish, which can be rationalized as follows: Note that $\frac{\mathbf{R}'\mathbf{R}}{\mathbf{N}}$ is decomposed into a component due to systematic factors (whether this component exists or not) and a component due to idiosyncratic shocks. Since the minimum eigenvalue of $\frac{\mathbf{R}'\mathbf{R}}{\mathbf{N}}$ is larger than that of each component, the condition is guaranteed by simply imposing that the cross-sectional variance of idiosyncratic shocks do not vanish. The last condition is that aggregate mispricing in the economy is bounded, which is in parallel with APT of Ross (1976).

The agnostic the SDF estimator $\hat{\mathbf{m}}$ converges to the true SDF in mean squared error sense by exploiting a large panel return data.

Theorem 2.1. Under Assumption 1, as N, T $\rightarrow \infty$,

$$\frac{1}{\mathrm{T}}(\hat{\mathbf{m}} - \mathbf{m})'(\hat{\mathbf{m}} - \mathbf{m}) \xrightarrow{\mathrm{p}} 0.$$

All proofs are in Appendix B or in Kim (2017).

From the asymptotic analysis, our estimator is not a projection on observed sample returns. At first, it might seem to be such a projection because the estimator does employ returns; hence, one could easily infer that the estimator is akin to a sample mean/variance efficient portfolio, which, of course, is composed differently across various sub-samples of assets.[14] But a close examination of our estimator belies such intuition. Instead of a projection on asset returns, it is actually a projection on time periods. As a consequence, for finite N, it does not depend on by the distributions of returns or even by their identity as long as a unique SDF prices all assets in the cross-section.

III. A more generalized framework for SDF

In this section, we construct the time-varying SDF in a more general framework. We also discuss a few extensions of the methods that can alleviate a concern that N has to be greater than T. The finite sample bias and estimation error issues are examined in the simulations. Finally, we discuss the applications that depend and not depend on the finite sample bias of the agnostic SDF.

III.A Model Setup

Let $\mathrm{m}(\mathbf{Z_t})$ be the SDF that depends on the economic state variables $\mathbf{Z_t}$ (a K by 1 vector). $\mathbf{Z_t}$ can be macro variables or conditioning information following Hansen and Richards (1987), Ferson and Siegel (2001), and Cochrane (2005). Denote $\mathbf{R_t}$ (an N by 1 vector) the raw stock returns. The conditional asset pricing model can be written as follows:

$$\mathrm{E}(\mathrm{m}(\mathbf{Z_t})\mathbf{R_t}|\mathbf{Z_t}) = \mathbf{1}_\mathrm{N}. \tag{14}$$

Here, $\mathbf{1_N}$ is an N by 1 vector, with each element equal to 1. The above equation can be rewritten as

$$\mathrm{E}((\mathrm{m}(\mathbf{Z_t})\mathbf{R_t} - \mathbf{1}_\mathrm{N})|\mathbf{Z_t}) = \mathbf{0}_\mathrm{N},$$

where $\mathbf{0_N}$ is an N by 1 vector, with each element equal to 0. The conditional expectation above implies that for any function (f) measurable with respect to the space of state variable $\mathbf{Z_t}$, $\mathrm{E}((\mathrm{m}(\mathbf{Z_t})\mathbf{R_t} - \mathbf{1}_\mathrm{N})\mathrm{f}(\mathbf{Z_t})) = \mathbf{0}_\mathrm{N}$. Rewriting it, we obtain

$$\mathrm{E}(\mathrm{m}(\mathbf{Z_t})\mathbf{R_t}\mathrm{f}(\mathbf{Z_t})) = \mathrm{E}(\mathrm{f}(\mathbf{Z_t}))$$

The space of measurable function contains the linear functions, polynomials functions, and even functions that are not differentiable (but still measurable with respect to the information at time t). Theoretically, we could

[14] Such intuition is readily overturned by thinking about the SDF as an unknown function in an integral equation; see Section II.A.

choose any number (even uncountable number) of functions. To simplify, assume that we only select M functions. Stack these functions into a vector $\mathbf{f}(\mathbf{Z_t}) = \begin{pmatrix} f_1(\mathbf{Z_t}) \\ \vdots \\ f_M(\mathbf{Z_t}) \end{pmatrix}$, we can write the model as

$$E(m(\mathbf{Z_t})(\mathbf{R_t} \otimes \mathbf{f}(\mathbf{Z_t}))) = E(\mathbf{1}_N \otimes \mathbf{f}(\mathbf{Z_t}))$$

Here $\otimes$ is the Kronecker product. Given that $\mathbf{R_t}$ is an N by 1 vector and $\mathbf{f}$ is an M by 1 vector, the Kronecker product $\mathbf{R_t} \otimes \mathbf{f}(\mathbf{Z_t})$ and $\mathbf{1}_N \otimes \mathbf{f}(\mathbf{Z_t})$ are NM by 1 vectors. Define $\widetilde{\mathbf{R}}_t = \mathbf{R_t} \otimes \mathbf{f}(\mathbf{Z_t})$ and call it the augmented asset returns, the above equation becomes

$$E(m(\mathbf{Z_t})\widetilde{\mathbf{R}}_t) = E(\mathbf{1}_N \otimes \mathbf{f}(\mathbf{Z_t})) \tag{15}$$

Thus, we obtain a model that should price MN augmented assets. Even if N is not very large, a substantial value of M can significantly expand the total number of assets. Two implications arises: first, we do not require N to be larger than T in this case. Second, we should establish the statistical property for large MN (which can be much larger than T) augmented assets.

We can further simplify the model. Define $\bar{\mathbf{f}}(\mathbf{Z_t}) = \begin{pmatrix} \bar{f}_1(\mathbf{Z_t}) \\ \vdots \\ \bar{f}_M(\mathbf{Z_t}) \end{pmatrix}$, where for any i, $\bar{f}_i(\mathbf{Z_t}) = f_i(\mathbf{Z_t})/E(f(\mathbf{Z_t}))$. Let $\overline{\mathbf{R}}_t = \mathbf{R_t} \otimes \bar{\mathbf{f}}(\mathbf{Z_t})$ be the adjustd augmented return. Hence, the above equation can be written as

$$E(m(\mathbf{Z_t})\overline{\mathbf{R}}_t) = \mathbf{1}_{NM}. \tag{16}$$

Here, $\overline{\mathbf{R}}_t$ and $\mathbf{1}_{NM}$ are both NM by 1 vectors. There is a special case of the above model, when conditioning information is a constant. Then the model reduces to the model without conditioning information. I.e., $\bar{\mathbf{f}}(\mathbf{Z_t}) = 1$, where 1 is a scalar. Hence, the model becomes $E(m\mathbf{R}_t) = \mathbf{1}_N$.

To estimate SDF, we following the same agnostic approach introduced in section II. We can regress $\mathbf{1}_{NM}$ on adjustd augmented return $\overline{\mathbf{R}}_t$. Specifically, define $\overline{\mathbf{R}} = (\overline{\mathbf{R}}_1, \cdots, \overline{\mathbf{R}}_T)$. The agnostic estimator of the SDF can be written as:

$$\widehat{\mathbf{m}}(\mathbf{Z}) = T(\overline{\mathbf{R}}'\overline{\mathbf{R}})^{-1}(\overline{\mathbf{R}}'\mathbf{1}_{NM}). \tag{17}$$

Here $\widehat{\mathbf{m}}(\mathbf{Z})$ is an T by 1 vector, which is the time-series estimation of the SDF. And $\mathbf{1}_{\mathbf{NM}}$ is an MN by 1 vector of ones.

When conditioning information is not incorporated, we obtain the neoclassical factor model:

$$\mathbf{R_t} = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{F_t} + \boldsymbol{\varepsilon}. \tag{18}$$

In this case, the SDF estimator becomes:

$$\widehat{\mathbf{m}} = \mathrm{T}(\mathbf{R}'\mathbf{R})^{-1}(\mathbf{R}'\mathbf{1}_{\mathrm{N}}). \tag{19}$$

We can also regress $\mathbf{1}_{\mathrm{N}} \otimes \mathbf{f}(\mathbf{Z_t})$ on augmented return $\widetilde{\mathbf{R}}_{\mathbf{t}}$ to obtain $\mathrm{m}(\mathbf{Z_t})$. See the Internet Appendix A for the implementation details.

III.B Factor Structure

Kim and Korajczyk (2021) assume that there exists a known a factor structure that generates asset returns. This is a critical distinction relative to our assumption. Although a factor structure may exist, its form and composition unknown to investors and researchers. For example, Kozak, Nagel and Santosh (2019) find that the SDF can be determined by a large set of factors; thus, looking for only a few factors can be futile. It is natural to remain agnostic about the SDF and attempt to estimate it without any presumption about the true factor structure. Moreover, we will show later in the simulation section that an error in the assumption about factors can lead to an the SDF that is uncorrelated with the true SDF.

Nonetheless, it may be useful theoretically to assume a set of ***unknown*** factors represented by $\mathbf{F_t}$, a K×1 vector. The true return process can then be written as

$$\overline{\mathbf{R}}_{\mathbf{t}} = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{F_t} + \boldsymbol{\vartheta}. \tag{20}$$

Here $\boldsymbol{\alpha}$ (a NM by 1 vector) and $\boldsymbol{\beta}$ (a NM by K matrix) are the mispricing and factor loadings for all assets, and $\boldsymbol{\vartheta}$ (a NM by 1 vector) is the residual of the model, representing the idiosyncratic error.

Recall that $\overline{\mathbf{R}}_{\mathbf{t}} = \mathbf{R_t} \otimes \overline{\mathbf{f}}(\mathbf{Z_t})$, so the factors $\mathbf{F_t}$ can price not only $\mathbf{R_t}$, but also the interaction of returns and any measurable functions of conditioning information.

Following Kim and Korajczyk (2021), there is a unique SDF determined by this factor model:

$$\mathbf{m}(\mathbf{Z_t}) = \delta + \mathbf{F_t}\boldsymbol{\delta}_\mathbf{F}. \tag{21}$$

Here $\delta = \frac{1}{\lambda_0}(1 + \boldsymbol{\mu}_\mathrm{F}{}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_\mathrm{F})$ and $\boldsymbol{\delta}_\mathbf{F} = -\frac{1}{\lambda_0}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}_\mathrm{F})$, with $\boldsymbol{\mu}_\mathbf{F}$ the mean of factors, $\boldsymbol{\lambda}_\mathbf{F}$ the risk premium associated with factors, λ_0 is the intercept of the APT model,[15] and $\boldsymbol{\Sigma} = \mathbf{E}\big((\mathbf{F_t} - \boldsymbol{\mu}_\mathrm{F})(\mathbf{F_t} - \boldsymbol{\mu}_\mathrm{F})'\big)$.

III.C Interpretation of the agnostic method

III.C.1 A ridge regression interpretation

The agnostic estimator has a natural interpretation in terms of the ridge regression. To see this, we decompose $\overline{\mathbf{R}}'\overline{\mathbf{R}}$ and obtain

$$\overline{\mathbf{R}}'\overline{\mathbf{R}} = \mathbf{F}'\boldsymbol{\beta}'\boldsymbol{\beta}\mathbf{F} + \boldsymbol{\vartheta}'\boldsymbol{\vartheta}. \tag{22}$$

When N converges to infinity and T is finite,

$$\widehat{\mathbf{m}}(\mathbf{Z}) = \mathrm{T}(\overline{\mathbf{R}}'\overline{\mathbf{R}})^{-1}(\overline{\mathbf{R}}'\mathbf{1}_{\mathrm{NM}}) = \mathrm{T}(\mathbf{F}'\boldsymbol{\beta}'\boldsymbol{\beta}\mathbf{F} + \boldsymbol{\vartheta}'\boldsymbol{\vartheta})^{-1}(\mathbf{F}'\boldsymbol{\beta}'\mathbf{1}_{\mathrm{NM}} + \boldsymbol{\vartheta}'\mathbf{1}_{\mathrm{NM}}) \to \mathrm{T}(\mathbf{F}'\boldsymbol{\beta}'\boldsymbol{\beta}\mathbf{F} + \boldsymbol{\vartheta}'\boldsymbol{\vartheta})^{-1}(\mathbf{F}'\boldsymbol{\beta}'\mathbf{1}_{\mathrm{NM}}).$$

If we assume that the aggregate idiosyncratic volatility is not time-varying, and returns are not auto-correlated,[16] $\boldsymbol{\vartheta}'\boldsymbol{\vartheta} \to \psi^2\mathbf{I}$, where, ψ is the aggregate idiosyncratic volatility, and $\mathbf{I}$ is a T by T identity matrix. Thus, the above equation is the estimator of the following Ridge Regression:

$$\mathbf{1}_{\mathrm{NM}} = \mathbf{c}(\mathbf{F}\boldsymbol{\beta}) + \boldsymbol{\vartheta} + \gamma(\psi^2 \parallel \mathbf{c} \parallel^2). \tag{23}$$

Here, $\mathbf{1}_{\mathrm{NM}}$ is the dependent variable, $\mathbf{F}\boldsymbol{\beta}$ is the independent variable, $\boldsymbol{\vartheta}$ is the regression residual, and $\|\cdot\|^2$ is the L-2 norm. Thus, $\gamma(\psi^2 \parallel \mathbf{c} \parallel^2)$ represents the constraint for the information selection criterion. For $\gamma = 1$, this is equivalent to our agnostic estimator.

The ridge regression essentially shrinks the estimated SDF over time. The level of shrinkage depends on the relative importance of $\mathbf{F}'\boldsymbol{\beta}'\boldsymbol{\beta}\mathbf{F}$ and $\psi^2\mathbf{I}$. Intuitively, if systematic risk $\mathbf{F}'\boldsymbol{\beta}'\boldsymbol{\beta}\mathbf{F}$ is relatively more important during a period, the shrinkage term is relatively less important. If idiosyncratic risk $\psi^2\mathbf{I}$ is more important, the SDF in that period shrinks more.

---

[15] From the APT, we know that $\mathrm{E}(\mathbf{R}) = \lambda_0 + \boldsymbol{\beta}\boldsymbol{\lambda}_\mathrm{F}$.

[16] The assumption is made through this section. Note that the relaxation of this assumption does not change the analysis.

The correspondence between our agnostic method and the Ridge Regression can naturally control the multicollinearity as well as the overfitting issues of the regression

$$\mathbf{1}_{\mathrm{NM}} = \mathbf{c}(\mathbf{F}\boldsymbol{\beta}) + \boldsymbol{\varepsilon}. \tag{24}$$

Specifically, if the number of factors (K) is smaller than the number of periods (T), there is a multicollinearity issue. If the number of assets (NM) is similar to the number of periods (T), there is an overfitting issue. Ridge regression can mitigate these problems.

III.C.2 A ridge regression extension

We can extend the ridge regression above to estimate the SDF to further control multicollinearity and overfitting issues. This is possible by choosing various values of γ. One possible method is to follow Kim (2017), in choosing $\gamma = \psi^2 + \mathrm{s(NM)}$, where s(NM) converges to zero when N converges to infinity.[17] For example, definining s(NM)=log(NM)/NM. The estimator becomes:

$$\widehat{\mathbf{m}}(\mathbf{Z}) = \mathrm{T}(\overline{\mathbf{R}}'\overline{\mathbf{R}} + \mathrm{s(NM)}\mathbf{I})^{-1}(\overline{\mathbf{R}}'\mathbf{1}_{\mathrm{NM}}) \tag{25}$$

In this case, the matrix $\mathrm{s(NM)}\mathbf{I}$ is invertible even if $\overline{\mathbf{R}}'\overline{\mathbf{R}}$ is not, which can control multicollinearity. Moreover, when NM is smaller than or close to T, the agnostic method has unidentifiable or overfitting issues. Ridge regression can control these issues.

III.D Bias and estimation errors

III.D.1 Bias vs. Estimation error of the SDF

Ridge regression usually produces biased estimators. This is illustrated by Kim (2017) and Kim and Korajczyk (2021). They show that when the number of the estimation period, T, is finite, the bias cannot be ignored. Therefore, to reduce both the bias and estimation error, both T and N should be large.

In this section, we show that when the number of assets (NM) is finite, T dictates a tradeoff between bias and estimation error; a higher (lower) T results in lower (higher) bias, but higher (lower) the estimation error.

The estimated SDF can be written as:

[17] There can be other choices of the parameter. For example, we can choose it so there is Bayesian interpretation for the SDF estimation.

$$\widehat{\mathbf{m}}(\mathbf{Z}) = \mathrm{T}(\mathbf{F}'\boldsymbol{\beta}'\boldsymbol{\beta}\mathbf{F} + \psi^2\mathbf{I})^{-1}(\mathbf{F}'\boldsymbol{\beta}\mathbf{1} + \boldsymbol{\vartheta}'\mathbf{1}_{\mathrm{NM}}). \tag{26}$$

Kim (2017) shows that if $\frac{1}{\mathrm{T}}\psi^2\mathbf{I} \to \mathbf{0}$, the above the SDF prices asset correctly asymptotically, i.e., they satisfy $\mathrm{E}(\widehat{\mathbf{m}}\overline{\mathbf{R}}) \to \mathbf{1}_{\mathrm{NM}}$ when both NM and T converge to infinity.[18] Hence, increasing the sample size (T becomes larger but still smaller than NM) results in a smaller bias.

On the other hand, the estimation error when NM is finite can be written as $\mathrm{T}(\mathbf{F}'\boldsymbol{\beta}'\boldsymbol{\beta}\mathbf{F} + \psi^2\mathbf{I})^{-1}(\boldsymbol{\vartheta}'\mathbf{1}_{\mathrm{NM}}) = (\mathbf{F}'\boldsymbol{\beta}'\boldsymbol{\beta}\mathbf{F}/\mathrm{T} + \psi^2\mathbf{I}/\mathrm{T})^{-1}(\boldsymbol{\vartheta}'\mathbf{1}_{\mathrm{NM}})$. When the number of factors (K) is smaller than the number of periods, $\mathbf{F}'\boldsymbol{\beta}'\boldsymbol{\beta}\mathbf{F}$ has zero rank. Thus, the invertibility of the matrix depends on $\psi^2\mathbf{I}/\mathrm{T}$. If T is large, $\psi^2\mathbf{I}/\mathrm{T}$ becomes smaller (determinant of the matrix is closer to zero), making the estimation error larger.

Thus, when the number of periods is large, the smaller the $\psi^2\mathbf{I}$, the smaller the bias. On the other hand, the estimation error becomes larger. Hence, there is a tradeoff between the bias and estimation error of the estimated SDF.

III.D.2 A ridge regression method to reduce the estimation errors

Based on the analysis before, we can lower the estimation error by choosing a smaller T. However, choosing a smaller T leads to a larger bias in pricing. Moreover, when idiosyncratic volatilities of the assets are low, the matrix $\mathbf{F}'\boldsymbol{\beta}'\boldsymbol{\beta}\mathbf{F} + \psi^2\mathbf{I}$ can still be close to singular, even if T is not very large. In these scenarios, a multicollinearity issue emerges. Hence, the ridge regression method in section **III.C.2** can avoid the issues. Specifically, we could adopt the following estimator to estimate the SDF:

$$\widehat{\mathbf{m}}(\mathbf{Z}) = \mathrm{T}(\overline{\mathbf{R}}'\overline{\mathbf{R}} + \mathrm{s}(\mathrm{NM})\mathbf{I})^{-1}(\overline{\mathbf{R}}'\mathbf{1}_{\mathrm{NM}}) \tag{27}$$

In this case, the estimation error becomes $\mathrm{T}(\mathbf{F}'\boldsymbol{\beta}'\boldsymbol{\beta}\mathbf{F} + \psi^2\mathbf{I} + \mathrm{s}(\mathrm{NM})\mathbf{I})^{-1}(\boldsymbol{\vartheta}'\mathbf{1}_{\mathrm{NM}}) = (\mathbf{F}'\boldsymbol{\beta}'\boldsymbol{\beta}\mathbf{F}/\mathrm{T} + \psi^2\mathbf{I}/\mathrm{T} + \mathrm{s}(\mathrm{NM})\mathbf{I}/\mathrm{T})^{-1}(\boldsymbol{\vartheta}'\mathbf{1}_{\mathrm{NM}})$. Since $\mathbf{F}'\boldsymbol{\beta}'\boldsymbol{\beta}\mathbf{F} + \psi^2\mathbf{I} + \mathrm{s}(\mathrm{NM})\mathbf{I} > \mathbf{F}'\boldsymbol{\beta}'\boldsymbol{\beta}\mathbf{F} + \psi^2\mathbf{I}$ (i.e., the difference between two matrices is positive definite), the standard deviation of $\mathrm{T}(\mathbf{F}'\boldsymbol{\beta}'\boldsymbol{\beta}\mathbf{F} + \psi^2\mathbf{I} + \mathrm{s}(\mathrm{NM})\mathbf{I})^{-1}(\boldsymbol{\vartheta}'\mathbf{1}_{\mathrm{NM}})$ should be smaller than that of $\mathrm{T}(\mathbf{F}'\boldsymbol{\beta}'\boldsymbol{\beta}\mathbf{F} + \psi^2\mathbf{I})^{-1}(\boldsymbol{\vartheta}'\mathbf{1}_{\mathrm{NM}})$. The cost of this method, however, is a larger bias. This is because the correction term $\mathrm{s}(\mathrm{NM})\mathbf{I}$ adds on $\psi^2\mathbf{I}$, making the $(\mathbf{F}'\boldsymbol{\beta}'\boldsymbol{\beta}\mathbf{F} + \psi^2\mathbf{I} + \mathrm{s}(\mathrm{NM})\mathbf{I})$ deviate more from $\mathbf{F}'\boldsymbol{\beta}'\boldsymbol{\beta}\mathbf{F}$.

[18] Another possible way is to make $\psi^2 \to 0$ (idiosyncratic variance converges to zero) and to achieve that, one can construct portfolios.

Why is it important to reduce the estimation error at the cost of the bias? Noisier SDFs are detrimental for examining the correlation between the SDF and the factors, while the bias does not (as we will show it in Section III.F.2).

III.D.3 Simulation evidence: bias vs. estimation error

To gauge the bias and estimation errors, as well as the tradeoff between them, we rely on the simulations. We follow the design of Kim and Korajczyk (2021). To keep it parsimonious, we assume that there is no conditioning information. All asset returns are available over all periods (a balanced panel).[19] The main messages from simulations do not depend on these assumptions.

We simulate portfolio returns from 1,720 anomaly returns following Jones, Lv, Pukthuanthong, and Wang (2020). These portfolios are built from the 172 anomalies in Hou, Xue, and Zhang (2019). Thus, they are based on the characteristics similar to that of Fama-French five factors. We also construct 2,190 portfolios with non-linear characteristics. Recent researchers examine non-linear factor structures (such as Kirby (2019) and Bryzgalova, Pelger, and Zhu (2020)). Specifically, Kirby (2019) applies a factor structure with non-linear characteristics. Following this paper, we construct portfolios based on these non-linear characteristics. Specifically, we select 9 characteristics that are known to affect cross-sectional asset prices (agr, chcsho, mom1m, mom12m, mom36m, operprof, mve, retvol, and turn; See Freyberger, Neiurehl and Weber, 2020). See the Appendix E for variable descriptions. The non-linear characteristics are second and third power cross multiplications of these 9 characteristics. This will lead to 219 of non-linear characteristics in addition to nine linear characteristics. For each of them, we sort all stocks into ten portfolios, leading to 2,190 portfolios. These portfolios are assets we use to construct the SDF in simulation as well as the empirical analysis (See Appendix F for the description of these portfolios).

We assume that all portfolios follow a Fama-French five-factor model:

$$R_{it} = \alpha + \beta_{mktrf,i}F_{mktrf,t} + \beta_{smb,i}F_{smb,t} + \beta_{hml,i}F_{hml,t} + \beta_{rmw,i}F_{rmw,t} + \beta_{cma,i}F_{cma,t} + \varepsilon_{it}$$

Here R_{it} is the return of asset i at time t, $F_{mktrf,t}$, $F_{smb,t}$, $F_{hml,t}$, $F_{rmw,t}$, and $F_{cma,t}$ represent five factors and $\beta_{mktrf,i}$, $\beta_{smb,i}$, $\beta_{hml,i}$, $\beta_{rmw,i}$, and $\beta_{cma,i}$ are the corresponding factor loadings. We estimate these loadings and keep them as the true betas of assets. The residual term ε_{it} can be calculated from the regression. For each asset, the standard error of residuals is calculated and preserved for simulations.

[19] We choose a balanced panel in the simulation, because all of our applications are based on the portfolio or individual stocks with a full sample of data.

The Monte-Carlo simulation begins with the factors. They are assumed to follow a multinomial normal distribution with mean and variances the same as the sample mean and variances of the five factors from 1962 to 2019. The residuals of asset returns are simulated similarly, following a normal distribution with mean zero and standard error calculated from the real data. The asset returns are then calculated using the betas estimated from the regression, factors, and residuals, with the Fama-French five-factor model above.

The SDF can be estimated using these asset returns. There are six SDF estimators we use for comparison. The first method is the original agnostic method. The second method is the ridge regression method or shrinkage method described in section III A. The third and fourth methods are based on Kim (2017).

$$\widehat{\mathbf{m}} = \mathrm{T}(\mathbf{R}'\mathbf{R} - \widehat{\psi}^2\mathbf{I} + \mathrm{s(N)}\mathbf{I})^{-1}(\mathbf{R}'\mathbf{1}). \tag{28}$$

Here $\widehat{\psi}$ is the estimation of the standard error of residuals. To calculate this value, we regress simulated returns on five factors and calculate the standard error of the residuals. This correction method contains two parts. The first part is to subtract the bias from the $\mathbf{R}'\mathbf{R}$, which we illustrate as our third method and it is a pure bias-correction approach for comparison in the simulations, i.e.,

$$\widehat{\mathbf{m}} = \mathrm{T}(\mathbf{R}'\mathbf{R} - \widehat{\psi}^2\mathbf{I})^{-1}(\mathbf{R}'\mathbf{1}). \tag{29}$$

If the factor structure can be correctly specified, we can estimate $\widehat{\psi}^2$ consistently and the above method can correct the bias. However, if the factor structure is misspecified, this method can still be biased. The fourth method is to include the shrinkage term $\mathrm{s(N)}\mathbf{I}$ [20] and bias-adjustment term.

The fifth method is Kim and Korajczyk (2021). They assume that there is a factor structure, and the SDF is a linear combination of these factors. Hence, the SDF is obtained through estimating the coefficients of the linear combination. Specifically, the esimtated SDF can be written as

$$\widehat{\mathbf{m}} = \mathbf{F}\mathrm{T}(\mathbf{F}\mathbf{R}'\mathbf{R}\mathbf{F}' - \widehat{\psi}^2\mathbf{F}'\mathbf{I}\mathbf{F})^{-1}(\mathbf{F}\mathbf{R}'\mathbf{1}). \tag{30}$$

Theoretically, if returns of all stocks follow the same factor model and the SDF is linear on the factors, the SDF should also price all asset returns. Following this logic, one can apply Hansen and Jagannathan method to estimate the SDF (which is essentially the GMM with identity weighing matrix) only using the factors as

[20] Note that shrinkage term $s(N)\boldsymbol{I}$ only depends on N because we assume the conditioning information is constant in the simulations.

testing assets. Thus, theoretical foundations of Kim and Korajczyk (2021) approach is equivalent to that of Hansen and Jagannathan approach. We specify this approach in term of HJ approach. Instead of using all asset returns, the SDF can be estimated using only factors. i.e. we estimate the coefficients δ and $\boldsymbol{\delta_F}$ following moment conditions:

$$\frac{1}{T}\sum_{t=1}^{T}\left((\delta + \mathbf{F_t}\boldsymbol{\delta_F})\mathbf{F_t}\right) = \mathbf{1}, \tag{31}$$

$$\frac{1}{T}\sum_{t=1}^{T}\left((\delta + \mathbf{F_t}\boldsymbol{\delta_F})R_f\right) = 1. \tag{32}$$

The estimated SDF is $\hat{\delta} + \mathbf{F_t}\hat{\boldsymbol{\delta}}_{\mathbf{F}}$. This method (we name it as HJ approach) is our final method for comparison.

The examination criteria for the methods are following Kim and Korajczyk (2021). Given that the true SDF can be calculated with the known factor structure, we regress each estimated the SDF on the true SDF:

$$\hat{\mathbf{m}} = a + b\mathbf{m} + \epsilon. \tag{33}$$

Suppose the estimated the SDF has no bias, $a = 0$ and $b = 1$. If the estimation error is small, the R-square of the regression will be large.

We repeatedly simulate returns and factors for 1000 times, estimate a, b, and R-square in each trial, and then average them over 1,000 replications. We report eight combinations of N, T, standard deviation of SDF, and the standard error of residuals ($\hat{\psi}$).

Panel A shows when N is 10,000 and T=500, the R^2 of the agnostic estimator is 47% and Panel B shows the averages are $a = 0.09$ and $b = 0.91$ for the original agnostic method, indicating that the bias is not very large. Since the portfolios are constructed by characteristics similar to that of Fama-French factors, the idiosyncratic risks (standard deviations of residuals from the factor models) of these portfolios are relatively small. On the oth er hand, The Shrinkage method (ridge regression) approach improves R^2 significantly to 79%. However, it also implies a much larger biased estimator ($a = 0.25$, and $b = 0.75$). The bias-adjustment is unreliable with almost zero R^2 in all specifications. The reason is that the idiosyncratic risks, in this case, are small, so $\hat{\psi}^2\mathbf{I}$ is small, thus blowing up standard errors. Kim's or shrinkage-bias adjustment is similar to ridge regression. Kim and Korajczyk (2021) achieve almost no bias with an R^2 of 92%. However, a much simpler method, classical GMM, achieves almost the same result. When T decreases, the R^2 increases for the agnostic method (see Section III.D.1) while it decreases for other methods. This supports our argment in Section III.D.1 where we show a tradeoff between bias and errors for the agnostic estimator when T increases. When true standard deviation of the SDF increases (doubles from the original based on our example), the R^2 decreases by 25%, 20%, 24%, 17% and 17% for the agnostic, ridge, shrinkage-bias adjustment, KK and classical GMM. When the standard deviation of idiosyncratic risk doubles, the R^2 decreases for all methods except KK and classifcal GMM. It is noteworthy the R^2 of the original agnostic estimator, the shrinkage, and the shrinkage-bias adjusted approaches decreases by 11%, 29%, and 56%. When both standard deviation of the SDF and perturbation volatility increase, R^2 decreases by 28%, 47%, 67%, 16%, and 16% for the agnostic SDF, ridge, shrinkage-bias adjustment, KK and classifcal GMM. So far, the shrinkage-bias adjustment approach is the most sensitive to an increase in standard deviations of SDF and idiosyncratic risk whereas the KK and classical GMM are similar and least sensitive of all SDF estimators.

Considering the slope and intercept, T decreases the bias of SDF eimator (the slope of all measures to be closer one) except KK and classical GMM. The perturbation volatility also increases the bias for the agnostic, shrinkage and bias adjustment method. Interestingly, although the standard deviation of SDF increases the estimation errors (decrease R^2) of the first four approach, it makes the estimator less biased for the agnostic, ridge, and the bias-shrinkage adjustment approaches. Noticably, the slope of bias-adjustment dramatically decreases from 1.12 to 0.15. Similar to the R^2 result, the slope and intercept of the KK and classical GMM are insensitive to any change in T, and SDF volatility and purtubation volatility.

III.E Why the Agnostic method is essential

Kim and Korajczyk (2021) and classical GMM can adjust for bias and enhance R-square. KK impose a factor structure to estimate the SDF. If the factor model is correctly specified, KK's method can achieve almost no bias, with very low estimation error (characterized by a high R-square) in Table 1.

In this section, however, we show that the correct factor structure is essential for KK's results. In Table 2, we repeat the same simulation as in Table 1 with one different assumption. The factor model is *not* correctly identified. Panel A assumes that the true model follows a five-factor structure, but researchers miss factors when they estimate the SDF. That is, we show the cases when researchers miss one, two three and four factors when they estimate the SDF. In Panel B, we assume that the true model follows a four-factor structure (without the HML factor), but to estimate the SDF, researchers assume one to four more factors than the true number of factors of asset model.

We apply the same methods as before. For Panel A when researchers miss factors, we can see that R-square is generally not very good for KK (unless we only miss one factor). The KK method can lead to a 100% bias ($\mathrm{a} = 1.06$, and $\mathrm{b} = -0.07$) with almost zero R-square when we miss four factors. Similar findings are shown for classical GMM. However, for the agnostic approach, the bias and the R-square are approximately the same as they are in Table 1. The reason is that the agnostic approach does not depend on the presumption of the factor structure. Note that the bias-adjusted method and Kim's method do not work well in this case and this is because the bias-adjusted term $\widehat{\psi}^2\mathbf{I}$ depends on identifying the model correctly. KK and classical GMM are fragile to misspecification. For KK, R^2 decreases from 60% to 15% and slope decreases from 0.67 to -0.07 when number of missing factors increases from one to four. The classical GMM marginally performs better than the KK but still dramatically underperforms the agnostic estimator. For the classical GMM, R^2 decreases from 64% to 15% and slope decreases from 0.69 to 0.15. For the agnostic approach, the bias is only 9% regardless of number of assets (n).

For Panel B when researchers add more factors than the true factor model, the bias is less severe than missing factors. This is because redundant factors are just noise. Under KK and HJ, the coefficients of these noise factors will be zero. However, given the existence of these factors, the SDF will be much noisier, thereby implying a small correlation with the true SDF and low R^2 in the simulations. In this case, KK performs similarly to the agnostic SDF estimators in term of R^2 (about 1% to 16%) and slope/intercept when n is 10^4. When n increases to 10^6, the R^2 of the agnostic SDF estimator increases to 91% while R^2 of KK is only between 24% to 34%. Although KK and classifical GMM perform similarly in the case of missing factors, the R^2 of the classical GMM is much higher than that of KK in all cases.

In all cases, the bias adjustment performs poorest with R^2 equal to zero or negative. The ridge or shrinkage method has higher R^2 than the agnostic estimator but its estimator is more biased than the agnostic SDF estimator when n is less than 10^6. The shrinkage-bias adjusted estimator performs similarly to the ridge estimator, but when n increases to 10^6, the R^2 of the shrinkage-bias adjusted estimator becomes zero.

We also examine if we can achieve similar simulation results (R^2 or coefficients) using two sets of assets (like one from anomaly portfolios and another from nonlinear portfolios). However, unless N and T are extremely large (N=1,000,000 T=5000), the R-squares are quite different for SDFs constructed by different assets. Thus, we hesitate to draw a conclusion that the SDF does not depend on the subsample of the assets or distribution of the assets. The simulations are in the finite sample; thus, we can still emphasize our method works well in theory (or asymptotically).

From the simulations above, even though the agnostic the SDF and the ridge regression methods contain larger estimation error and bias, it can preserve the critical pricing features of the actual model. In asset pricing tests, some research questions are more sensitive to bias, while others are not. If the application is not highly susceptible to bias, the agnostic method can still be applied. In the next section, we will discuss the applications that depend on bias and those that do not rely on the bias.

III.F Applications and the Bias

III.F.1 Applications depending on the bias

If we want to test the null hypothesis "$\mathrm{E}(\mathrm{m}\mathbf{R}) = \mathbf{1}$", bias is an important issue. This can be shown through simulations below.

We follow the same simulation design as in the previous section. In each trial, we estimate the SDF, then calculate the following pricing error $\frac{1}{T}\left(\sum_{t=1}^{T} m_t R_{it}\right) - 1$ for each asset. Given that there are 1000 replications, we can estimate the mean and standard error of the above pricing error. The T-stats for pricing errors for each asset can be calculated. If the method has no bias, and the distribution of the mean pricing error is normal and the distribution of T-stats should be close to a standard normal. Hence, if we sort t-stats and pick 5%, 10%, 50%, 90%, and 95% critical values, they should be close to -1.96, 1.69, 0, 1.69 and 1.96.

We study four scenarios. First, we use anomaly asset returns to construct the SDF and calculate the pricing errors with these assets. Second, we use non-linear asset returns to create the SDF and calculate the pricing errors. Third, we use anomaly assets to construct the SDF and calculate the pricing errors for nonlinear assets. Fourth, we use nonlinear assets to build the SDF and calculate the pricing errors for anomaly assets. The third and fourth scenarios are expected to result in large bias as we do not use the same assets used to construct the SDF and test pricing. We use portfolios from non-linear characteristics because lately they are found to explain variation of cross-sectional stock returns (see Kirby (2020) and Bryzgalova, Pelgers and Zhu (2020), for instance).

There are four methods in our comparison: agnostic, ridge regression, bias-adjusted , and Kim.[21] The results are shown in Table 3.

From Table 3, when we use the SDF to price the same assets used to construct it, the distribution of the T-ratio of pricing errors for the agnostic method is close to a normal distribution (Panels A and B). For ridge regression and Kim's adjustment, the bias can be large. The bias-adjusted approach can lead to a small absolute

[21] We drop KK (2020) method and Classical GMM. If their model assumption is correct, they will not lead to a bias in pricing errors. Although this additional structure reduces estimation error and corrects the bias when sample size is large, the assumption that all assets and the SDF itself follow an identical factor process is obviously not the only possibility or even the most realistic one. Some assets might not conform to a factor structure at all and the number of factors need not be the same for each and every asset. Moreover, it is less likely that the SDF would necessarily follow such a process. Our goal is to evaluate the bias for agnostic method.

critical value. This is because the method has to invert a close-to-singular matrix, which can produce a large error of the estimated SDF and the corresponding pricing error. These outliers enlarge the standard error, leading to a smaller T-ratios.

On the other hand, if we use one set of assets to create the SDF and use that to price another set of assets, the bias can be substantial. For example, in panels C and D of Table 3, the bias in T-stat can be significant even for the original method. The bias-adjusted method produces reasonable T-distribution. However, the method is based on the assumption that the factor structure is correctly specified. When it is misspecified, the results are much more biased.

The distributions of t-stats for the four methods in Table 3 are plotted in Appendix A Figure I. In each plot, there are four subplots; the top left represents our agnostic method, the top right represents the ridge regression method, the bottom left represents the bias-adjusted method, and bottom right represents Kim's method. The most interesting finding is that the distribution of the bias-adjusted method is not bell-shaped. Given that the inversion of the singular matrix can lead to a significant error in the estimated SDF, the distribution of the T-ratios can be non-normal.

From the simulations, we have shown that the estimation of pricing errors contains a large bias. The estimation of the standard error of $m(\mathbf{Z})$ and $m(\mathbf{Z})\mathbf{R}$ can suffer similar biases. Thus, sorting on the pricing errors (such as He, Huang, Rapach, and Zhou (2020) with an agnostic SDF), or the calculating the idiosyncratic volatility of the pricing errors (Ang et al. (2006)) all suffer biases.

III.F.2 Applications that do <u>not</u> depend on the bias

As mentioned in Section III.D.2 that, it is important to reduce the estimation error at the cost of the bias. Noisier SDFs are detrimental for examining the correlation between the SDF and the factors, while the bias does not. In this section, we present three applications of an agnostic SDF estimator. These applications do not suffer from the bias in the SDF. Moreover, the agnostic method is critical for these application as explained below.

(1) Find the factors that determine the SDF. There are many factors proposed in asset pricing literature. We want to examine whether they can be correlated with the SDF. Moreover, recent papers declare that the returns are priced by factors from non-linear combinations characteristics (Kirby (2019) and Bryzgalova, Pelger, and Zhu (2020)). Examining whether the SDF is correlated with non-linear characteristics is also important. We apply the regularization techniques (See Table 4 where we apply Lasso or Appendix H1 and H2 for Elastic Net method) to select the most critical factors. We can also obtain the significance of coefficients of regressing the SDF on selected factors.

(2) Examine which characteristics are related to asset returns. Kelly et al. (2019) (the IPCA paper), Fama, and French (2018) and Bessembinder et al. (2019) all emphasize the importance of characteristics in asset pricing. We can test it using agnostic the SDF. Specifically, the time-varying covariance $\mathrm{cov}(\mathrm{m}(\mathbf{Z_t}), \overline{\mathbf{R}}_\mathrm{t})$ should be a linear combination of characteristics if characteristics can proxy for beta (as IPCA paper suggests). We regress $\mathrm{cov}(\mathrm{m}(\mathbf{Z_t}), \overline{\mathbf{R}}_\mathrm{t})$ on the characteristics from Green, Hand, and Zhang (2017). With daily data and a rolling window, we can estimate a time-varying SDF without conditioning information ($\mathrm{cov}(\mathrm{m}, \mathbf{R}_\mathrm{t})$). Specifically, in each month, we use the daily stock returns from the past three years to construct the SDF and calculate the covariance between the estimated the SDF and the stock return. We then examine the correlation between this covariance and the characteristics using a panel regression method.

(3) We can test whether different assets can be explained by the same SDF. In theory, if estimated slope is one with 100% R-square, two asset markets should be perfectly integrated. But it is unlikely to see perfectly correlated SDFs from two sets of assets, because of the bias and the estimation error from the estimated SDFs. On the other hand, if SDFs estimated from two sets of assets are uncorrelated, it is unlikely that the markets for these two sets of assets are integrated. Thus, if we regress the SDF constructed from one market on that from another market, we can examine whether the slope coefficient is insignificantly different from zero.

In Internet Appendix B, we show that three applications do not depend on the bias. The application (3) requires more discussion because it is related to testing market integration. To make it simple, assume that conditioning information is not incorporated. Then the estimated SDF is in Equation (6). The vector on the right side of this equation is an estimate based on N assets and a sample period of length T, a combination of cross-sectional and time series observations. SDF paradigm also contends that any other set of assets within the same integrated market should produce, aside from sampling variation, the same m from concurrent time series observations. Hence, if we denote by m(k) a sample m computed according to Eq. (6) (where k indicates a set of K assets) and then, from the same calendar observations, choose a complement set $j \not\subseteq k$ with J assets, the SDF null hypothesis of market integration can be expressed as

$$H_0: E[m(k)-m(j)]=0 \tag{34}$$

Notice that K and J need not be equal. This test is reminiscent of DeSantis (1993) and Ferson (1995), who suggest comparing SDFs derived from a subset of assets to SDFs derived from all available assets. Testing for the equivalence of pricing operators across two groups of assets is also explored by Chen and Knez (1994)[22] and, in the context of the APT, by Brown and Weinstein (1983).[23]

It is important to emphasize that the philosophy of the above test is standard; i.e., we will never be able to prove that the two SDFs are exactly the same and that compared markets are indeed completely integrated, but we do have the possibility of rejecting these implications. If markets are not complete and integrated, an infinite number of stochastic discount factors sastisfy Eq(1) because $E_{t-1}[(\tilde{m}_t + \tilde{\omega}_t)\tilde{p}_t] = E_{t-1}(\tilde{m}_t\tilde{p}_t)$ whenever ω and p are orthogonal; Cf. Cochrane (2005, Section 4.1). But $\tilde{m}_t + \tilde{\omega}_t$ looks just like the true SDF plus an estimation error. Indeed, if markets are complete, $\tilde{\omega}_t$ is an estimation error because $\tilde{m}_t$ is unique. On the contrary, if markets are incomplete $\tilde{\omega}_t$ can differ across groups of assets and hence the null hypothesis in Eq (34) can potentially be rejected.

Ideally, if market is integrated and we regress one SDF on another, the slope coefficient should be 1. However, given the bias in SDF, the coefficient could be deviated from 1. In Internet Appendix B, we show that coefficient should still be nonzero if the market is integrated almost surely. Therefore, if we cannot reject the hypothesis that the coefficient is not zero, it is likely that the market is not integrated.

IV. Data

We employ five different sets of data. First, stock returns are from CRSP from 1962 to 2018. Similar to the standard procedure of screening stock returns in the existing literature, we collect monthly returns from the Center for Research in Security Prices (CRSP) and accounting information from the Compustat Annual and Quarterly Fundamental Files. We exclude financial firms and firms with negative book equity. See Appendix C for descriptive statistics of stocks in each year.

Second, we construct 1,720 portfolios or 172 anomalies based on candidate factors that are constructed similarly to those in Hou, Xue, and Zhang (2019). These anomalies cover six dimensions including 10 momentum, 16 value versus growth, 34 investment, 34 profitability, 49 intangibles, and 29 trading frictions. At this stage, we limit our candidate factors estimated for one month and they are value-weighted average returns. See Appendix D for the list of anomalies, their mean, t-stat of mean equal to zero, and standard deviation.

---

[22] Chen and Knez (1995) derive a measure of market integration as the minimal amount that two pricing operators differ. Their measure too does not rely on any restrictive pricing model assumptions and its applications do not depend on the validity of any asset pricing model. They use a similar framework to develop a general approach to portfolio performance measurement in Chen and Knez (1996).

[23] The Arbitrage Pricing Theory due to Ross (1976).

Third, we construct 94 firm characteristics based on Green, Hand and Zhang (2018). Jeremiah Green generously make a code available on his website. Initially, there are 102 characteristics; however, eight of them have VIF greater than 7 and thus to alleviate the multicollinearity concern, we do not include them. See Appendix E for the list of characteristics and their descriptive statistics.

Fourth, we construct 2,190 portfolios 219 long-short portfolios based on the nonlinear functions of nine characteristics that Freyberger, Nieurehl and Weber (2020) find significantly explain cross-sectional stocks returns. We call them nonlinear portfolios or factors. We apply the following procedure to construct these portfolios. See Appendix F for the list of our nonlinear factors. As an indication, we use 3 characteristics (x1, x2, and x3) as an example.

1. We generate the following characteristics up to polynomials of degree 3, including X1, X2, X3, X1X2, X1X3, X2X3, X1X2X3, X1X1X3, X1X1X2, X1X2X2, X2X2X3, X1X3X3, X2X3X3, $X1^2$, $X2^2$, $X3^2$, $X1^3$, $X2^3$, $X3^3$)

We alleviate multicollinearity concern among these characteristics by orthogonalizing each characteristics using a residual from regressing characteristics on its linear and non-linear component. For instance, we use the residual of regressing X1X2X2 on X1, X2, X1X2, and X2X2 instead of using X1X2X2 directly, or the residual of regressing X1X2 on X1 and X2, instead of using x1x2. Generally, we use X^3-C1*X-C2*X^2 where C1 and C2 are estimated coefficients from regressing X^3 on X and X^2, respectively to eliminate the impact of both X and X^2 from X^3, or X^2-C2*X, which is a residual from regressing X^2 on X.

There are two benefits from using residual. First, the residual methods can remove all the possible correlations between X and X^2. Second, if X have different sign (for instance, X1=-2, X2 = 0, and X3=1), X1<X2<X3 but $X2^2<X3^2<X1^2$ and $X1^3<X2^3<X3^3$. The relation is not monotonic if X1 to Xn have different signs. Using residuals will take care of this regardless of the sign.

2. We standardize across stocks each time to avoid look-ahead bias. Also, we could avoid the mis-ranking issue when we standardize chars firm by firm.
3. We sort stocks into deciles based on transformed chars above and calculate the average returns next period by group and assign to the corresponding chars and decile (for example, X1_1)
4. We create long-short portfolios i.e., portfolios 10 minus 1 for each transformed characteristic.

Fifth, we use 360 tree-factor portfolios developed by Bryzgalova, Pelger, and Zhu (2020)[24] (see Appendix G for the list and description). These portfolios can capture the higher dimensional non-linear information from characteristics. Lastly, we obtain the state variables from Amit Goyal's website. Those conditioning information includes such as dividend price ratio, default yield spread, risk-free rate, term spread, corporate bond returns, stock variance, and net equity expansion and Cay. The last is from Lettau and Ludvigson (2001).

Lastly, we collect the returns from multiple assets including stocks, bonds, commodities, real estate and currencies that have the data during 2002 to 2013. Our goal for this data is to test the application 3 whether SDFs across assets have the same factor structure. We start from 2002 because that is when TRACE starts offering the data and we need bond trading prices from TRACE. TRACE is the most common source researchers use for bond data. We collect 45 currencies pair against US dollar and 47 commodities from Datastream. Stocks and REITs are collected from CRSP and Ziman REIT. For a comparison, the data of other assets also start in 2002 and thus we end up with 225 bonds, 323 stocks, 47 commodities, 45 currencies, and 97 REITs representing real estate from 2002 to 2013; this sample satisfies the requirement that N exceeds T with a comfortable level of degrees of freedom. The data of 12 years might be too short and with the annual data, the statistical power might be low. We thus augment our agnostic estimator with ten conditioning information. Thus we end up with 10 times larger number of observations for each asset with 144 months (see Section III.A for our model setup for conditioning agnostic SDF estimator).

For returns of corporate bonds, we use transaction records in the Trade Reporting and Compliance Engine (TRACE). TRACE provides high-quality corporate bond intraday information for trading price, trading volume, and sell and buy indicators etc. Our sample period is from August 2002 to June 2017. We follow Bai, Bali and Wen (2018)'s data screening procedure and return estimation approach. The monthly corporate bond returns are computed from the quoted price, accrued interest, and coupon payment for a month divided by the quoted price in the previous month. A bond's excess return is the difference its computed total return and the risk-free rate, where the latter is proxied by the one-month Treasury bill rate. Our final sample consists of 331,728 observations, and the cross-sectional mean monthly excess return is 0.389%, which is comparable to Bai, Bali and Wen (2018)'s sample. We include only bonds that have at least 30 continuous monthly returns;[25] 6421 bonds remain in our final sample.

[24] We thank Marcus Pelger for generously providing us the portfolios data.
[25] The results are robust using different windows.

V. Empirical results

V.A Application 1

The tested factors are long-short portfolios constructed using the higher order of characteristics. We exclude the first moment to make sure this set of factors present only nonlinear information. We first regress our estimated SDF, $\hat{\mathbf{m}}_{\mathrm{i,t}}$, on the market factor and its higher-order term up to six polynomials (since they are not in long-short portfolios). Then, we use the residuals of this regression, $X\hat{\mathbf{m}}_{\mathrm{i,t}}$ to run the following regression:

$$X\hat{\mathbf{m}}_{\mathrm{i,t}} = \alpha + \beta_{\mathrm{N}} f_{\mathrm{N,t}} \tag{35}$$

where $\mathbf{f}_{\mathrm{N}}$ are the non-linear factors chosen by Lasso. Portfolios i includes 1720 anomalies portfolios, 360 trees portfolios, 2,100 nonlinear portfolios, and 87 stocks (see Section IV for the data description and construction). We examine two SDFs, an SDF that does not contain a shrinkage term (0^{th} percentile), and an SDF that fully does (100^{th} percentile), i.e.,

$$\hat{\mathbf{m}}(\mathbf{Z}) = \mathrm{T}(\bar{\mathbf{R}}'\bar{\mathbf{R}} + \mathrm{s}(\mathrm{N})\mathbf{I})^{-1}(\bar{\mathbf{R}}'\mathbf{1}_{\mathrm{NM}}) \tag{36}$$

We augment all of the returns by 9 state variables described in section IV to relax the requirement that N>T.[26]For instance, the tree-based returns are augmented by which means there are 360 * (9 + 1) or 3,600 assets. We do the same for stocks, anomalies and nonlinear portfolios. Table 4 show application 1 using Lasso. We also apply Elastic Net with lambda equal to 0.03 and 0.05 and the results are shown in the Appendices H1 and H2. The detailed implementation of three regularization methods (Lasso and Elastic Net) are introduced in the Internet Appendix C.

We find that many factors are selected when the tree and nonlinear portfolios are used to construct the SDF, but no factors selected when anomalies or stocks are used. There are two points to be made. First, although most of the asset pricing literature finds that there are only small number of factors, recent researches show that the number of factors could be much larger. For example, Kozak, Nagel and Santosh (2018) show that the factors constructed through many characteristics affect the SDF. This is more likely to be true if non-linear characteristics are taken into consideration. Bryzgalova, Pelger, and Zhu (2020) find that the majority component of the SDF is determined by non-linear factors. The above results are consistent with this conclusion. Second, a large number of nonlinear factors can potentially be correlated to the SDF, if the goal is to price non-traditional assets, such as portfolios constructed by the tree method or with nonlinear

[26] Those conditioning information includes such as dividend price ratio, default yield spread, risk-free rate, term spread, CORPR, stock variance, and net equity expansion and Cay. The last is from Lettau and Ludvigson.

characteristics. This result is not mechanical and in fact supports Giglio, Xiu and Deng (2021) in that using all assets to construct does not help identify the pricing factors. The pricing factors should be constructed from the assets that we want those factors to price. The anomaly portfolios and 87 stocks existing for the whole sample are likely driven by the linear factors. In our Lasso or Elastic Net regression, we only include non-linear factors; thus, it is reasonable to expect that none of them are selected. When we use the ridge regression to construct the SDF (the 100th percentile), there are more factors selected. This is because these SDFs are less noisy, leading to higher and more significant coefficients in the regression.

When we use all four portfolios to construct the SDF ($\mathbf{m}$ mixed), there is no factor selected from $\mathbf{m}$ with a shrinkage term of zero. This seemingly surprising result could indicate a fundamental important issue for the financial markets. The return for the markets are not integrated. Specifically, traders in tree and non-linear portfolios (which are likely to be mutual funds or hedge funds that explore non-linear characteristics pattern) and long-term existing stocks (these are the stocks with data over 500 months, so they might appear in many investors' portfolio) might be fundamentally different, resulting in pricing factors being unrelated. When we use a shrinkage term of 100th percentile, there are 7 factors selected, with two of them being significant. This result is likely due to the lower estimation error when we construct the SDF using ridge regression.

When we use nonlinear and tree portfolios combined to construct $\mathbf{m}$ (see the last two columns of Table 4), we find 4 factors, none of which is significant, and 7 factors, four of which are significant at least at 5% , are selected for $\mathbf{m}$ with 0th and 100th percentile of shrinkage term, respectively. This result show that the market for nonlinear and tree portfolios are still distinct. We also present the results for Elastic Net method in the Appendices H1 and H2 . These results are similar to Table 4.

Our result in application 1 presents the evidence that is contradictory to KK's underlying assumption of knowing a true factor model ex ante applied to all assets. Our results here show the number of factors can be large and depends which assets we want to price. It supports the underlying assumption of the agnostic estimator as we do not rely on knowing any factor model ex ante, but just returns.

V.B Application 2

This section examines which characteristics can price asset returns. We estimate the SDF using the daily data of stock returns over the last three years following Jegadeesh et al. (2019). The chosen stocks are those that exist for at least 90% of the days. We have more than 4000 stocks on average in each rolling window.[27] Next, we calculate the covariance between the SDF with each stock return to obtain $\mathbf{cov}(\hat{\mathbf{m}}, \mathbf{R})$ in each month. Given that we have covariance and characteristics for all stocks and all time, we can run panel predictive regression by using the past firm characteristics as predictors.

In all regressions, we exclude characteristics that have high correlation. Multicollinearity leads to unreliable T-stats; thus, we exclude eight characteristics that are most strongly related to other characteristics including betasq, dolvol, lgr, maxret, mom6m, pchquick, quick, and stdacc, each of which has a VIF > 7 (see Appendix E for the description of these characteristics).

We run the regression on selected characteristics using cross-sectional data or panel data. Table 5 shows the results of Application 2. Model (1) presents OLS and applies clustered covariance matrix estimation for panel data, estimation of sandwich covariance a la Newey-West (1987) and Driscoll and Kraay (1998) for panel data. We cluster standard errors by stocks and count the serial correlation among periods. Model (2) includes time fixed effect without inference HAC and HAC standard errors for 36 lag since the covariances are estimated using 3-year data.

Among 94 characteristics, Earnings to price (ep), bid-ask spread (baspread), 12-month momentum (mom12m), % change in depreciation (pchdepr), 1-month momentum (mom1m), 36-month momentum (mom36m), % change in gross margin - % change in sales (pchgm_pchsale), change in inventory (chinv), idiosyncratic volatility (idiovol), illiquidity (ill), price delay (pricedelay), change in number of analysts (chnanalyst), asset growth (agr), industry-adjusted change in asset turnover (chatoia), earnings announcement returns (ear), returns on assets (roaq), capital expenditures and inventory (invest), gross profitability (gma), organizational capital (orgcap) are chosen across models at least with 5% significance level.

If we want to avoid the multi-testing issues (Harvey, Liu and Zhu (2016)) by imposing a T-stat greater than 3, we find earnings to price (ep), bid-ask spread (baspread), 12-month, 1-month, and 36-month momentum (mom12m, mom1m, mom36m), change in inventory (chinv), % change in gross margin, % change in sales (pchgm_pchsale), idiosyncratic return volatility (idiovol), illiquidity (ill), change in number of analysts (chnanalyst), asset growth (agr), earnings announcement returns (ear), gross profitability (gma), and organizational capital (orgcap) are chosen across two models.

[27] At month t, we calculate the covariance using daily data from month t-36 to month t-1. There are approximately 250 times 3 or 750 days for each month t.

These results are consistent with Kozak, Nagel and Santosh (2018). The SDF can contain many linear characteristics. It might be difficult to ignore some economic meaningful characteristics in asset pricing model.

V.C Application 3

We construct SDFs using different test assets, and test whether the market of these assets are integrated. Our testing assets include SDFs constructed from different asset classes (such as stocks, bonds, commodities, real estate, and currencies),[28] and portfolios (such as anomaly portfolios, portfolios based on non-linear characteristics, individual stocks, and tree portfolios). These assets are augmented by conditioning information.

We run a univariate regression on all other SDFs and test whether it is significantly different from zero. If two markets are integrated, SDFs are the same, which implies that beta should be nonzero. Thus, if beta is not significantly different from zero, it is very likely that two markets are disintegrated. The results of these univariate regressions are shown in Table 6. Panel A presents the results for different asset classes. With shrinkage term of zero, we find bonds and REIT and commodity and stocks present very high negative coefficient. For $\mathbf{m}$ with shrinkage term of one, we find currency and bonds, and currency and REITs have significant and negative coefficient estimate. But in most cases, the correlations are still small and insignificant, thereby suggestion asset classes are disintegrated.

Rather than applying individual assets, Panel B presents $\mathbf{m}$ constructed from different portfolios. Consistent with the results in Table 4, we find augmented $\mathbf{m}$ constructed with stock returns (the portfolios constructed based primarily on linear characteristics in some cases) are not integrated with SDFs constructed by tree and nonlinear portfolios based on non-linear characteristics. This is likely to be true because these stocks in our sample are likely to be large stocks and should be explained by common linear factors. On the other hand, the slope coefficient between tree portfolios and non-linear portfolios is significantly different from zero, thereby suggesting tree portfolios are likely to be priced by the non-linear factors. In short, the SDF from non-linear factors could be quite different from the SDF constructed from the large stocks. Therefore, if two SDFs estimated from two sets of assets are uncorrelated, it is likely that the markets for these two sets of assets are not integrated.

Our result here supports those in Table 4, where we extract the SDF from stocks. We show that the SDF implied by individual stocks are not correlated with any nonlinear factors and trees (more than 200 factors and

[28] There are 323 stocks, 225 bonds, 47 commodities, 19 pairs of currencies against US dollar and 97 REITs to present real estate. We start the sample period in 2002 when bond returns are available in TRACE. Since we augment our SDF by states variables, we can increase to N*10, which allows us to use monthly return data.

none is selected). Hence, there may not be a common factor structure for both assets, confirming that the two markets are not integrated. In general, the significant level in panel B does not change when we apply different shrinkage/ridge approach to construct the SDF. This implies that the integration of the markets is unlikely to be driven by the estimation error in SDFs.

IX. Conclusions

The stochastic discount factor (SDF) paradigm predicts that the same SDF should price all assets in a given period when markets are complete. This implies that an SDF estimated from one subset of assets can be used to price assets from other subsets within the same (complete) market.

We derive an SDF estimator that can be calculated from observed returns only and is not a function of macroeconomic state variables, preferences, and the form of the multivariate distribution of returns, including its factor structure. We emphasize that an SDF estimator needs not depend on the above-mentioned traditional elements because the SDF is a mathematical function within an integral equation. This property of the SDF implies that it does not depend on the distribution of returns, including their factor structure, if any.

Our SDF estimator is consistent but biased in finite samples and has a standard error that depends on both the number of asset, N, and the number of time periods, T, used in its construction. Although there is a finite sample bias, the applications of this method do not need to depend on the bias. We apply the agnostic SDF estimator in three important asset pricing questions. Our results suggest that a large number of characteristics can affect asset prices, and assets in the equity markets might not be integrated.

Table 1: Simulation to show the bias and R-square

The portfolio returns are simulated using Fama-French five factors. We regress returns on factors and collect their loadings, and regression residuals. The mean and standard deviation of the factors and residuals are calculated and used for the simulations. In each month, the factors and residuals are simulated following a normal distribution with these mean and standard deviation. The return is then calculated following the five-factor model using factor loadings as well as the simulated factors and residuals. $\hat{\psi}$ is the estimation of the standard error of residuals. There are six methods to estimate the SDF: No-adjust (original agnostic method), Shrinkage (ridge regression approach), bias-adjust method, Shrinkage and bias-adjust (Kim (2017) method), KK method (Kim and Korajczyk (2021)), and classical GMM. The detail of these methods is introduced in Section III.C and III.D. We assume researchers make the correct assumption in that the true model follows a five-factor structure and they apply the same model to estimate the SDF.

Panel A

				R^2					
T	**N**	σ_m	$\hat{\psi}$	**Agnostic**	**Shrinkage**	**Bias-adjust**	**Shrinkage and bias-adjust**	**KK**	**Classical GMM**
T=500	10,000	org	org	47%	79%	0%	75%	92%	92%
T=250	10,000	org	org	54%	69%	0%	69%	85%	85%
T=150	10,000	org	org	54%	57%	0%	60%	77%	78%
T=500	1,000	org	org	6%	55%	0%	56%	90%	91%
T=500	1,000,000	org	org	90%	90%	0%	0%	92%	92%
T=500	10,000	org	2org	36%	50%	0%	19%	92%	92%
T=500	10,000	2org	org	22%	59%	0%	51%	75%	75%
T=500	10,000	2org	2org	19%	32%	0%	8%	76%	76%

Panel B

					Regression Coefficients					
T	N	σ_m	std_error	Intercept/ slope	no adjust	Shrinkage	Bias-adjust	Shrinkage and bias-adjust	KK	Classical GMM
500	10,000	org	org	intercept	0.09	0.25	-0.12	0.20	0.00	0.00
				slope	0.91	0.75	1.12	0.80	1.00	1.01
250	10,000	org	org	intercept	0.16	0.39	-0.33	0.32	-0.01	-0.01
				slope	0.84	0.61	1.33	0.68	1.01	1.02
150	10,000	org	org	intercept	0.24	0.51	-0.38	0.43	0.00	-0.02
				slope	0.76	0.49	1.40	0.57	1.01	1.02
500	1,000	org	org	intercept	0.09	0.63	0.13	0.62	0.01	0.00
				slope	0.91	0.36	0.87	0.37	0.99	1.00
500	1,000,000	org	org	intercept	0.10	0.10	0.00	0.01	0.01	0.01
				slope	0.90	0.90	1.00	0.99	0.99	1.00
500	10,000	org	2org	intercept	0.28	0.38	-0.41	0.18	0.00	0.00
				slope	0.72	0.62	1.42	0.81	1.00	1.00
500	10,000	2org	org	intercept	0.02	0.09	0.84	0.06	-0.01	-0.01
				slope	0.98	0.91	0.15	0.94	1.02	1.02
500	10,000	2org	2org	intercept	0.10	0.16	3.06	0.06	-0.01	-0.01
				slope	0.90	0.84	-2.10	0.94	1.01	1.02

Table 2: Misspecified model

The portfolio returns are simulated using Fama-French five factors. We regress returns on factors and collect their loadings, and regression residuals. The mean and standard deviation of the factors and residuals are calculated and used for the simulations. In each month, the factors and residuals are simulated following a normal distribution with these mean and standard deviation. The return is then calculated following the five-factor model using factor loadings as well as the simulated factors and residuals. There are six methods to estimate the SDF: No-adjust (original agnostic method), Shrinkage (ridge regression approach), bias-adjust method, Shrinkage and bias-adjust (Kim (2017) method), KK method (Kim and Korajczyk (2021)), and classical GMM. The detail of these methods is introduced in Section III.C and III.D. There are 1000 trials for simulations. Panel A presents the results when the true model follows a five-factor structure, but to estimate the SDF, researchers assume that we have four, three, two and one factor. Similarly, in Panel B, we assume that the true model follows a four-factor structure, but researchers assume a five-, six- and seven-factor model to estimate the SDF. a and b are intercept and slope, respectively.

Panel A: Missing factors

N	**T**	**no adjust**	**Shrinkage**	**Bias-adjust**	**Shrinkage and bias-adjust**	**KK**	**Classical GMM**	**KK**	**Classical GMM**	**KK**	**Classical GMM**	**KK**	**Classical GMM**
						1-factor missing		**2-factor missing**		**3-factor missing**		**4-factor missing**	
							R^2						
1000	500	6%	57%	0%	57%	62%	64%	32%	39%	9%	15%	15%	15%
							Coeffs						
a		0.08	0.63	0.39	0.62	0.30	0.30	0.59	0.57	1.09	0.83	1.06	0.85
b		0.92	0.37	0.62	0.38	0.70	0.70	0.41	0.43	-0.10	0.17	-0.07	0.15
							R^2						
10^4	500	46%	79%	0%	75%	62%	64%	32%	39%	8%	15%	15%	15%
							Coeffs						
a		0.09	0.25	0.33	0.20	0.30	0.30	0.59	0.57	1.09	0.84	1.06	0.85
b		0.91	0.75	0.67	0.80	0.69	0.70	0.41	0.43	-0.10	0.16	-0.07	0.15
							R^2						
10^5	500	89%	89%	0%	0%	60%	62%	30%	37%	9%	14%	15%	15%
							Coeffs						
a		0.10	0.10	0.01	-0.03	0.33	0.33	0.60	0.59	1.09	0.84	1.07	0.86
b		0.90	0.90	0.99	1.03	0.67	0.68	0.39	0.41	-0.10	0.16	-0.07	0.15

Panel B: More factors

N	T	no adjust	Shrinkage	Bias-adjust	Shrinkage and bias-adjust	KK	Classical GMM	KK	Classical GMM	KK	Classical GMM	KK	Classical GMM
						1 more factor		2 more factors		3 more factors		4 more factors	
							R^2						
10^3	500	1%	46%	0%	46%	1%	15%	2%	22%	4%	35%	16%	82%
							Coeffs						
a		0.06	0.63	9.84	0.61	0.00	0.00	-0.14	0.00	0.05	0.00	-0.10	0.00
b		0.94	0.37	-8.83	0.38	1.00	1.00	1.15	1.00	0.95	1.00	1.10	1.00
							R^2						
10^4	500	15%	58%	$6.60*10^{-5}$	46%	1%	16%	2%	22%	5%	35%	15%	83%
							Coeffs						
a		0.05	0.17	-1.86	0.13	-0.01	-0.01	-0.03	-0.01	-0.01	-0.01	0.01	-0.01
b		0.95	0.83	2.86	0.88	1.01	1.01	1.03	1.01	1.01	1.01	0.99	1.01
							R^2						
10^6	500	91%	92%	$6.03*10^{-5}$	0%	24%	16%	24%	23%	27%	36%	34%	81%
							Coeffs						
a		0.02	0.02	-0.21	0.09	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04
b		0.98	0.98	1.21	0.92	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04

Table 3: Bias in the asset price model

The simulation procedure is the same as in Tables 1 and 4. We examine the bias in the equation "$\mathrm{E}(\mathrm{m}\mathbf{R}) = \mathbf{1}$," We calculate the following pricing error $\frac{1}{T}\left(\sum_{t=1}^{T} m_t R_{it}\right) - 1$ for each asset in each trial. Then estimate the mean and standard error of the pricing error over 1000 trials. The T-ratios for pricing errors for each asset can be calculated. We then sort T-ratios and pick 5%, 10%, 50%, 90%, and 95% critical values. Panel A presents the case to use anomaly asset returns to construct the SDF and calculate the pricing errors with these assets. Panel B presents the case to use non-linear asset returns to construct the SDF and calculate the pricing errors. Panel C presents the case to use anomaly assets to construct the SDF and calculate the pricing errors for nonlinear assets. Panel D presents the case to use nonlinear assets to construct the SDF and calculate the pricing errors for anomaly assets. There are four methods in our comparison: original agnostic method, shrinkage (ridge regression method), bias-adjust method, and bias-adjust and shrinkage (Kim's adjustment) method. See Section III.D.3 for the description of each method.

Panel A: Anomaly portfolios

			T-ratios		
Critical value	5%	10%	50%	90%	95%
Original	-2.00	-1.70	-0.05	1.58	1.86
Shrinkage	-2.60	-2.31	-0.63	1.07	1.36
Bias-adjust	-1.90	-1.61	-0.01	1.66	1.88
Bias-adjust+Shrink	-2.56	-2.25	-0.59	1.12	1.42

Panel B: Non-linear portfolios

Critical value	5%	10%	50%	90%	95%
Original	-2.00	-1.75	-0.05	1.57	1.84
Shrink	-2.60	-2.31	-0.58	1.06	1.38
Bias-adjust	-1.46	-1.37	-0.05	1.37	1.46
Bias-adjust+Shrink	-2.55	-2.25	-0.52	1.12	1.44

Panel C: Anomaly the SDF on non-linear portfolios

Critical value	5%	10%	50%	90%	95%
Original	-1.61	-1.27	0.44	1.95	2.27
Shrink	13.63	13.92	15.73	17.50	17.82
Bias-adjust	-1.89	-1.59	0.01	1.62	1.88
Bias-adjust+Shrink	12.28	12.56	14.35	16.10	16.41

Panel D: Non-linear the SDF on anomaly portfolios

Critical value	5%	10%	50%	90%	95%
Original	-12.29	-11.88	-10.26	-8.57	-8.20
Shrink	-67.67	-67.37	-65.35	-63.24	-62.76
Bias-adjust	-1.72	-1.54	0.11	1.67	1.88
Bias-adjust+Shrink	-64.65	-64.35	-62.35	-60.25	-59.84

Table 4: Application #1- Do nonlinear factors determine SDF?

This table presents the nonlinear characteristics that are selected by Lasso and that are tested to be significant under the OLS regression where stochastic discount factor ($\mathbf{m}$) is extracted from multiple assets. See Section IV for the construction of non-linear factors and the Appendix F for the full name of each non-linear factors. $\mathbf{m}$ is constructed from 1) 1,720 portfolios from 172 anomalies constructed similarly to Hou et al (2019), $\mathbf{m}$_anomaly (2) 87 individual stocks that survive for the whole sample period, $\mathbf{m}$_stock, and (3) 2,100 nonlinear portfolios we construct from 9 characteristics up to 3 polynomials, $\mathbf{m}$_nonlinear. The 9 characteristics are those that Freybeyger, Nieurehl and Webers (2019) find significant. (4) is 360 tree portfolios, $\mathbf{m}$_Tree, (5) is nonlinear portfolios and trees portfolios, $\mathbf{m}$_nonlinear and trees, (6) is all four portfolios, $\mathbf{m}$_Mixed. $\mathbf{m}$ is augmented by conditioning information ($\mathbf{Z}$) including dividend price ratio, default yield spread, risk-free rate, risk-free rate, term spread, corporate bond returns, stock variance, inflation, and net equity expansion available from Amit Goyal's website and Cay by Lettau and Ludvigson. We denote this augmented $\mathbf{m}$ as $\widehat{\mathbf{m}}(\mathbf{Z})$. We apply ridge regression to $\mathbf{m}$ and present the conditioning the SDF ($\widehat{\mathbf{m}}(\mathbf{Z})$) that has shrinkage term, $\mathrm{s(N)}\mathbf{I}$, of zero and at 100^{th} percentile.

$$\widehat{\mathbf{m}}(\mathbf{Z}) = \mathrm{T}(\bar{\mathbf{R}}'\bar{\mathbf{R}} + \mathrm{s(N)}\mathbf{I})^{-1}(\bar{\mathbf{R}}'\mathbf{1}_{\mathbf{NM}})$$

We perform three steps. First we regress each $\widehat{\mathbf{m}}(\mathbf{Z})$ on six polynomials of market factors where market factor is market rate of return subtracted by risk-free rate. Second, we use the residual as a dependent variable, apply Lasso to select nonlinear factors, and present them in the table. Third, we perform an OLS regression of the residuals of $\mathbf{m}$ on the selected nonlinear factors. ***, **, and * present 1%, 5% and 10% significance, respectively.

Nonlinear factor	$\boldsymbol{m}$_anomaly			$\boldsymbol{m}$ _stock			$\boldsymbol{m}$_nonlinear			$\boldsymbol{m}$_Tree			$\boldsymbol{m}$_Mixed			$\boldsymbol{m}$_nonlinear and trees	
	0^{th}	100^{th}		0^{th}	100^{th}		0^{th}	100^{th}		0^{th}	100^{th}		0^{th}	100^{th}		0^{th}	100^{th}
X11							1.180	0.927									
X13								-1.089									
X14								3.139***									
X16							-0.721	-0.051			-3.849***						
X17								1.089									
X18							3.200***	2.987***			1.631						0.372
X21								-2.460**									
X22											0.542						
X26							-1.286	-2.932***									
X29							1.259	1.328									
X31											-0.574						
X32								0.539									
X33								-1.483									
X36							1.119	1.264									
X38							-0.808	-1.174									
X40								-0.800									

Nonlinear factors	$\boldsymbol{m}$_anomaly			$\boldsymbol{m}$ _stock			$\boldsymbol{m}$_nonlinear			$\boldsymbol{m}$_Tree			$\boldsymbol{m}$_Mixed			$\boldsymbol{m}$_nonlinear and trees	
	Shrinkage term																
	0th	100th		0th	100th		0th	100th		0th	100th		0th	100th		0th	100th
X41											0.232			1.281			
X47								1.336			1.399					1.335	1.361
X50							-0.336	-0.378			-2.177**			-2.054**			-2.831***
X57								-0.763									
X59								0.814									
X61							0.580	0.249									
X63								1.289									
X69								0.794									
X70							1.378	0.562									
X73							-1.173	-1.243									
X74								1.811*									
X75							0.413	1.074									
X77								0.912			-3.066***						
X78								-1.387									
X79											2.381**						
X82							0.859	1.613									
X83											1.018						
X84							1.762*	1.831*									
X85								-3.033***									
X86							-1.918*	-2.518**									
X87											-1.198						
X88								-0.992									
X90								-0.298									
X91							-1.475	-2.104**									
X92								-1.579									
X93							0.071	-0.549									
X94											-0.676			-1.503			
X95								0.626			0.821						
X96								-1.476									
X97								-1.399			-3.530***						
X98							2.067**	2.331**									
X100							1.499	0.547									
X102								-2.120**									
X104																	-1.093
X105							1.074	2.067**									

Nonlinear factors	$\boldsymbol{m}$_anomaly		$\boldsymbol{m}$ _stock		$\boldsymbol{m}$_nonlinear		$\boldsymbol{m}$_Tree		$\boldsymbol{m}$_Mixed		$\boldsymbol{m}$_nonlinear and trees	
	Shrinkage term											
	0^{th}	100^{th}	0^{th}	100^{th}	0^{th}	100^{th}	0^{th}	100^{th}	0^{th}	100^{th}	0^{th}	100^{th}
X106					-1.467	0.084					-1.437	
X107						-1.106		1.633				
X108						0.718						
X110					-0.567	-0.490		-2.268**		-1.254	-1.401	-1.723
X111					2.462**	1.943*		3.059***		2.042**	1.973*	2.585**
X112					-1.620	-2.385**						
X116					1.274	1.531						
X117						0.434						
X118						-0.285						
X119					2.697**	2.528**						
X122						-1.306						
X123						-1.266						
X125						3.189***						
X133						2.885***						
X134					1.035	0.579		1.389		1.688		1.674
X135						-1.094						
X137						-0.779						
X142						-0.165						
X145						0.913						
X146								-0.579				
X148					-3.101***	-3.427***						
X150						0.613						
X152						1.336		2.134**				
X154					-0.758	-1.280						
X159								-1.073				
X164						0.949						
X170						-3.295***						
X173					-1.224	-1.602						
X175						2.308**						
X181						0.731						
X183						1.040						
X185						1.346						
X186						1.117						
X187								-1.425				
X188						-1.407						

Nonlinear factors	$\boldsymbol{m}$_anomaly			$\boldsymbol{m}$ _stock			$\boldsymbol{m}$_nonlinear			$\boldsymbol{m}$_Tree			$\boldsymbol{m}$_Mixed			$\boldsymbol{m}$_nonlinear and trees	
	Shrinkage term																
	0th	100th		0th	100th		0th	100th		0th	100th		0th	100th		0th	100th
X189								-1.462									
X191								0.117			2.155**						
X193								-1.366						-1.289			
X195							-1.519	-2.957***									
X196								-1.395									
X198								0.848									
X199								1.225									
X202							-0.427	-0.111									
X203								-0.364			1.812*						
X204								-0.745									
X211								-1.241									
X215								0.956									
X218								-1.479									

Table 5: Application 2 – Do characteristics explain systematic risk?

This table examines whether a time-varying covariance between SDF and returns should also be a linear combination of characteristics. It shows the coefficient estimates and associated t-stat from regressing covariance between returns and the SDF ($\widehat{\mathbf{m}}$) on lagged firm characteristics. We estimate the SDF using the daily data of stock returns over the last three years and then calculate the covariance between the SDF with each stock returns to obtain $\mathrm{cov}(\widehat{\mathbf{m}}, \mathbf{R})$ in each month. We apply Elastic Net on 94 characteristics and use the selected ones to run panel predictive regression by using the past firm chars as predictors. Model (1) presents OLS and applies clustered covariance matrix estimation for panel data, estimation of sandwich covariances a la Newey-West (1987) and Driscoll and Kraay (1998) for panel data, clustered standard errors by stocks and counts the serial correlation among periods. Model 2 includes time fixed effect without inference HAC and HAC standard errors for 36 lags. ***, **, and * present 1%, 5% and 10% significance level. See Appendix E for the definition of each characteristics.

Coefficients:	Type	OLS			Time Fixed Effect Model		
		(1)			(2)		
		Estimate	T-stat	Sig.	Estimate	T-stat	Sig.
Intercept		-61.29	-6.930	***			
mve	Trading friction	-17.78	-1.468		-35.46	-4.138	***
beta	Trading friction	2.52	0.480		10.36	2.510	*
baspread	Trading friction	12.33	4.982	***	14.88	6.387	***
turn	Trading friction	2.11	0.552		4.58	1.429	
mve_ia	Trading friction	-1.06	-0.548		0.39	0.308	
std_dolvol	Trading friction	0.76	0.192		-2.04	-1.381	
std_turn	Trading friction	3.85	2.254	**	3.57	2.086	**
idiovol	Trading friction	-81.07	-6.176	***	-97.42	-8.930	***
ill	Trading friction	-22.81	-4.066	***	-19.57	-4.067	***
pricedelay	Trading friction	-1.18	-3.307	***	-1.06	-2.581	**
retvol	Trading friction	-11.73	-2.021	**	-15.72	-2.859	**
zerotrade	Trading friction	-8.19	-2.323	**	-12.87	-4.502	***
aeavol	Trading friction	0.31	0.353		0.80	1.407	
bm	Value/growth	2.41	0.867		3.52	1.565	
ep	Value/growth	-16.82	-3.130	***	-16.45	-3.340	***
dy	Value/growth	-4.21	-1.632		0.45	0.185	
sgr	Value/growth	-1.36	-1.061		-0.55	-0.457	
sp	Value/growth	-3.58	-1.285		-2.07	-0.816	
cfp_ia	Value/growth	3.30	0.698		-2.16	-0.892	
bm_ia	Value/growth	-4.62	-1.169		0.56	0.215	
cfp	Value/growth	-2.44	-2.741	***	-1.18	-1.480	
egr	Value/growth	-2.45	-2.818	***	-2.30	-2.534	**
spi	Value/growth	-4.93	-2.225	**	-4.55	-2.255	**
cashdebt	Value/growth	-1.44	-0.737		-0.71	-0.420	
sue	Momentum	-0.89	-0.533		-1.59	-1.090	
mom12m	Momentum	-53.92	-10.993	***	-51.34	-9.982	***
mom1m	Momentum	-20.45	-10.875	***	-19.04	-10.030	***
mom36m	Momentum	-66.74	-9.279	***	-62.54	-8.836	***
nincr	Momentum	3.20	3.162	***	1.84	1.986	**
indmom	Momentum	-5.90	-1.683	*	-10.81	-3.330	***
ear	Momentum	-2.53	-2.711	***	-3.08	-3.700	***

Coefficients:	Type	OLS				Time Fixed Effect Model		
		(1)				(2)		
		Estimate	T-stat	Sig.		Estimate	T-stat	Sig.
rsup	Momentum	-2.39	-1.774	*		-1.49	-1.216	
chtx	Momentum	-0.21	-0.342			-0.24	-0.418	
chmom	Momentum	-4.89	-1.769	*		-2.29	-1.141	
lev	Profitability	-0.80	-0.241			1.16	0.338	
ps	Profitability	-2.83	-1.407			-2.06	-1.198	
chpmia	Profitability	1.79	2.157	*		1.96	2.143	**
chatoia	Profitability	-4.27	-3.816	***		-3.71	-3.903	***
gma	Profitability	9.05	3.532	***		6.51	3.275	***
roe	Profitability	0.36	0.200			-0.86	-0.629	
roeq	Profitability	-1.33	-1.492			-1.50	-1.520	
cashpr	Profitability	-0.21	-0.173			0.41	0.454	
roaq	Profitability	-7.37	-2.806	***		-6.52	-2.353	**
tb	Profitability	-0.60	-0.660			-0.24	-0.386	
operprof	Profitability	0.98	1.054			0.69	0.826	
ms	Profitability	-0.33	-0.133			-2.01	-1.215	
roic	Profitability	-4.11	-1.849	*		-4.95	-2.658	***
currat	Intangibles	0.79	0.861			-0.59	-0.699	
pchcurrat	Intangibles	-2.67	-1.629			-2.53	-1.555	
acc	Intangibles	1.32	0.353			4.46	2.634	**
pchsale_pchin	Intangibles	-1.56	-1.669	*		0.55	0.871	
pchsale_pchre	Intangibles	1.65	1.233			1.40	1.318	
pchgm_pchsale	Intangibles	-2.38	-3.083	***		-2.22	-3.175	***
pchsale_pchxs	Intangibles	2.27	1.681	*		2.09	1.683	*
roavol	Intangibles	1.99	0.781			1.94	0.872	
nanalyst	Intangibles	7.69	1.381			14.65	3.803	***
age	Intangibles	-2.53	-1.423			-3.17	-3.372	***
herf	Intangibles	-4.78	-2.298	**		-1.66	-0.980	
chnanalyst	Intangibles	-2.78	-3.624	***		-3.98	-6.627	***
tang	Intangibles	-3.54	-2.284	**		-4.00	-3.403	***
hire	Intangibles	-2.95	-1.654	*		-3.12	-2.726	**
absacc	Intangibles	2.07	0.949			2.81	2.733	**
stdcf	Intangibles	0.64	0.432			1.69	1.199	
pctacc	Intangibles	-3.31	-2.354	**		-2.56	-2.404	**
cash	Intangibles	1.02	0.383			2.73	1.153	
orgcap	Intangibles	-14.05	-4.917	***		-6.36	-3.173	***
salecash	Intangibles	2.00	2.747	***		0.44	1.150	
salerec	Intangibles	2.82	2.533	**		2.04	2.261	**
saleinv	Intangibles	-0.10	-0.105			-1.75	-2.220	**
pchsaleinv	Intangibles	0.00	0.003			-0.34	-0.955	

Coefficients	Type	OLS				Time Fixed Effect Model		
		(1)				(2)		
		Estimate	T-stat	Sig.		Estimate	T-stat	Sig,
grltnoa	Investment	-2.51	-2.019	**		-1.44	-1.456	
cinvest	Investment	-0.86	-1.914	*		-0.54	-1.525	
grcapx	Investment	-2.20	-1.483			-1.42	-1.213	
agr	Investment	-6.68	-3.758	***		-7.50	-5.540	***
pchcapx_ia	Investment	-0.11	-0.095			-0.06	-0.078	
pchdepr	Investment	4.31	2.702	***		4.12	3.096	**
chinv	Investment	-6.53	-3.778	***		-5.48	-4.170	***
chcsho	Investment	2.37	1.436			1.80	1.390	
invest	Investment	8.85	3.465	***		10.17	5.371	***
depr	Investment	0.96	0.507			0.50	0.390	
R^2			27.55%				27.80%	
Adj. R^2			27.54%				27.77%	

Table 6: Application 3 – Market integration

This table presents coefficient estimates and associated t-stat from regressing the SDF constructed from one asset class on the SDF constructed from another asset class. SDFs are constructed from individual stocks, bonds, commodities, currencies as well as the anomaly, and non-linear portfolios augmented by conditioning information on monthly frequency. Panel A presents the estimated coefficients from regressing the SDF constructed from one type of assets (stocks, bonds, commodities and currencies) on the SDF constructed from another types of portfolios. Panel B presents the results for the similar regression, where the SDF is constructed from (1) 1,720 portfolios from 172 anomalies constructed similarly to Hou et al (2019), (2) 87 individual stocks that survive for 1973-2016, and (3) 2,190 nonlinear portfolios we constructed from 9 characteristics up to 3 polynomials. The 9 characteristics are those that Freybeyger, Nieurehl and Webers (2019) find significant. (4) is 360 tree portfolios. Augmented $\widehat{\mathbf{m}}(\mathbf{Z})$ is augmented by conditioning information (Z) including dividend price ratio, default yield spread, risk-free rate, term spread, corporate bond returns, stock variance, inflation, and net equity expansion provided by Amit Goyal's website and Cay by Lettau and Ludvigson. We denote this augmented $\mathbf{m}$ as $\widehat{\mathbf{m}}(\mathbf{Z})$. We apply ridge regression and present the SDF ($\widehat{\mathbf{m}}(\mathbf{Z})$) that has shrinkage term, ($\mathrm{s}(\mathrm{N})\mathbf{I}$, with 0th and 100th percentile.

$$\widehat{\mathbf{m}}(\mathbf{Z}) = \mathrm{T}(\overline{\mathbf{R}}'\overline{\mathbf{R}} + \mathrm{s}(\mathrm{N})\mathbf{I})^{-1}(\overline{\mathbf{R}}'\mathbf{1}_{\mathrm{NM}})$$

Our sample period is from 1960 to 2016. We perform the test using annual and monthly frequency. "Year" presents the results after we convert the returns into annual frequency. "Month" presents the results from monthly returns. Decade" presents the results using monthly returns to construct the SDF for four different intervals: 1973 – 1983, 1984 – 1994, 1995-2005, 2006-2016.

Panel A: Market integration using augmented $\mathbf{m}$ or $\widehat{\mathbf{m}}(\mathbf{Z})$.

Dependent variable	$\widehat{\boldsymbol{m}}(\boldsymbol{Z})$ with 138 monthly observations									
	Shrinkage level					Shrinkage level				
	0^{th}		100^{th}	Independent variable		Dependent variable	0^{th}		100^{th}	Independent variable
Bond	-0.001		2.102	commodity		commodity	-7.771		0.000	Bond
	(-1.040)		(0.231)				(-1.040)		(0.231)	
Bond	-6.726**		0.065	REIT		REIT	-0.006**		0.110	Bond
	(-2.344)		(0.988)				(-2.344)		(0.988)	
Bond	-0.058		0.190*	Stocks		Stocks	-0.061		0.119*	Bond
	(-0.693)		(1.776)				(-0.693)		(1.776)	
Bond	0.061		-0.345**	Currency		Currency	0.017		-0.132**	Bond
	(0.374)		-(2.546)				(0.374)		(-2.546)	
Commodity	0.000		0.020	REIT		REIT	-50.220		0.006	Commodity
	(-0.403)		(0.123)				(-0.403)		(0.123)	
Commodity	-273.106**		0.000	Stocks		Stocks	0.000**		54.938	Commodity
	(-2.108)		(0.340)				(-2.108)		(0.340)	
Commodity	0.000		0.004	Currency		Currency	27.217		0.006	Commodity
	(0.356)		(0.059)				(0.356)		(0.059)	
REIT	-0.020		0.124	stocks		stocks	-0.034		0.129	REIT
	(-0.305)		(1.487)				(-0.305)		(1.487)	
REIT	0.173		-0.401**	Currency		Currency	0.034		-0.116**	REIT
	(0.891)		(-2.574)				(0.891)		(-2.574)	
Stock	0.000		0.039	Currency		Currency	-122.037		0.006	Stock
	(-0.511)		(0.185)				(-0.511)		(0.185)	

Panel B: Market integration using augmented $\mathbf{m}$ or $\hat{\mathbf{m}}(\mathbf{Z})$ of different portfolios

	Year											
Shrinkage level		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ anomaly		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ anomaly		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ anomaly		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ nonlinear		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ nonlinear		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ tree
0th	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ nonlinear	0.435***	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ stock	0.015	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ tree	0.664**	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ stock	-0.034	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ tree	1.498***	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ stock	0.058*
		(3.980)		(0.226)		(2.042)		(-0.424)		(4.382)		(1.995)
20th		0.433***		0.014		0.663**		-0.034		1.498***		0.057*
		(3.976)		(0.215)		(2.065)		(-0.428)		(4.431)		(1.953)
40th		0.430***		0.013		0.662**		-0.034		1.500***		0.056*
		(3.973)		(0.204)		(2.087)		(-0.432)		(4.478)		(1.913)
60th		0.428***		0.012		0.661**		-0.035		1.501***		0.055*
		(3.969)		(0.193)		(2.107)		(-0.436)		(4.523)		(1.875)
80th		0.426***		0.012		0.661**		-0.035		1.503***		0.054*
		(3.965)		(0.182)		(2.127)		(-0.440)		(4.566)		(1.838)
100th		0.424***		0.011		0.660***		-0.035		1.506***		0.053*
		(3.962)		(0.172)		(2.146)		(-0.444)		(4.607)		(1.803)

Shrinkage level		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ nonlinear		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ stock		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ tree		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ stock		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ tree		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ stock
0th	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ anomaly	0.641***	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ anomaly	0.083	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ anomaly	0.139**	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ nonlinear	-0.128	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ nonlinear	0.213***	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ tree	1.527*
		(3.980)		(0.226)		(2.042)		(-0.424)		(4.382)		(1.995)
20th		0.643***		0.080		0.142**		-0.130		0.216***		1.494*
		(3.976)		(0.215)		(2.065)		(-0.428)		(4.431)		(1.953)
40th		0.646***		0.076		0.145**		-0.131		0.219***		1.463*
		(3.973)		(0.204)		(2.087)		(-0.432)		(4.478)		(1.913)
60th		0.648***		0.073		0.148**		-0.133		0.222***		1.434*
		(3.969)		(0.193)		(2.107)		(-0.436)		(4.523)		(1.875)
80th		0.650***		0.069		0.150**		-0.135		0.224***		1.406*
		(3.965)		(0.182)		(2.127)		(-0.440)		(4.566)		(1.838)
100th		0.653***		0.066		0.153**		-0.137		0.227***		1.380*
		(3.962)		(0.172)		(2.146)		(-0.444)		(4.607)		(1.803)

Monthly												
Shrinkage level		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ anomaly		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ anomaly		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ anomaly		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ nonlinear		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ nonlinear		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ tree
0^{th}	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ nonlinear	0.411***	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ stocks	-0.001	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ tree	0.045	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ stocks	-0.018	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ tree	0.035**	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ stock	-0.099
		(4.162)		(-0.020)		(1.444)		(-0.659)		(2.612)		(-1.104)
20^{th}		0.461***		0.742*		0.378***		0.254		0.185***		0.581**
		(4.980)		(1.792)		(6.015)		(1.328)		(6.415)		(2.091)
40^{th}		0.486***		0.602		0.411***		0.277		0.207***		0.434
		(5.375)		(1.528)		(6.724		(1.496)		(7.244)		(1.608
60^{th}		0.501***		0.528		0.425***		0.282		0.217***		0.359
		(5.606)		(1.411)		(7.162)		(1.587)		(7.778)		(1.362)
80^{th}		0.511***		0.484		0.432***		0.282		0.223***		0.315
		(5.754)		(1.346)		(7.472)		(1.647)		(8.159)		(1.220)
100^{th}		0.517***		0.453		0.436***		0.280		0.226***		0.287
		(5.853)		(1.305)		(7.706)		(1.690)		(8.448)		(1.131)

Shrinkage level		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ nonlinear		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ stocks		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ tree		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ stocks		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ tree		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ stocks
0^{th}	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ anomaly	0.078***	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ anomaly	-0.001	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ anomaly	0.088	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ nonlinear	-0.045	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ nonlinear	0.366**	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ tree	-0.023
		(4.162)		(-0.020)		(1.444)		(-0.659)		(2.612)		(-1.104)
20^{th}		0.098***		0.008*		0.171***		0.013		0.393***		0.014**
		(4.980)		(1.792)		(6.015)		(1.328)		(6.415)		(2.091)
40^{th}		0.107***		0.007		0.193***		0.015		0.440***		0.011
		(5.375)		(1.528)		(6.724)		(1.496)		(7.244)		(1.608)
60^{th}		0.113***		0.007		0.210***		0.017		0.476***		0.010
		(5.606)		(1.411)		(7.162)		(1.587)		(7.778)		(1.362)
80^{th}		0.116***		0.007		0.223***		0.016		0.505		0.009
		(5.754)		(1.346)		(7.472)		(1.562)		(8.159)		(1.220)
100^{th}		0.118***		0.007		0.233***		0.019		0.529		0.008
		(5.853)		(1.305)		(7.706)		(1.690)		(8.448)		(1.131)

Decade											
Shrinkage level	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ anomaly		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ Anomaly		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ anomaly		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ nonlinear		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ nonlinear		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ tree
$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ nonlinear	0.383***	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ stocks	0.514***	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ tree	0.073***	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ stocks	0.010	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ tree	0.036***	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ stock	-0.600
	(5.307)		(2.792)		(3.384)		(0.091)		(2.831)		(-1.615)
	0.402***		0.492***		0.134***		0.017		0.091***		-0.391
	(5.530)		(2.895)		(4.497)		(0.172)		(5.283)		(-1.591)
	0.414***		0.474***		0.151***		0.022		0.109***		-0.280
	(5.686)		(2.978)		(4.826)		(0.232)		(6.072)		(-1.280)
	0.423***		0.458***		0.160***		0.025		0.119***		-0.215
	(5.808)		(3.047)		(5.037)		(0.280)		(6.565)		(-1.056)
	0.430***		0.445***		0.165***		0.027		0.126***		-0.173
	(5.910)		(3.106)		(5.195)		(0.319)		(6.934)		(-0.892)
	0.436***		0.433***		0.168***		0.028		0.130***		-0.143
	(6.000)		(3.157)		(5.321)		(0.351)		(7.232)		(-0.769)

Shrinkage level	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ nonlinear		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ stocks		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ tree		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ stocks		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ tree		$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ Stocks
$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ anomaly	0.133***	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ anomaly	0.028***	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ anomaly	0.294***	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ nonlinear	0.002	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ nonlinear	0.419***	$\hat{\boldsymbol{m}}(\boldsymbol{Z})$_ tree	-0.008
	(5.307)		(2.792)		(3.384)		(0.091)		(2.831)		(-1.615)
	0.137***		0.032***		0.277***		0.003		0.553***		-0.012
	(5.530)		(2.895)		(4.497)		(0.172)		(5.283)		(-1.591)
	0.140***		0.035***		0.281***		0.005		0.600***		-0.011
	(5.686)		(2.978)		(4.826)		(0.232)		(6.072)		(-1.280)
	0.143***		0.038***		0.288***		0.006		0.636***		-0.010
	(5.808)		(3.047)		(5.037)		(0.280)		(6.565)		(-1.056)
	0.145***		0.041***		0.296***		0.007		0.667***		-0.009
	(5.910)		(3.106)		(5.195)		(0.319)		(6.934)		(-0.892)
	0.147***		0.043***		0.304***		0.008		0.696***		-0.008
	(6.000)		(3.157)		(5.321)		(0.351)		(7.232)		(-0.769)

References

Adrian, Tobias, Richard K. Crump, and Emanuel Moench. "Pricing the Term Structure with Linear Regressions," *Journal of Financial Economics* 110(1) (2013), 110-138.

Aït-Sahalia, Yacine and Andrew W. Lo. "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *Journal of Finance* 53 (1998), 499-547.

Aït-Sahalia, Yacine and Andrew W. Lo. "Nonparametric Risk Management and Implied Risk Aversion," *Journal of Econometrics* 94 (2000), 9-51.

Alvareza, Fernando, and Urban J. Jermann. 2005. "Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth." *Econometrica*, 73(6): 1977-2016.

Ang, Andrew, Robert J. Hodrick, Yuhang Zing, and Xiaoyan Zhang. 2006. "The Cross-Section of Volatility and Expected Returns." *Journal of Finance*, 61(1): 259-299.

Araujo, Fabio, Joao Victor Issler, and Marcelo Fernandes. "Estimating the Stochastic Discount Factor without a Utility Function," Working Paper, Graduate School of Economics EPGE Getulio Vargas Foundation, March 2005.

Araujo, Fabio and Joao Victor Issler. "A Stochastic Discount Factor Approach to Asset Pricing Using Panel Data Asymptotics," Graduate School of Economics EPGE Getulio Vargas Foundation, 2011.

Back, Kerry, *Asset Pricing and Portfolio Choice Theory*, 2nd edition, 2017 Oxford University Press.

Bai, Jennie, Tarun Bali, and Quan Wen. "Common risk factors in the. cross-section of corporate bond returns," *Journal of Financial Economics* 131(3), 619-642.

Bessembinder, Hendrick, Michael J Cooper and Feng Zhang. "Characteristic-Based Benchmark Returns and Corporate Events" *Review of Financial Studies* 32(1): 75-125

Bossaerts, Peter L., *The Paradox of Asset Pricing*, Princeton, Princeton University Press, 2002.

Britten-Jones, Mark, "The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights," *Journal of Finance* 54 (1999), 655-671.

Brown, Morton B. and Alan B. Forsythe. "Robust Tests for Equality of Variances," *Journal of the American Statistical Association* 69 (1974), 364–367.

Brown, Stephen J. and Mark I. Weinstein. "A New Approach to Testing Asset Pricing Models: The Bilinear Paradigm," *Journal of Finance* 38(3) (1983), 711-43.

Bryzgalova, S., Marcus Pelger, and Jason Zhu. "Forest through the trees: Buidling Cross-Sections of Stock Returns," Stanford working paper (2020).

Campbell, John Y. "Intertemporal Asset Pricing without Consumption Data," *American Economic Review* 83(1993), 487-512.

Campbell, John Y. "Empirical Asset Pricing: Eugene Fama, Lars Peter Hansen, and Robert Shiller," *the Scandinavian Journal of Economics* 116 (2014), 593-634.

Chapman, David A. "Approximating the Asset Pricing Kernel," *Journal of Finance* 52 (1997), 1383-1410.

Chen, Xiaohong and Sydney Ludvigson,. "Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models," *Journal of Applied Econometrics* 24 (2009), 1057-1093.

Chen, Zhiwu, and Peter J. Knez, "A Pricing Operator-Based Testing Foundation for a Class of Factor Pricing Models," *Mathematical Finance* 4 (1994), 121-141.

Chen, Zhiwu, and Peter J. Knez. "Measurement of Market Integration and Arbitrage," *Review of Financial Studies* 8 (1995), 287-325.

Chen, Zhiwu, and Peter J. Knez. "Portfolio Performance Measurement: Theory and Applications," *Review of Financial Studies* 9 (1996), 511-555.

Cochrane, John H. "A Cross-Sectional Test of an Investment-Based Asset Pricing Model," *Journal of Political Economy* 104 (1996), 241-255.

Cochrane, John H. "Asset Pricing," Revised Edition, Princeton: Princeton University Press (2005).

Cochrane, John H., "Financial Markets and the Real Economy," *Foundations and Trends(R) in Finance*, now publishers 1(1) 2005, 1-101.

Cochrane, John H., and Christopher L. Culp. "Equilibrium Asset Pricing and Discount Factors: Overview and Implications for Derivatives Valuation and Risk Management," ch. 5, in *Modern Risk Management: A History.* Peter Field, ed. London: Risk Books (2003).

Cochrane, John H., and Lars Peter Hansen. "Asset Pricing Explorations for Macroeconomics," Working Paper No. 4408, National Bureau of Economic Research (June 1992.)

Da, Zhi and Hayong Yun. "Industrial Electricity Usage and Stock Returns," *Journal of Financial and Quantitative Analysis* 52 (2017), 37-69.

DeSantis, Gerard. "Volatility Bounds for Stochastic Discount Factors: Tests and Implications from International Financial Markets," Ph.D. dissertation, Department of Economics, University of Chicago (1993).

Dew-Becker, Ian, and Stefano Giglio, 2016. "Asset Pricing in the Frequency Domain: Theory and Empirics." *Review of Financial Studies* 29(8): 2029-2068.

Driscoll, John, Aart Kraay, 1998. "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data," *Review of Economics and Statistics* 80, 549-560.

Engle, Robert F., and Sharon Kozicki. "Testing for Common Features," *Journal of Business and Economic Statistics* 11 (1993), 369-80.

Fama, Eugene, Kenneth French. 2018. "Choosing Factors," *Journal of Financial Economics* 128(2), 234-252.

Ferson, Wayne. "Theory and Empirical Testing of Asset Pricing Models," in Jarrow, R. A., V. Maksimovic, and W. T. Ziemba (eds.), *Handbooks in OR and MS,* Vol. 9, Finance, Amsterdam: Elsevier Science, 1995.

Ferson, Wayne and Andrew F. Siegel. "The Efficient Use of Conditioning Information in Portfolios." *Journal of Finance* 56(3), 967-982.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber. 2020. "Dissecting Characteristics Nonparametrically," *Review of Financial Studies* 33(5), 2326-2377.

Green, Jeremiah, John R.M. Hand, and X. Frank Zhang. 2017. "The Characteristics that Provide Independent Information above Average U.S. Monthly Stock Returns," *Review of Financial Studies* 30(12), 4389-4436.

Han, Yufeng, Ai He, David Rapach, and Guofu Zhou. 2020. "Firm Characteristics and Expected Stock Returns." Washington University working paper.

Hansen, Lars Peter, and Richard, Scott F, 1987. "The Role of Conditioning Information in Deducing Testable," *Econometrica,* Econometric Society, 55(3), 587-613.

Hansen, Lars Peter, and Jose A. Scheinkman. 2009. "Long-term Risk: An Operator Approach." *Econometrica*, 77(1): 177-234.

Hansen, Lars Peter, and Ravi Jagannathan. "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy* 99 (1991), 225-262.

Hausman, J.A. "Specification Tests in Econometrics," *Econometrica* 46 (6) (1978), 1251-1271.

He, Jia, Lilian Ng and Chu Zhang. "Asset Pricing Specification Errors and Performance Evaluation," *European Finance Review* 3 (1999), 205-232.

Hotelling, Harold. "The Generalization of Student's Ratio," *Annals of Mathematical Statistics* 2 (3) (1931), 360–378.

Hou, Kewei, Chen Xue and Lu Zhang. "Replicating Anomalies," *Review of Financial Studies* 33(5), 2019-2033.

Jegadeesh, Narasimhan, Joonki Noh, Kuntara Pukthuanthong, Richard Roll, and Junbo Wang. "Empirical tests of asset pricing models with individual assets: Resolving the errors-in-variables bias in risk premium estimation" *Journal of Financial Economics* 133(2), 273-298.

Jones, Christopher, Jinchi Lv, Kuntara Pukthuanthong, and Junbo Wang (2020). "A Generalized Machine Learning Framework for Linear Factor Model Test" Working paper.

Kelly, Bryan T., Seth Pruittt and Yinan Su, "Characteristics are covariances: A unified model of risk and return," *Journal of Financial Economics* 134(3), 501-524.

Kim, Soohun, "Properties of the Pukthuanthong-Roll Stochastic Discount Factor," Georgia Tech working paper, 2017.

Kim, Soohun, and Robert Korajczyk, "Large Sample Estimator of the Stochastic Discount Factor," Georgia Institute of Technology, Scheller College of Business working paper, 2021.

Kirby, Chris. "Firm Characteristics, Cross-Sectional Regression Estimates, and Asset Pricing Tests," Review of Asset Pricing Studies 10 (2020), 335-395.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh. "Interpreting Factor Models," *Journal of Finance* 73 (2017), 1183-1223.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh. "Shrinking the cross-section," *Journal of Finance Economics* 135 (2020), 271-292.

Lettau, Martin, and Sydney Ludvigson. "Consumption, Aggregate Wealth and Expected Stock Returns," *Journal of Finance* 56 (2001), 815-849.

Long, John. "The Numeraire Portfolio," *Journal of Financial Economics* 26(1) (1990), 29-69.

Martin, Ian. 2017. "What Is the Expected Return on the Market?" *Quarterly Journal of Economics*, 132(1): 367-433.

McCulloch, J. Huston, "The Risk-Neutral Measure and Option Pricing under Log-Stable Uncertainty," Ohio State University Economics Department working paper, (2003).

Mehra, Raj, and Edward Prescott. "The Equity Premium: A Puzzle," *Journal of Monetary Economics* 15 (1985), 145-161.

Newey, Whitney and Kenneth D. West. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55 (1987), 703-708.

Pelger, Markus and Martin Lettau. "Estimating Latent Asset-Pricing Factors," forthcoming *Journal of Econometrics*, 2019.

Polyanin, Andrei D., and Alexander V. Manzhirov, *Handbook of Integral Equations*, Boca Raton, CRC Press, 1998.

Roll, Richard. "Evidence on the Growth Optimum Model," *Journal of Finance* 28(3) (1973), 551-566.

Rosenberg, Joshua, and Robert F. Engle. "Empirical Pricing Kernels," *Journal of Financial Economics* 64 (2002), 341-372.

Ross, Stephen A. "The Arbitrage Theory Of Capital Asset Pricing," *Journal of Economic Theory* 13 (1976), 341–360.

Ross, Stephen A. "The Recovery Theorem." *Journal of Finance* 70(2): 615-648.

Shumway, Tyler, The Delisting Bias in CRSP Data, *Journal of Finance* 52 (1997), 327-340.

Shumway, Tyler and Vincent Warther, 1999, The Delisting Bias in CRSP's Nasdaq Data and Its Implications for the Size Effect, *Journal of Finance* 54 (1999), 2361-2379.

Singleton, Kenneth J. "Specification and Estimation of Intertemporal Asset Pricing Models," in Briedman, B., and F. Hahn (eds.), *Handbook of Monetary Economics*, Vol, 1, Amsterdam, Elsevier Science, 1990.

Singleton, Kenneth J. "*Empirical Dynamic Asset Pricing*," Princeton: Princeton University Press, 2006.

Snow, Karl N. "Diagnosing Asset Pricing Models Using the Distribution of Asset Returns." *Journal of Finance,* Papers and Proceedings 46 (1991), 955-983.

Theil, Henri. "*Applied Economic Forecasting*," Chicago: Rand McNally, 1966.

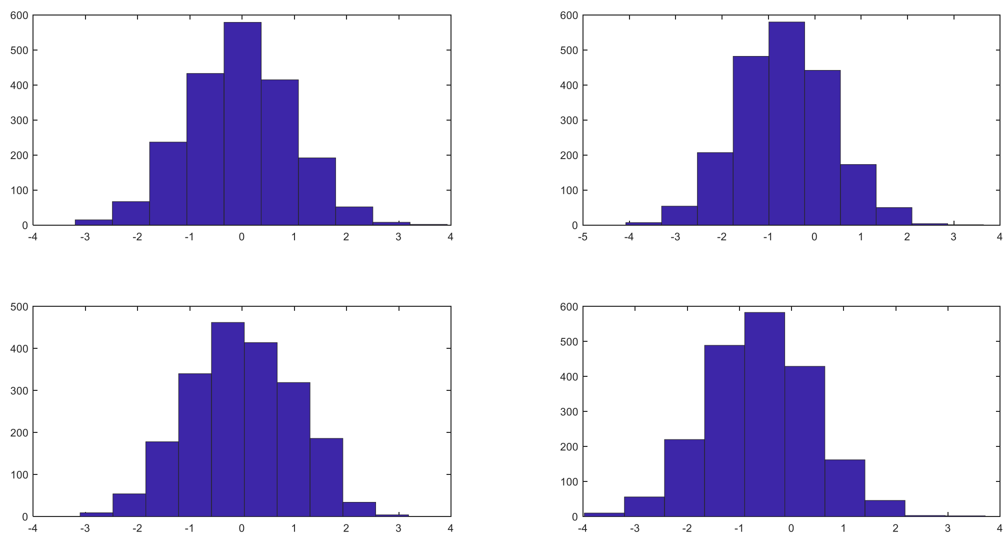
Welch, B. L. "On the Comparison of Several Mean Values: An Alternative Approach," *Biometrika* 38 (1951), 330-336.

Appendix A Figure I

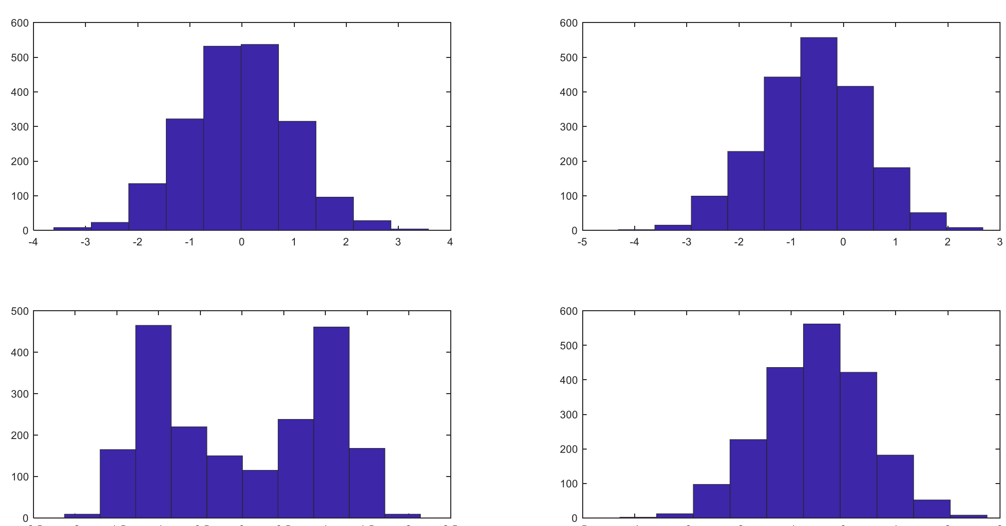
Distribution of T- E(MR)-1 in simulations

The figure contains plots of the T-ratios for pricing error $\frac{1}{T}(\sum_{t=1}^{T} m_t R_{it}) - 1$ in Table 3. There are four subplots in each figure; the top left represents the original agnostic method, the top right represents the ridge regression method, the bottom left represents the bias-adjusted method, and the bottom right represents both ridge and bias-adjust method. Four cases are considered: First, we use anomaly asset returns to construct the SDF and calculate the pricing errors with these assets. Second, we use non-linear asset returns to construct the SDF and calculate the pricing errors. Third, we use anomaly assets to construct the SDF and calculate the pricing errors for nonlinear assets. Four, we use nonlinear assets to construct the SDF and calculate the pricing errors for anomaly assets.

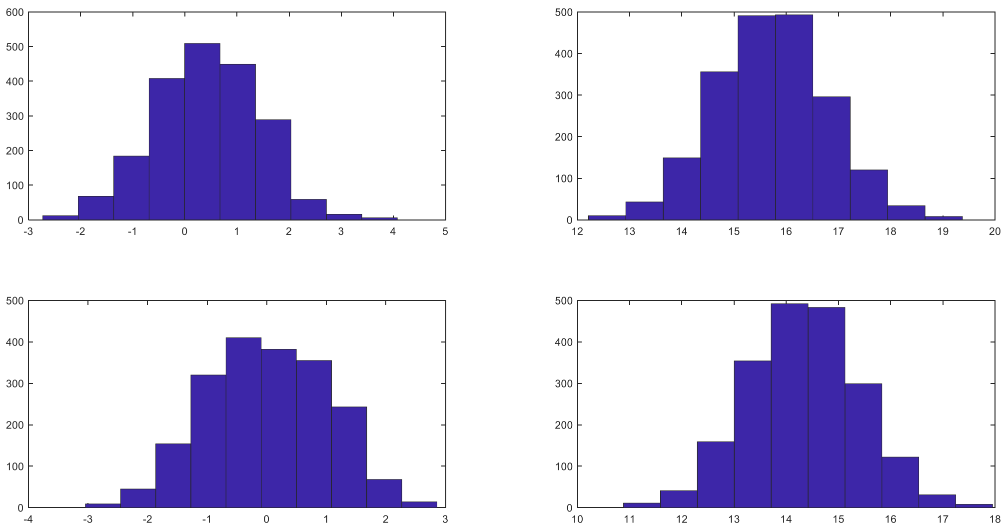
Anomaly portfolios



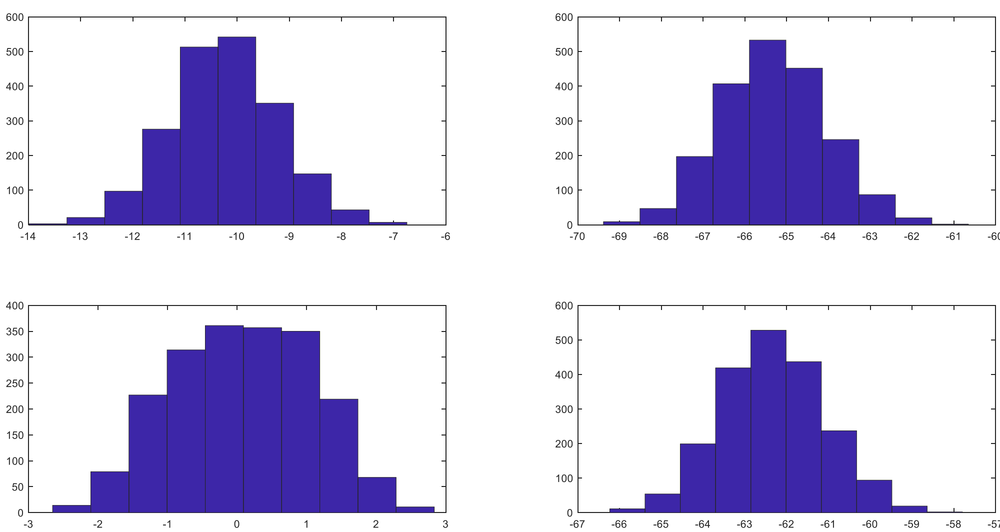
Non-linear portfolios



Anomaly the SDF on non-linear portfolios



Non-linear the SDF on anomaly portfolios



Appendix B
Proof of Theorem 2.1, from Kim (2017)

The pricing equation of $\frac{1}{T}\mathbf{Rm} = \mathbf{1}_N + \varepsilon$ implies that

$$\hat{\mathbf{m}} - \mathbf{m} = -T\left(\mathbf{R}'\ \mathbf{R}\right)^{-1}\mathbf{R}'\ \varepsilon.$$

Let $\lambda_{\max}(\mathbf{A}), \lambda_{\min}(\mathbf{A})$ and tr ($\mathbf{A}$) denote the maximum eigenvalue, the minimum eigenvalue, and the trace of a square matrix $\mathbf{A}$, respectively. Then, we have that

$$\frac{1}{T}(\hat{\mathbf{m}} - \mathbf{m})'\ (\hat{\mathbf{m}} - \mathbf{m}) = -T\varepsilon'\ \mathbf{R}\left(\mathbf{R}'\ \mathbf{R}\right)^{-2}\mathbf{R}'\ \boldsymbol{\varepsilon} = T\mathrm{tr}(\mathbf{R}\left(\mathbf{R}'\ \mathbf{R}\right)^{-2}\mathbf{R}'\ \varepsilon\varepsilon'\ )$$
$$\leq T\lambda_{\max}\left(\mathbf{R}\left(\mathbf{R}'\ \mathbf{R}\right)^{-2}\mathbf{R}'\ \right)\ \mathrm{tr}\ (\varepsilon\varepsilon'\ )$$

$$= T\lambda_{\max}(\left(\left(\mathbf{R}'\ \mathbf{R}\right)^{-1}\right)\varepsilon'\ \varepsilon = \frac{T}{N}(\lambda_{\min}\left(\frac{\mathbf{R}'\ \mathbf{R}}{N}\right)\ (\varepsilon\varepsilon'\ )\left(\lambda_{\min}\left(\frac{\mathbf{R}'\ \mathbf{R}}{N}\right)\right)^{-1}\ \varepsilon\ '\ \ \varepsilon$$

where the second equality is from (A.1), the inequality is from (A.2) and the third equality is from (A.1) and the property of eigenvalue (Greene (2008) page 270.)

In addition, Assumption 1 (i), (ii), and (iii) imply $\frac{T}{N} \to 0$, $\lambda_{\min}\left(\frac{\mathbf{R}'\ \mathbf{R}}{N}\right) > c$, the boundedness of $\varepsilon'\ \varepsilon$, respectively, which, in conjunction with (B.1), yield

$$\frac{1}{T}(\hat{\mathbf{m}} - \mathbf{m})'\ (\hat{\mathbf{m}} - \mathbf{m}) \to 0$$

This completes the proof of theorem.

Supplementary Proofs:

Let $\lambda_{\max}(\mathbf{A})$ and tr ($\mathbf{A}$) denote the maximum eigenvalue and the trace of a square matrix A, respectively. The following properties of eigenvalues and trace operator, vectorize operator are useful for the proof of the lemmas:

(A.1) Consider a ($L\times M$) matrix of $\mathbf{A}$ and ($M\times L$) matrix of $\mathbf{B}$. Then, it holds that

tr ($\mathbf{AB}$) = tr ($\mathbf{BA}$).

(A.2) Consider ($L\times L$) positive semidefinite matrices of $\mathbf{A}$, $\mathbf{B}$. Then, it holds

tr ($\mathbf{AB}$) $\leq \lambda_{\max}(\mathbf{A})\ \mathrm{tr}(\mathbf{B})$

Appendix C Summary statistics for individual stock returns

This table presents descriptive statistics of the individual stock returns, and characteristics we apply in this study. Our sample period is from 1962 to 2018. Panel A presents the statistics of firm stock returns, including year, number of observations (# obs), number of stocks (# stocks), mean median, standard deviation (stdev), skewness, kurtosis, min and max of returns.

year	# obs	# stocks	mean	median	stdev	Skew ness	kurtosis	min	max
1962	5	1	-0.003	-0.044	0.077	1.321	0.786	-0.060	0.120
1963	1741	334	0.007	0.000	0.065	0.864	2.848	-0.201	0.380
1964	3433	413	0.017	0.011	0.066	2.672	34.045	-0.361	1.106
1965	3327	313	0.022	0.015	0.072	1.472	8.720	-0.224	0.787
1966	4289	426	-0.004	-0.012	0.084	0.902	3.736	-0.410	0.602
1967	8284	1022	0.036	0.017	0.112	1.755	8.154	-0.317	1.143
1968	13413	1184	0.029	0.015	0.113	1.401	6.226	-0.368	1.284
1969	15023	1355	-0.021	-0.026	0.108	0.711	2.996	-0.573	0.866
1970	17221	1545	-0.003	-0.006	0.137	0.555	2.738	-1	1.240
1971	19514	1737	0.017	0.002	0.121	1.121	4.413	-0.887	1.000
1972	20997	1849	0.005	-0.004	0.110	1.087	7.244	-1	1.222
1973	25139	2714	-0.032	-0.037	0.141	0.567	3.064	-1	1.265
1974	35396	3216	-0.022	-0.035	0.154	1.834	16.816	-1	3.167
1975	37163	3230	0.049	0.013	0.178	2.358	15.083	-1	3.600
1976	36616	3224	0.037	0.012	0.142	1.994	12.297	-1	2.455
1977	36937	3208	0.016	0.000	0.118	4.332	95.630	-1	4.600
1978	36003	3165	0.021	0.012	0.143	1.510	17.649	-1	3.185
1979	36169	3296	0.030	0.013	0.131	1.974	15.909	-1	2.364
1980	38540	3530	0.031	0.015	0.157	2.209	22.061	-1	2.923
1981	39934	3596	0.004	0.000	0.136	1.905	19.609	-1	2.750
1982	41697	3836	0.023	0.000	0.163	2.884	35.633	-0.937	3.955
1983	43461	3896	0.029	0.010	0.170	3.469	40.525	-1	4.000
1984	44718	4208	-0.008	-0.005	0.143	1.922	29.865	-1	3.267
1985	46889	4304	0.019	0.003	0.218	66.773	9546.964	-1	31.764
1986	46475	4274	0.005	0.000	0.172	7.145	312.761	-1	10.200
1987	46944	4388	-0.003	0.000	0.193	3.215	89.871	-1	8.071
1988	49028	4583	0.017	0.000	0.164	4.423	98.139	-1	6.167
1989	48966	4456	0.011	0.000	0.161	4.055	88.530	-1	6.385
1990	48058	4353	-0.017	-0.016	0.195	7.449	288.474	-1	11.000
1991	47597	4296	0.038	0.010	0.230	10.021	394.427	-1	14.000
1992	47518	4349	0.021	0.000	0.205	7.289	197.868	-1	10.000
1993	49704	4527	0.018	0.000	0.170	5.377	133.287	-1	7.480
1994	55815	5448	-0.002	-0.005	0.154	11.954	816.860	-1	12.500
1995	62743	5794	0.024	0.013	0.160	3.339	51.066	-0.992	4.667
1996	64675	5977	0.016	0.005	0.166	3.994	93.064	-0.944	7.000
1997	67007	6295	0.020	0.010	0.173	3.560	80.276	-0.990	6.077
1998	67413	6273	0.002	-0.006	0.223	9.316	381.403	-0.974	12.667

year	# obs	# stocks	mean	median	stdev	Skew ness	kurtosis	min	max
1999	64586	6006	0.024	-0.004	0.232	5.167	96.251	-0.984	9.500
2000	62001	5841	-0.001	-0.011	0.256	3.848	72.772	-1	10.344
2001	59706	5590	0.026	0.006	0.266	3.921	54.947	-1	8.667
2002	56636	5053	-0.008	-0.009	0.212	2.669	36.205	-1	5.640
2003	52840	4710	0.048	0.024	0.186	3.745	49.567	-1	5.179
2004	50300	4463	0.018	0.010	0.149	5.980	175.177	-1	6.908
2005	49277	4412	0.004	0.000	0.131	2.075	31.445	-1	3.303
2006	48744	4364	0.014	0.007	0.128	2.687	45.132	-0.947	4.023
2007	47751	4318	-0.004	-0.005	0.131	3.825	124.273	-0.981	5.801
2008	46884	4232	-0.044	-0.035	0.199	1.490	24.403	-1	4.900
2009	45551	4013	0.046	0.020	0.274	10.771	427.327	-0.998	15.774
2010	43574	3829	0.023	0.015	0.157	3.312	72.655	-0.965	6.107
2011	41992	3714	-0.005	-0.008	0.147	2.690	44.459	-1	3.994
2012	40797	3587	0.016	0.010	0.145	4.732	103.288	-0.957	5.210
2013	39728	3491	0.032	0.023	0.131	3.586	59.621	-0.936	3.746
2014	39588	3510	0.005	0.004	0.135	6.295	192.536	-0.966	6.080
2015	40390	3643	-0.005	-0.005	0.154	11.459	609.285	-0.954	9.564
2016	40523	3602	0.016	0.010	0.170	8.124	273.583	-0.953	7.635
2017	39808	3509	0.013	0.007	0.142	3.251	47.289	-0.962	3.273
2018	39346	3483	-0.011	-0.008	0.154	4.699	129.833	-0.994	6.428

Appendix D: Descriptive statistics of anomalies

This table presents descriptive statistics of the candidate factors we apply in this study. Our sample period is from 1962 to 2018. The candidate factors are constructed in a similar vein as Hou et al. (2018). We use the same screening criteria, delisting procedure, and period similar to what they do. The first column presents the identification numbers and names of the candidate factors according to their papers. The second column presents the paper introducing these candidate factors. The last four columns present the number of observations, the mean of candidate factors, t-stat testing the mean is statistically different from zero, and the standard deviation of candidate factors. All candidate factors are based on 1-month calculation, and these portfolios are value-weighted returns. ***, **, and * present 1%, 5%, and 10% significance level.

Candidate factors	Reference papers	# obs	mean	t-stat	std.dev
	A. Momentum				
A.1.1 Standardized unexpected earnings	Foster, Olsen, and Shevlin (1984)	534	0.009	4.966***	0.04
A.1.2 Cumulative abnormal returns around earnings announcement dates	Chan, Jegadeesh, and Lakonishok (1996)	521	0.016	8.571***	0.043
A.1.4 Price momentum, prior 6-month returns	Jegadeesh and Titman (1993)	534	0.01	3.126***	0.077
A.1.5 Price momentum, prior 11-month returns	Fama and French (1996)	534	0.014	4.244***	0.077
A.1.6 Industry momentum	Moskowitz and Grinblatt (1999)	534	0.569	2.227**	5.905
A.1.7 Revenue surprises	Jegadeesh and Livnat (2006)	534	0.002	1.371	0.037
A.1.10 The number of quarters with consecutive earnings increase	Barth, Elliott, and Finn (1999)	533	0.005	1.758*	0.071
A.1.11 52-week high	George and Hwang (2004)	529	-0.001	-0.181	0.068
A.1.12 Residual momentum, prior 6-month returns	Blitz, Huij, and Martens (2011)	534	0.003	1.398	0.056
A.1.13 Residual momentum, prior 11-month returns	Blitz, Huij, and Martens (2011)	534	0.010	3.887***	0.06
	B. Value versus growth				
B.2.1 Book-to-market equity	Rosenberg, Reid, and Lanstein (1985)	534	0.005	2.170**	0.05
B.2.2 Book-to-June-end market equity	Asness and Frazzini (2013)	534	0.005	2.327**	0.053
B.2.3 Quarterly book-to-market equity		534	0.018	6.875***	0.061
B.2.6 Assets-to-market	Rosenberg, Reid, and Lanstein (1985)	534	0.005	2.066**	0.056
B.2.8 Reversal.	De Bondt and Thaler (1985)	534	-0.004	-1.808*	0.056
B.2.9 Earnings-to-price	Basu (1983)	534	0.002	0.933	0.057
B.2.12 Cash flow-to-price	Lakonishok, Shleifer, and Vishny (1994)	534	0.000	0.123	0.048
B.2.14 Dividend yield	Litzenberger and Ramaswamy (1979)	534	0.002	1.008	0.039
B.2.16 Payout yield	Boudoukh et al. (2007)	529	0.005	2.587**	0.046
B.2.16 Net payout yield	Boudoukh et al. (2007)	529	0.005	2.506**	0.048
B.2.18 5-year sales growth rank	Lakonishok, Shleifer, and Vishny (1994)	534	-0.001	-0.715	0.044
B.2.19 Sales growth	Lakonishok, Shleifer, and Vishny (1994)	534	-0.002	-1.132	0.042

Candidate factors	Reference papers	# obs	mean	t-stat	std.dev
B.2.20 Enterprise multiple	Loughran and Wellman (2011)	534	-0.005	-2.000**	0.056
B.2.22 Sales-to-price	Barbee, Mukherji, and Raines (1996)	534	0.007	2.683**	0.058
B.2.26 Intangible return	Daniel and Titman (2006)	534	-0.009	-4.884***	0.044
B.2.30 Equity duration	Dechow, Sloan, and Soliman (2004)	534	-0.008	-3.180***	0.056
	C. Investment				
C.3.1 Abnormal corporate investment	Titman, Wei, and Xie (2004)	534	-0.003	-2.114**	0.031
C.3.2 Investment-to-assets	Cooper, Gulen, and Schill (2008)	534	0.002	4.031***	0.011
C.3.3 Quarterly investment-to-assets		522	-0.001	-0.672	0.026
C.3.4 Δ in PPE and inventory-to-assets	Lyandres, Sun, and Zhang (2008)	534	-0.004	-3.012***	0.033
C.3.5 Noa and dNoa, (changes in) net operating assets	Hirshleifer et al. (2004)	534	-0.006	-4.058***	0.032
C.3.6 Changes in long-term net operating assets.	Fairfield, Whisenant, and Yohn (2003)	534	-0.004	-3.005**	0.034
C.3.7 Investment growth	Xie (2008)	534	-0.004	-3.485***	0.028
C.3.8 2-year investment growth	Anderson and Garcia-Feijoo (2006)	534	-0.003	-1.930**	0.032
C.3.9 3-year investment growth	Anderson and Garcia-Feijoo (2006)	534	-0.002	-1.383	0.034
C.3.10 Net stock issues	Pontiff and Woodgate (2008)	534	-0.004	-3.554***	0.026
C.3.11 % change in investment relative to industry	Abarbanell and Bushee (1998)	534	-0.003	-2.336**	0.034
C.3.12 Composite equity issuance	Daniel and Titman (2006)	534	-0.002	-0.817	0.044
C.3.13 Composite debt issuance	Lyandres, Sun, and Zhang (2008)	534	-0.001	-0.425	0.038
C.3.14 Inventory growth	Belo and Lin (2011)	534	-0.003	-2.061**	0.033
C.3.15 Inventory changes	Thomas and Zhang (2002)	534	-0.004	-2.922***	0.031
C.3.16 Operating accruals	Sloan (1996)	534	-0.003	-2.169**	0.033
C.3.17 Total accruals	Richardson, Sloan, Soliman, and Tuna (2005)	534	-0.003	-1.962*	0.035
C.3.18 Δ in net noncash working capital, in current operating assets, and in current operating liabilities.	Richardson, Sloan, Soliman, and Tuna (2005)	534	-0.002	-1.118	0.036
C.3.19 Δ in noncurrent operating assets	Richardson, Sloan, Soliman, and Tuna (2005)	534	-0.005	-3.415***	0.032
C.3.19 Δ in noncurrent operating liabilities	Richardson, Sloan, Soliman, and Tuna (2005)	534	-0.001	-0.867	0.03
C.3.19 Δ in net noncurrent operating assets	Richardson, Sloan, Soliman, and Tuna (2005)	534	-0.004	-3.340***	0.03
C.3.20 Δ in book equity	Richardson, Sloan, Soliman, and Tuna (2005)	534	-0.001	-0.285	0.051
C.3.20 Δ in net financial assets	Richardson, Sloan, Soliman, and Tuna (2005)	534	0.002	2.040**	0.028
C.3.20 Δ in financial liabilities	Richardson, Sloan, Soliman, and Tuna (2005)	534	-0.001	-1.339	0.023
C.3.20 Δ in in long-term investments	Richardson, Sloan, Soliman, and Tuna (2005)	534	-0.001	-1.357	0.025
C.3.20 Δ in short-term investments	Richardson, Sloan, Soliman, and Tuna (2005)	534	0.000	0.393	0.024
C.3.21 Discretionary accruals computed from Nasdaq Index	Xie (2001)	516	-0.003	-1.936*	0.036

Candidate factors	Reference papers	# obs	mean	t-stat	std.dev
C.3.21 Discretionary accruals computed from NYSE and Amex	Xie (2001)	534	-0.002	-1.407	0.03
C.3.22 % operating accruals	Hafzalla, Lundholm, and Van Winkle (2011)	534	-0.004	-3.059***	0.031
C.3.23 % total accruals	Hafzalla, Lundholm, and Van Winkle (2011)	534	-0.002	-1.417	0.026
C.3.24 % discretionary accruals	Hafzalla, Lundholm, and Van Winkle (2011)	534	-0.003	-2.314**	0.034
C.3.25 Net debt financing	Bradshaw, Richardson, and Sloan (2006), 1972/7	528	-0.002	-1.942*	0.026
C.3.25 Net equity financing	Bradshaw, Richardson, and Sloan (2006), 1972/8	528	-0.002	-0.797	0.047
C.3.25 Net external financing	Bradshaw, Richardson, and Sloan (2006)	528	-0.003	-1.834*	0.042
D. Profitability					
D.4.1 Return on equity	Hou, Xue, and Zhang (2015)	534	0.018	8.135***	0.051
D.4.2 4-quarter change in return on equity		528	0.004	2.715**	0.035
D.4.3 Roa1, Roa6, and Return on assets	Balakrishnan, Bartov, and Faurel (2010), 1972/1	534	0.015	7.404***	0.048
D.4.4 4-quarter change in return on assets.		522	0.005	2.814***	0.037
D.4.5 Assets turnover	Soliman (2008)	534	0.001	0.540	0.045
D.4.5 Profit margin	Soliman (2008)	534	0.001	0.244	0.049
D.4.5 Return on net operating assets	Soliman (2008)	534	0.001	0.485	0.043
D.4.6 Capital turnover	Haugen and Baker (1996)	534	0.001	0.891	0.039
D.4.7 Quarterly assets turnover		534	0.004	2.176**	0.043
D.4.7 Quarterly profit margin		534	0.005	2.429**	0.048
D.4.7 Quarterly return on net operating assets		486	0.005	2.374**	0.045
D.4.8 Quarterly capital turnover	Haugen and Baker (1996)	534	0.006	3.634***	0.041
D.4.9 Gross profits-to-assets.	Novy-Marx (2013)	534	0.003	1.902*	0.034
D.4.10 Gross profits-to-lagged assets		534	0.000	0.196	0.036
D.4.11 Quarterly gross profits-to-lagged assets	Fama and French (2015)	486	0.004	3.141***	0.031
D.4.12 Operating profits to equity	Fama and French (2015)	534	0.002	1.140	0.05
D.4.13 Operating profits-to-lagged equity		534	0.001	0.400	0.045
D.4.14 Quarterly operating profits-to-lagged equity		534	0.008	3.392***	0.055
D.4.15 Operating profits-to-assets	Ball et al. (2015)	534	0.004	2.037**	0.044
D.4.16 Operating profits-to-lagged assets		534	0.003	1.411	0.043
D.4.17 Quarterly operating profits-to-lagged assets		486	0.009	4.303***	0.044
D.4.18 Cash-based operating profitability	Ball et al. (2016)	534	0.007	3.530***	0.043

Candidate factors	Reference papers	# obs	mean	t-stat	std.dev
D.4.19 Cash-based operating profits-to-lagged assets		534	0.005	2.761***	0.041
D.4.20 Quarterly cash-based operating profits-to-lagged assets		486	0.007	4.324***	0.035
D.4.21 Fundamental score.	Piotroski (2000)	528	0.002	1.700*	0.033
D.4.24 Ohlson's O-score	Dichev (1998)	534	0.001	0.317	0.043
D.4.25 Quarterly O-score		486	-0.002	-1.260	0.034
D.4.26 Altman's Z-score	Dichev (1998)	534	-0.004	-2.011**	0.046
D.4.27 Quarterly Z-score		486	-0.005	-2.072**	0.052
D.4.29 Taxable income-to-book income.	Lev and Nissim (2004)	534	0.000	0.244	0.03
D.4.30 Quarterly taxable income-to-book income		534	0.001	0.581	0.038
D.4.31 Growth score	Mohanram (2005)	348	0.004	1.079	0.077
D.4.32 Book leverage	Fama and French (1992)	534	0.001	0.438	0.039
D.4.33 Quarterly book leverage		534	0.000	0.136	0.042
E. Intangibles					
E.5.1 Industry adjusted organizational capital-to-assets	Eisfeldt and Papanikolaou (2013)	534	0.001	0.353	0.042
E.5.2 Advertising expense-to-market	Chan, Lakonishok, and Sougiannis (2001)	534	0.000	0.171	0.029
E.5.3 Growth in advertising expense.	Lou (2014)	534	0.002	3.384***	0.014
E.5.4 R&D expense-to-market	Chan, Lakonishok, and Sougiannis (2001)	534	-0.002	-0.933	0.045
E.5.8 Operating leverage	Novy-Marx (2011)	534	0.001	0.438	0.033
E.5.9 Olq1, Olq6, and Olq12, quarterly operating leverage	Novy-Marx (2011)	522	0.004	2.549**	0.032
E.5.10 Hiring rate	Belo, Lin, and Bazdresch (2014)	534	0.002	2.939***	0.014
E.5.11 R&D capital-to-assets	Li (2011)	534	0.000	0.254	0.041
E.5.12 Bca, brand capital-to-assets.	Belo, Lin, and Vitorino (2014)	516	0.007	2.046**	0.074
E.5.17 Ha, industry concentration (assets)	Hou and Robinson (2006)	534	-0.003	-1.248	0.047
E.5.17 He, industry concentration (book equity)	Hou and Robinson (2006)	534	-0.002	-1.098	0.044
E.5.17 Hs, industry concentration (sales)	Hou and Robinson (2006)	534	-0.003	-1.394	0.043
E.5.19 D1, price delay	Hou and Moskowitz (2005)	534	0.002	0.976	0.043
E.5.19 D2, price delay	Hou and Moskowitz (2005)	534	0.000	-0.107	0.023
E.5.19 D3, price delay	Hou and Moskowitz (2005)	534	0.000	-0.405	0.023
E.5.20 % Δ in sales minus % Δin inventory	Abarbanell and Bushee (1998)	534	0.000	0.438	0.004
E.5.21 % Δ in sales minus % Δ in accounts receivable	Abarbanell and Bushee (1998)	534	0.000	1.085	0.006
E.5.22 % Δ in gross margin minus % Δ in sales	Abarbanell and Bushee (1998)	534	0.001	2.281**	0.006
E.5.23 % Δ in sales minus % Δ in SG&A	Abarbanell and Bushee (1998)	534	0.000	1.425	0.005

Candidate factors	Reference papers	# obs	mean	t-stat	std.dev
E.5.24 Effective tax rate	Abarbanell and Bushee (1998)	534	0.000	1.425	0.005
E.5.25 Labor force efficiency	Abarbanell and Bushee (1998)	534	0.000	1.173	0.005
E.5.26 Analysts coverage	Elgers, Lo, and Pfeiffer (2001)	485	-0.001	-0.432	0.03
E.5.27 Tangibility	Hahn and Lee (2009)	534	-0.001	-0.828	0.026
E.5.28 Quarterly tangibility.	Hahn and Lee (2009)	534	0.000	0.164	0.033
E.5.29 Industry-adjusted real estate ratio	Tuzel (2010)	534	0.001	0.467	0.037
E.5.30 Financial constraints (the Kaplan-Zingales index)	Lamont, Polk, and Saa-Requejo (2001)	534	0.002	1.521	0.03
E.5.32 Financial constraints (the Whited-Wu index)	Whited and Wu (2006)	534	0.000	0.125	0.028
E.5.33 Wwq1, Wwq6, and Wwq12, the quarterly Whited-Wu index	Whited and Wu (2006)	534	0.001	0.33	0.039
E.5.34 Secured debt-to-total debt	Valta (2016)	534	-0.001	-0.646	0.03
E.5.35 Convertible debt-to-total debt	Valta (2016)	534	0.001	0.826	0.042
E.5.37 Cta1, Cta6, and Cta12, cash-to-assets	Palazzo (2012)	534	0.002	1.079	0.045
E.5.41 Earnings persistence	Rajgopal, Shevlin, and Venkatachalam (2003)	534	-0.001	-0.665	0.032
E.5.41 Earnings predictability	Rajgopal, Shevlin, and Venkatachalam (2003)	534	-0.004	-2.162**	0.041
E.5.42 Earnings smoothness	Francis et al. (2004)	534	-0.001	-1.012	0.027
E.5.44 Earnings conservatism	Francis et al. (2004)	534	-0.002	-1.479	0.027
E.5.44 Earnings timeliness	Francis et al. (2004)	534	0.000	0.103	0.032
E.5.44 Earnings conservatism	Francis et al. (2004)	534	0.001	0.757	0.019
E.5.44 Earnings timeliness	Francis et al. (2004)	534	0.001	1.108	0.022
E.5.45 FRM, Pension plan funding rate	Franzoni and Martin (2006)	534	0.001	0.977	0.024
E.5.45 FRA, Pension plan funding rate	Franzoni and Martin (2006)	534	-0.002	-1.695*	0.03
E.5.46 Ala, asset liquidity	Ortiz-Molina and Phillips (2014)	486	0.000	-0.122	0.045
E.5.46 Alm, asset liquidity	Ortiz-Molina and Phillips (2014)	486	0.004	1.692	0.051
E.5.51 Average returns Ra1	Heston and Sadka (2008)	534	0.001	7.550***	0.002
E. 5.51 Average returns Ra[2,5]	Heston and Sadka (2008)	534	0.001	3.604***	0.004
E.5.51 Average returns Ra[6,10]	Heston and Sadka (2008)	534	0.000	3.554***	0.003
E.5.51 Average returns Rn1	Heston and Sadka (2008)	534	0.002	4.953***	0.007
E. 5.51 Average returns Rn[2,5]	Heston and Sadka (2008)	534	0.002	3.354***	0.012
E.5.51 Average returns Rn[6,10]	Heston and Sadka (2008)	534	0.001	3.143***	0.009
E.5.51 Average returns Rn[16,20]	Heston and Sadka (2008)	534	0.002	1.152	0.044
F. Trading frictions					
F.6.1 Me, market equity	Banz (1981)	534	-0.001	-0.485	0.05
F.6.2 Ivff1, Ivff6, and Ivff12, idiosyncratic volatility per the Fama and French (1993) 3-factor model	Ang, Hodrick, Xing, and Zhang (2006)	534	-0.010	-2.542**	0.088

Candidate factors	Reference papers	# obs	mean	t-stat	std.dev
F.6.3 Iv, idiosyncratic volatility	Ali, Hwang, and Trombley (2003)	534	-0.012	-3.406***	0.079
F.6.5 Ivq1, Ivq6, and Ivq12, idiosyncratic volatility		534	-0.011	-3.340***	0.079
F.6.6 Tv1, Tv6, and Tv12, total volatility	Ang, Hodrick, Xing, and Zhang (2006)	534	-0.013	-3.427***	0.089
F.6.8 β1, β6, and β12, market beta	Fama and MacBeth (1973)	534	0.000	-0.125	0.08
F.6.9 βFP1, βFP6, and βFP12, the Frazzini-Pedersen beta	Frazzini and Pedersen (2013)	534	-0.006	-1.529	0.095
F.6.10 βD1, βD6, and βD12, the Dimson beta	Dimson (1979)	533	-0.001	-0.51	0.058
F.6.11 Tur1, Tur6, and Tur12, share turnover	Datar, Naik, and Radcliffe (1998)	534	-0.002	-0.823	0.063
F.6.12 Cvt1, Cvt6, and Cvt12, coefficient of variation of share turnover	Chordia, Subrahmanyam, and Anshuman (2001)	533	0.000	-0.106	0.034
F.6.13 Dtv1, Dtv6, and Dtv12, dollar trading volume	Brennan, Chordia, and Subrahmanyam (1998)	533	-0.001	-0.605	0.034
F.6.14 Cvd1, Cvd6, and Cvd12, coefficient of variation of dollar trading volume.	Chordia, Subrahmanyam, and Anshuman (2001)	533	0.001	0.37	0.033
F.6.15 Pps1, Pps6, and Pps12, share price	Miller and Scholes (1982)	534	0.000	0.127	0.084
F.6.16 Ami1, Ami6, and Ami12, absolute return-to-volume	Amihud (2002)	533	-0.001	-0.396	0.05
F.6.17 Lm11, Lm16, Lm112, turnover-adjusted number of zero daily volume	Liu (2006)	533	0.000	-0.014	0.058
F.6.17. Lm121, Lm126, Lm1212, turnover-adjusted number of zero daily volume	Liu (2006)	533	0.002	0.695	0.06
F.6.17, Lm61, Lm66, Lm612, turnover-adjusted number of zero daily volume	Liu (2006)	533	0.002	0.711	0.061
F.6.18 Mdr1, Mdr6, and Mdr12, maximum daily return	Bali, Cakici, and Whitelaw (2011)	534	-0.008	-2.586**	0.074
F.6.20 Isc1, Isc6, and Isc12, idiosyncratic skewness per the CAPM		534	0.003	2.249**	0.027
F.6.21 Isff1, Isff6, and Isff12, idiosyncratic skewness per the Fama and French	Bali, Engel, and Murray (2016)	534	0.003	2.756***	0.025
F.6.23 Cs1, Cs6, and Cs12, coskewness	Harvey and Siddique (2000)	534	-0.001	-0.806	0.032
F.6.25 βlcc1, βlcc6, βlcc12, liquidity betas illiquidity-illiquidity	Kelly and Jiang (2014)	533	0.026	9.205***	0.065
F.6.25 βlcr1, βlcr6, βlcr12, liquidity betas (illiquidity-return)	Kelly and Jiang (2014)	533	0.001	0.438	0.042
F.6.25 βlrc1, βlrc6, βlrc12, liquidity betas return illiquidity	Kelly and Jiang (2014)	533	-0.003	-1.632	0.047
F.6.25 βnet1, βnet6, and βnet12, liquidity betas (net)	Kelly and Jiang (2014)	533	0.006	1.864*	0.077

Candidate factors	Reference papers	# obs	mean	t-stat	std.dev
F.6.25 βret1, βret6, and βret12, liquidity betas (return-return)	Kelly and Jiang (2014)	533	0.006	1.886*	0.078
F.6.26 Short-term reversal	Jegadeesh (1990)	533	0.003	1.307	0.051
F.6.27 β−1, β−6, and β−12, downside beta	Ang, Chen, and Xing (2006)	533	-0.002	-0.626	0.073
F.6.31 βPS1, βPS6, and βPS12, the Pastor-Stambaugh beta		534	0.001	0.304	0.04

Appendix E: Firm characteristics

This table presents characteristics constructed according to Green, Hand and Zhang (2018)

Name	Description	min	max	mean	median	skew	kurtosis
acc	Working capital accruals	-1.02	0.58	-0.02	-0.02	-0.88	4.81
aeavol	Abnormal earnings announcement volume	-1.00	21.69	0.87	0.30	3.64	18.51
age	# years since first Compustat coverage	1.00	56.00	12.72	9.00	1.36	1.55
agr	Asset growth	-0.68	5.85	0.15	0.08	4.26	30.02
baspread	Bid-ask spread	0.00	0.91	0.05	0.03	5.15	38.11
beta	Beta	-0.74	3.94	1.08	1.01	0.69	0.69
bm	Book-to-market	-2.35	7.81	0.77	0.60	2.48	11.46
cash	Cash holdings	0.00	0.98	0.16	0.07	1.89	3.15
cashdebt	Cash flow to debt	-7.71	2.23	0.07	0.13	-4.14	25.99
cashpr	Cash productivity	-520.62	600.28	-1.90	-0.73	0.89	29.15
cfp	Cash flow to price ratio	-513.56	156.76	0.05	0.05	-172.20	57390.53
cfp_ia	Industry-adjusted cash flow to price ratio	-449.37	7031.6	13.09	0.00	21.92	479.82
chatoia	Industry-adjusted change in asset turnover	-1.43	1.19	0.00	0.00	-0.15	4.74
chcsho	Δ shares outstanding	-0.89	2.57	0.11	0.01	3.28	13.85
chfeps	Δ forecasted EPS	-6.48	8.25	0.00	0.00	1.29	121.37
chinv	Δ inventory	-0.29	0.37	0.01	0.00	1.10	6.77
chnanalyst	Δ number of analysts	-12.00	9.00	-0.01	0.00	-0.60	9.46
chtx	Δ tax expense	-0.12	0.16	0.00	0.00	0.35	13.06
cinvest	Corporate investment	-26.83	27.87	-0.02	0.00	-2.17	244.24
currat	Current ratio	0.16	60.34	3.16	2.00	5.58	40.35
depr	Depreciation/PP&E	0.01	5.51	0.26	0.15	5.92	49.70
disp	Dispersion in forecasted EPS	0.00	10.00	0.15	0.04	6.48	58.18
dy	Dividend to price	0.00	0.35	0.02	0.00	2.67	10.86
ear	Earnings announcement return	-0.46	0.51	0.00	0.00	0.26	3.17
egr	Growth in common shareholder equity	-3.54	8.19	0.14	0.08	3.32	28.51
ep	Earnings to price	-7.66	0.68	-0.01	0.05	-8.11	107.23
fgr5yr	Forecasted growth in 5-year EPS	-43.50	99.41	16.35	14.50	1.50	5.47
gma	Gross profitability	-0.84	1.78	0.37	0.33	0.81	1.52
grcapx	Growth in capital expenditures	-13.89	55.54	0.89	0.14	5.60	45.95
grltnoa	Growth in long term net operating assets	-0.61	1.18	0.09	0.06	1.64	7.48
herf	Industry sales concentration	0.01	1.00	0.08	0.05	3.10	11.91
hire	Employee growth rate	-0.74	4.00	0.09	0.02	3.81	24.97
idiovol	Idiosyncratic return volatility	0.01	0.26	0.06	0.06	1.47	2.70
ill	Illiquidity	0.00	0.00	0.00	0.00	14.63	355.90
indmom	Industry momentum	-1.00	3.56	0.14	0.12	1.26	5.39
invest	Capital expenditures and inventory	-0.52	2.21	0.08	0.04	2.51	12.80
lev	Leverage	0.00	77.75	2.28	0.69	5.47	43.92
Meanrec	Mean number of analysts	1.00	4.50	2.22	2.20	-0.01	-0.32
mom12m	12-month momentum	-1.00	11.60	0.13	0.06	2.89	21.73
mom1m	1-month momentum	-0.70	2.11	0.01	0.00	1.16	7.77
mom36m	36-month momentum	-0.98	16.20	0.33	0.16	3.08	20.08
ms	Financial statement score	0.00	8.00	3.73	4.00	-0.03	-0.72
mve	Size	6.02	18.90	11.77	11.63	0.29	-0.31

Name	Description	min	max	mean	median	skew	kurtosis
mve_ia	Industry-adjusted size	-16,608	133,635	-158.3	-359.12	9.17	120.31
nanalyst	Number of analysts covering stock	0.00	34.00	5.17	3.00	1.75	2.80
nincr	Number of earnings increases	0.00	8.00	1.00	1.00	2.15	6.35
orgcap	Organizational capital	0.00	0.18	0.01	0.01	2.73	11.60
pchcapx_ia	Industry adjusted % Δ capital expenditures	-237.42	1640.09	6.50	-0.35	15.2	273.34
pchcurrat	% Δ current ratio	-0.89	6.72	0.06	-0.01	3.82	23.69
pchdepr	% Δ depreciation	-0.85	7.37	0.10	0.03	4.56	36.69
pchgm_pchsale	% Δ gross margin - % Δ sales	-12.26	4.77	-0.06	0.00	-5.49	54.90
pchsale_pchinvt	% Δ sales - % Δ inventory	-11.61	3.02	-0.06	0.01	-5.85	57.26
pchsale_pchrect	% Δ sales - %Δ A/R	-7.93	3.11	-0.04	0.00	-2.90	24.02
pchsale_pchxsga	% Δ sales - % Δ SG&A	-3.50	4.34	0.02	0.00	3.42	31.72
pctacc	Percent accruals	-64.75	71.43	-0.65	-0.27	-1.90	34.80
pricedelay	Price delay	-15.85	15.52	0.14	0.06	0.09	39.90
ps	Financial statement score	0.00	8.00	4.18	4.00	0.03	-0.55
rd_mve	R&D to market capitalization	0.00	2.23	0.06	0.03	5.12	48.03
rd_sale	R&D to sales	0.00	283.48	0.61	0.03	23.6	733.49
retvol	Return volatility	0.00	0.27	0.03	0.02	2.42	8.82
roaq	Return on assets	-0.48	0.16	0.00	0.01	-3.40	16.22
roavol	Earnings volatility	0.00	0.85	0.03	0.01	5.31	41.10
roe	Return on equity	-7.05	8.80	0.03	0.10	-2.56	35.80
roeq	Quarterly return on equity	-2.22	1.66	0.00	0.02	-2.80	29.85
roic	Return on invested capital	-21.24	1.01	-0.08	0.07	-10.4	149.98
rsup	Revenue surprise	-4.51	2.33	0.02	0.01	-3.83	64.41
salecash	Sales to cash	0.00	2,503.5	52.60	10.60	7.68	73.99
saleinv	Sales to inventory	0.29	1,031.2	25.91	7.59	6.96	62.95
salerec	Sales to receivables	0.00	594.00	11.68	5.94	5.25	31.16
sfe	Scaled earnings forecast	-36.23	1.09	-0.06	0.04	-14.8	296.41
sgr	Sales growth	-0.91	8.50	0.18	0.09	5.68	49.16
sp	Sales to price	0.00	54.59	2.32	1.10	4.51	29.91
spi	Industry-adjusted sales to price	-0.66	0.19	-0.01	0.00	-5.20	40.89
std_dolvol	Volatility of liquidity (dollar trading volume)	0.18	2.74	0.86	0.79	0.75	0.18
std_turn	Volatility of liquidity (share turnover)	0.02	184.01	3.90	1.90	6.07	64.87
stdcf	Cash flow volatility	0.00	1,882.9	9.88	0.14	11.9	178.00
sue	Unexpected quarterly earnings	-5.20	1.70	0.00	0.00	-13.4	550.18
tang	Debt capacity/firm tangibility	0.04	0.98	0.54	0.55	-0.14	0.99
tb	Tax income to book income	-27.70	15.36	-0.10	-0.03	-4.47	66.74
turn	Share turnover	0.00	195.94	1.02	0.52	20.4	1,384.13
zerotrade	Zero trading days	0.00	19.95	1.31	0.00	3.03	9.22

Appendix F: Nonlinear characteristics portfolios

This Appendix presents nine characteristics selected by Freyberger, Nieurehl and Weber (2020). and 2,120 non-linear characteristic-sorted decile portfolios constructed based on 212 characteristics. See Section IV for the detail construction. The nine characteristics are *agr* defined as annual percent change in total assets from Cooper, Gulen & Schill (2008), *chcsho* or annual percent change in shares outstanding from Pontiff & Woodgate (2008), *mom1m* defined as 1-month cumulative return from Jegadeesh & Titman (1993), *mom12m* defined as 11-month cumulative returns ending one month before month end from Jegadeesh (1990), *mom36m* defined as Cumulative returns from months *t*-36 to *t*-13, *operprof* or revenue minus cost of goods sold - SG&A expense - interest expense divided by lagged common shareholders' equity (Fama and French, 2015), *mve* or natural log of market capitalization at end of month t-1 from Banz (1981), *retvol* or standard deviation of daily returns from month t-1 from Ang, Hodrick, Xing & Zhang (2006) and *turn* or the average monthly trading volume for most recent 3 months scaled by number of shares outstanding in current month from Datar, Naik & Radcliffe (1998).

Portfolio name	**ID**
agr_agr	X8
agr_chcsho	X9
agr_mom12m	X10
agr_mom1m	X11
agr_mve	X12
agr_retvol	X13
agr_turn	X14
chcsho_chcsho	X15
chcsho_mom12m	X16
chcsho_mom1m	X17
chcsho_mve	X18
chcsho_retvol	X19
chcsho_turn	X20
mom12m_mom12m	X21
mom12m_mom1m	X22
mom12m_mve	X23
mom12m_retvol	X24
mom12m_turn	X25
mom1m_mom1m	X26
mom1m_mve	X27
mom1m_retvol	X28
mom1m_turn	X29
mve_mve	X30
mve_retvol	X31
mve_turn	X32
retvol_retvol	X33
retvol_turn	X34
turn_turn	X35
agr_agr_agr	X36
agr_agr_chcsho	X37
agr_agr_mom12m	X38
agr_agr_mom1m	X39

Portfolio name	ID
agr_agr_mve	X40
agr_agr_retvol	X41
agr_agr_turn	X42
agr_chcsho_chcsho	X43
agr_chcsho_mom12m	X44
agr_chcsho_mom1m	X45
agr_chcsho_mve	X46
agr_chcsho_retvol	X47
agr_chcsho_turn	X48
agr_mom12m_mom12m	X49
agr_mom12m_mom1m	X50
agr_mom12m_mve	X51
agr_mom12m_retvol	X52
agr_mom12m_turn	X53
agr_mom1m_mom1m	X54
agr_mom1m_mve	X55
agr_mom1m_retvol	X56
agr_mom1m_turn	X57
agr_mve_mve	X58
agr_mve_retvol	X59
agr_mve_turn	X60
agr_retvol_retvol	X61
agr_retvol_turn	X62
agr_turn_turn	X63
chcsho_chcsho_chcsho	X64
chcsho_chcsho_mom12m	X65
chcsho_chcsho_mom1m	X66
chcsho_chcsho_mve	X67
chcsho_chcsho_retvol	X68
chcsho_chcsho_turn	X69
chcsho_mom12m_mom12m	X70
chcsho_mom12m_mom1m	X71
chcsho_mom12m_mve	X72
chcsho_mom12m_retvol	X73
chcsho_mom12m_turn	X74
chcsho_mom1m_mom1m	X75
chcsho_mom1m_mve	X76
chcsho_mom1m_retvol	X77
chcsho_mom1m_turn	X78
chcsho_mve_mve	X79
chcsho_mve_retvol	X80
chcsho_mve_turn	X81
chcsho_retvol_retvol	X82
chcsho_retvol_turn	X83
chcsho_turn_turn	X84
mom12m_mom12m_mom12m	X85
mom12m_mom12m_mom1m	X86
mom12m_mom12m_mve	X87

Portfolio name	**ID**
mom12m_mom12m_retvol	X88
mom12m_mom12m_turn	X89
mom12m_mom1m_mom1m	X90
mom12m_mom1m_mve	X91
mom12m_mom1m_retvol	X92
mom12m_mom1m_turn	X93
mom12m_mve_mve	X94
mom12m_mve_retvol	X95
mom12m_mve_turn	X96
mom12m_retvol_retvol	X97
mom12m_retvol_turn	X98
mom12m_turn_turn	X99
mom1m_mom1m_mom1m	X100
mom1m_mom1m_mve	X101
mom1m_mom1m_retvol	X102
mom1m_mom1m_turn	X103
mom1m_mve_mve	X104
mom1m_mve_retvol	X105
mom1m_mve_turn	X106
mom1m_retvol_retvol	X107
mom1m_retvol_turn	X108
mom1m_turn_turn	X109
mve_mve_mve	X110
mve_mve_retvol	X111
mve_mve_turn	X112
mve_retvol_retvol	X113
mve_retvol_turn	X114
mve_turn_turn	X115
retvol_retvol_retvol	X116
retvol_retvol_turn	X117
retvol_turn_turn	X118
turn_turn_turn	X119
agr_operprof	X121
chcsho_operprof	X122
mom12m_operprof	X123
mom1m_operprof	X124
mve_operprof	X125
operprof_operprof	X126
operprof_retvol	X127
operprof_turn	X128
agr_agr_operprof	X129
agr_chcsho_operprof	X130
agr_mom12m_operprof	X131
agr_mom1m_operprof	X132
agr_mve_operprof	X133
agr_operprof_operprof	X134
agr_operprof_retvol	X135
agr_operprof_turn	X136

Portfolio name	ID
chcsho_chcsho_operprof	X137
chcsho_mom12m_operprof	X138
chcsho_mom1m_operprof	X139
chcsho_mve_operprof	X140
chcsho_operprof_operprof	X141
chcsho_operprof_retvol	X142
chcsho_operprof_turn	X143
mom12m_mom12m_operprof	X144
mom12m_mom1m_operprof	X145
mom12m_mve_operprof	X146
mom12m_operprof_operprof	X147
mom12m_operprof_retvol	X148
mom12m_operprof_turn	X149
mom1m_mom1m_operprof	X150
mom1m_mve_operprof	X151
mom1m_operprof_operprof	X152
mom1m_operprof_retvol	X153
mom1m_operprof_turn	X154
mve_mve_operprof	X155
mve_operprof_operprof	X156
mve_operprof_retvol	X157
mve_operprof_turn	X158
operprof_operprof_operprof	X159
operprof_operprof_retvol	X160
operprof_operprof_turn	X161
operprof_retvol_retvol	X162
operprof_retvol_turn	X163
operprof_turn_turn	X164
agr_mom36m	X166
chcsho_mom36m	X167
mom12m_mom36m	X168
mom1m_mom36m	X169
mom36m_mom36m	X170
mom36m_mve	X171
mom36m_operprof	X172
mom36m_retvol	X173
mom36m_turn	X174
agr_agr_mom36m	X175
agr_chcsho_mom36m	X176
agr_mom12m_mom36m	X177
agr_mom1m_mom36m	X178
agr_mom36m_mom36m	X179
agr_mom36m_mve	X180
agr_mom36m_operprof	X181
agr_mom36m_retvol	X182
agr_mom36m_turn	X183
chcsho_chcsho_mom36m	X184
chcsho_mom12m_mom36m	X185

Portfolio name	ID
chcsho_mom1m_mom36m	X186
chcsho_mom36m_mom36m	X187
chcsho_mom36m_mve	X188
chcsho_mom36m_operprof	X189
chcsho_mom36m_retvol	X190
chcsho_mom36m_turn	X191
mom2m_mom2m_mom36m	X192
mom12m_mom1m_mom36m	X193
mom12m_mom36m_mom36m	X194
mom12m_mom36m_mve	X195
mom12m_mom36m_operprof	X196
mom12m_mom36m_retvol	X197
mom12m_mom36m_turn	X198
mom1m_mom1m_mom36m	X199
mom1m_mom36m_mom36m	X200
mom1m_mom36m_mve	X201
mom1m_mom36m_operprof	X202
mom1m_mom36m_retvol	X203
mom1m_mom36m_turn	X204
mom36m_mom36m_mom36m	X205
mom36m_mom36m_mve	X206
mom36m_mom36m_operprof	X207
mom36m_mom36m_retvol	X208
mom36m_mom36m_turn	X209
mom36m_mve_mve	X210
mom36m_mve_operprof	X211
mom36m_mve_retvol	X212
mom36m_mve_turn	X213
mom36m_operprof_operprof	X214
mom36m_operprof_retvol	X215
mom36m_operprof_turn	X216
mom36m_retvol_retvol	X217
mom36m_retvol_turn	X218
mom36m_turn_turn	X219

Appendix G: Tree portfolios

This Appendix shows 36 tree characteristics groups, each of which contains 10 decile portfolios. See Bryzgalova, Pelger and Zhu (2020) for the detail construction and variable description on th next page.

LME_AC_IdioVol
LME_AC_Lturnover
LME_BEME_AC
LME_BEME_IdioVol
LME_BEME_Investment
LME_BEME_LT_Rev
LME_BEME_Lturnover
LME_BEME_OP
LME_BEME_r12_2
LME_BEME_ST_Rev
LME_IdioVol_Lturnover
LME_Investment_AC
LME_Investment_Idiovol
LME_Investment_LT_Rev
LME_investment_Lturnover
LME_Investment_ST_Rev
LME_LT_Rev_AC
LME_LT_Rev_IdioVol
LME_LT_Rev_Lturnover
LME_OP_AC
LME_OP_IdioVol
LME_OP_Investment
LME_OP_LT_Rev
LME_OP_Lturnover
LME_OP_ST_Rev
LME_r12_2_AC
LME_r12_2_IdioVol
LME_r12_2_Investment
LME_r12_2_LT_Rev
LME_r12_2_Lturnover
LME_r12_2_OP
LME_r12_2_ST_Rev
LME_ST_REV_AC
LME_ST_Rev_IdioVol
LME_ST_Rev_LT_Rev

Symbol	Names	Description	References
AC	Accrual	Change in operating working capital per split-adjusted share from the scal year	Sloan (1996)
BEME	Book-to-Market ratio	Book equity is shareholder equity (SH) plus deferred taxes and investment tax credit (TXDITC), minus preferred stock (PS). SH is shareholders equity (SEQ). If missing, SH is the sum of common equity (CEQ) and preferred stock (PS). If missing, SH is the difference between total assets (AT) and total liabilities (LT). Depending on availability, we use the redemption (item PSTKRV), liquidating (item PSTKL), or per value (item PSTK) for PS. The market value of equity (PRC*SHROUT) is as of December t-1.	Basu (1983), Fama and French (1992)
IdioVol	Idiosyncratic volatility	Standard deviation of the residuals from a regression of excess returns on the Fama and French three-factor model	Ang, Hodrick, Xing, and Zhang (2006)
Investment	Investment	Change in total assets (AT) from the fiscal year ending in year t-2 to the fiscal year ending in t-1, divided by t-2 total assets	Fama and French (2015)
LME	Size	Total market capitalization at the end of the previous month defined as price times shares outstanding	Banz (1981), Fama and French (1992)
LT_Rev	Long-term reversal	Cumulative return from 60 months before the return prediction to 13 months before	De Bondt and Thaler (1985)
LTurnover	Turnover	Last month's volume (VOL) over shares outstanding (SHROUT)	Datar, Naik, and Radcliffe (1998)
OP	Operating profitability	Annual revenues (REVT) minus cost of goods sold (COGS), interest expense (TIE), and selling, general, and administrative expenses (XSGA) divided by book equity (defined in BEME)	Fama and French (2015)
r12_2	Momentum	Return for the first 12 months except for the first month	Jegadeesh and Titman (1993)
ST_Rev	Short-term reversal	Prior month return	Jegadeesh (1990)

Appendix H1: Application 1 using Elastic Net with lambda = 0.05

This table presents the nonlinear characteristics that are selected by Elastic Net with lambda of 0.05 and that are tested to be significant under the OLS regression where stochastic discount factor ($\mathbf{m}$) is extracted from multiple assets. See Section xxx for the construction of non-linear factors and the appendix for the full name of each characteristics. $\mathbf{m}$ is constructed from 1) 1,720 portfolios from 172 anomalies constructed similarly to Hou et al (2019), $\mathbf{m}$_anomaly (2) 87 individual stocks that survive for the whole sample period, $\mathbf{m}$ _stock and (3) 2,100 nonlinear portfolios we construct from 9 characteristics up to 3 polynomials, $\mathbf{m}$_nonlinear. The 9 characteristics are those that Freybeyger, Nieurehl and Webers (2019) find significant. (4) is 360 tree portfolios, $\mathbf{m}$_Tree, (5) is nonlinear portfolios and trees portfolios, $\mathbf{m}$_nonlinear and trees, and (6) is all four portfolios, $\mathbf{m}$_Mixed. $\mathbf{m}$ is augmented by conditioning information ($\mathbf{Z}$) including dividend price ratio, default yield spread, risk-free rate, risk-free rate, term spread, CORPR, stock variance, inflation, and net equity expansion available from Amit Goyal's website and Cay by Lettau and Ludvigson. We denote this augmented $\mathbf{m}$ as $\widehat{\mathbf{m}}(\mathbf{Z})$. We apply ridge regression to $\mathbf{m}$ and present the conditioning the SDF ($\widehat{\mathbf{m}}(\mathbf{Z})$) that has shrinkage term, $\mathrm{s(N)}\mathbf{I}$, of zero and at 100th percentile.

$$\widehat{\mathbf{m}}(\mathbf{Z}) = \mathrm{T}(\overline{\mathbf{R}}'\overline{\mathbf{R}} + \mathrm{s(N)}\mathbf{I})^{-1}(\overline{\mathbf{R}}'\mathbf{1}_{\mathrm{NM}})$$

We perform three steps. First we regress each $\widehat{\mathbf{m}}(\mathbf{Z})$ on six polynomials of market factors where market factor is market rate of return subtracted by risk-free rate. Second, we use the residual as a dependent variable, apply Elastic Net to select nonlinear factors, and present them in the table. Third, we perform an OLS regression of the residuals of M on the selected nonlinear factors. ***, **, and * present 1%, 5% and 10% significance, respectively. "NA" means there is no nonlinear factors chosen by Elastic Net with lambda of 0.05.

Nonlinear factor	$\boldsymbol{m}$_anomaly			$\boldsymbol{m}$ _stock			$\boldsymbol{m}$_nonlinear			$\boldsymbol{m}$_Tree			$\boldsymbol{m}$_Mixed			$\boldsymbol{m}$_nonlinear and trees	
	Shrinkage level																
	0th	100th		0th	100th		0th	100th		0th	100th		0th	100th		0th	100th
X9								-0.136		0.24	-0.252						
X11										0.622	0.88						
X12								0.926		0.244	-0.974			1.766*			1.329
X13					1.947*			0.435			-1.211						
X14										0.672	1.902*						-0.793
X15								0.07									-0.868
X16								-2.52**		-0.506	-0.275			-3.35***			-3.16***
X17											0.929						
X18								1.64		2.468**	1.954*					0.028	1.175

Nonlinear factor	$\boldsymbol{m}$_anomaly			$\boldsymbol{m}$ _stock			$\boldsymbol{m}$_nonlinear			$\boldsymbol{m}$_Tree			$\boldsymbol{m}$_Mixed			$\boldsymbol{m}$_nonlinear and trees	
	Shrinkage level																
	0^{th}	100^{th}		0^{th}	100^{th}		0^{th}	100^{th}		0^{th}	100^{th}		0^{th}	100^{th}		0^{th}	100^{th}
X19					-1.274			-0.813									
X20					0.614						-1.487						
X21					-0.638					-0.9	-2.464**						
X22								1.319		-0.436	-0.473						1.317
X23								-0.260									
X24					-1.3						0.504						
X25								-1.400									
X26								-1.762		-1.438	-2.81***						-1.471
X27											-0.893						
X28								-0.747									
X29								0.679		1.426	1.4						
X31								-1.37									
X32					-0.729					0.208	0.943			0.103			0.095
X33								0.909			-0.891						
X34								-0.293									
X35								-0.369			-1.790						
X36										1.128	1.503						
X37											0.002						
X38										-0.759	-1.341					-0.289	1.03
X39								1.297			-1.255						
X40								1.281									
X41					-1.279			0.325						-0.488			-0.773
X42								-0.299									
X43					-0.975												
X46										0.253	-0.26						-0.376
X47								1.47		-0.082	0.72			0.84		0.967	1.391
X48											-1.732						
X49					0.372						0.031						
X50								-1.689		-0.899	-0.445			-2.01**		-1.97*	-1.222
X51					-3.006***												
X52								2.635**									
X53								-1.649			0.114						
X56											0.053						

Nonlinear factor	$\boldsymbol{m}$_anomaly			$\boldsymbol{m}$ _stock			$\boldsymbol{m}$_nonlinear			$\boldsymbol{m}$_Tree			$\boldsymbol{m}$_Mixed			$\boldsymbol{m}$_nonlinear and trees	
	Shrinkage level																
	0^{th}	100^{th}		0^{th}	100^{th}		0^{th}	100^{th}		0^{th}	100^{th}		0^{th}	100^{th}		0^{th}	100^{th}
X57								-2.002**		-0.427	-0.54			-1.012			-1.16
X58					2.246**						-0.674						
X59								-0.224			0.244						
X61										0.274	-0.072						0.103
X62																	-0.34
X63											1.078						
X64					-0.954					-0.812	-1.281						
X65					1.021												
X66								-0.122									
X67					-1.286												
X68								-0.813									
X69					-0.002						0.076						
X70										1.123	0.458						
X71					-1.352			-0.648		0.268	-0.094						
X73								-0.476		-1.734	-1.16						
X74					-0.584						1.341						
X75										0.059	1.094						1.3
X77					3.03***			-3.00***		1.155	1.309						
X78											-1.515						
X79					2.845***			2.917***									
X80								1.05									
X81					-1.401			-1.205			0.955						
X82								0.656		0.576	1.87*						
X83								-0.216						1.96*			0.898
X84										1.868*	1.111						1.735
X85								-1.452			-2.81***						
X86					-1.014			-0.211		-1.773	-2.012**						
X87								-0.748			-0.581						
X88											-1.14						
X89								-1.146									
X90					0.703			-0.365		-0.272	-1.24						
X91										-0.899	-1.38						
X92					0.41						-1.125						

Nonlinear factor	$\boldsymbol{m}$_anomaly			$\boldsymbol{m}$ _stock			$\boldsymbol{m}$_nonlinear			$\boldsymbol{m}$_Tree			$\boldsymbol{m}$_Mixed			$\boldsymbol{m}$_nonlinear and trees	
	Shrinkage level																
	0^{th}	100^{th}		0^{th}	100^{th}		0^{th}	100^{th}		0^{th}	100^{th}		0^{th}	100^{th}		0^{th}	100^{th}
X93								2.045**		-0.189	-0.722						
X94								-0.578		-0.316	0.005						-1.242
X95					-1.344			-0.12		1	0.757			-1.082			
X96					-1.319					-0.111	-1.277						
X97					-1.271			-2.78***			-1.159						
X98										1.847*	1.949*			-1.999*			
X99											1.01						
X100								0.235		0.317	0.559						
X101											0.539						
X102											-1.882*						
X104										-0.148	-0.785					-0.402	-0.495
X105								2.17**		0.903	0.853						
X106										-0.781	0.278					-1.067	-0.528
X107					-1.023			2.398**			-1.036						
X108								-1.57		0.058	0.943						
X110								-1.085		-0.076	-0.708			-1.404		-1.384	-1.631
X111								2.627**		2.432**	1.544			1.554		2.061**	1.652
X112								-0.407		-1.809	-2.475						
X113											-0.439						
X114					-0.935			1.262									
X115					0.071												
X116								1.728		1.646	1.49						
X117										0.68	0.174						
X118					1.238					-0.982	-0.193						
X119								-0.793		2.426	1.312						
X122					1.284						-0.915						
X123										-0.648	-1.244						
X125								0.999			2.609						
X126								-1.085		0.445	0.524						-0.251
X127					0.091			-0.963			0.926						
X132					1.715												
X133					-1.267						2.646						
X134								0.348		0.742	0.105						

Nonlinear factor	$\boldsymbol{m}$_anomaly			$\boldsymbol{m}$ _stock			$\boldsymbol{m}$_nonlinear			$\boldsymbol{m}$_Tree			$\boldsymbol{m}$_Mixed			$\boldsymbol{m}$_nonlinear and trees	
	Shrinkage level																
	0th	100th		0th	100th		0th	100th		0th	100th		0th	100th		0th	100th
X135					1.612			-1.58		-1.516	-0.673			1.574		1.144	0.96
X137					-0.977					-0.295	-1.105						
X138								-1.137									
X139											-0.025						
X141								0.232									
X142											0.387						
X143								1.196			-0.101						
X144								1.826*			1.025						
X145								1.254									
X146					1.212			-0.87			-1.052						
X148								-2.189**		-2.78***	-3.14***						
X150					-3.159						0.228						
X151								0.362									
X152								2.116**		1.629	0.772						
X154										-0.98	-0.742						
X156					-1.165			0.881		-0.496	0.467						
X157								0.388									
X158					-0.373						-0.658						
X159								-1.278			-0.109						
X160								1.13									
X161								-0.368			0.616						
X163					1.11												
X164											1.166						
X167					-2.131**			1.635									
X168								0.084			-0.417						
X169								-1.922*									
X170								-2.145**			-2.76***						
X171					1.284						0.26						
X172								-0.29			-0.657						
X173										-1.624	-1.332						
X174								0.345									
X175								0.664			1.949*						
X180								1.42									

Nonlinear factor	$\boldsymbol{m}$_anomaly			$\boldsymbol{m}$ _stock			$\boldsymbol{m}$_nonlinear			$\boldsymbol{m}$_Tree			$\boldsymbol{m}$_Mixed			$\boldsymbol{m}$_nonlinear and trees	
	Shrinkage level																
	0th	100th		0th	100th		0th	100th		0th	100th		0th	100th		0th	100th
X181											0.22						
X183											1.406						
X185											0.85						
X186								-0.01			0.908						
X187								-0.52									
X188					-0.078					-0.383	-1.516						
X189											-0.982						-2.378**
X191					-0.208			1.603			0.474						
X192					0.804												
X193								-1.087		0.108	-1.321			-1.757*			-1.175
X195										-1.682	-2.285**						
X196											-1.47						
X197								-1.482									
X198											0.165						
X199										1.049	0.484						
X200											0.862						
X201					-0.61			-1.469									
X202										-0.579	-0.536						
X203								0.435			-0.357						
X204											-0.402						
X206								-0.224									
X208														1.051			
X209					-0.523												
X210								0.382		0.764	0.979						
X211					-0.699						-0.187						
X213					0.507						0.075						
X214											0.223						
X215											1.865*						
X216								-0.755			-1.444						
X217					-0.948			-0.115			0.489						
X218								-1.426			-1.181						

Appendix H2: Application 1 using Elastic Net with lambda = 0.03

This table presents the nonlinear characteristics that are selected by Elastic Net with lambda of 0.05 and that are tested to be significant under the OLS regression where stochastic discount factor ($\mathbf{m}$) is extracted from multiple assets. See Section xxx for the construction of non-linear factors and the appendix for the full name of each characteristics. $\mathbf{m}$ is constructed from 1) 1,720 portfolios from 172 anomalies constructed similarly to Hou et al (2019), $\mathbf{m}$_anomaly (2) 87 individual stocks that survive for the whole sample period, $\mathbf{m}$ _stock and (3) 2,100 nonlinear portfolios we construct from 9 characteristics up to 3 polynomials, $\mathbf{m}$_nonlinear. The 9 characteristics are those that Freybeyger, Nieurehl and Webers (2019) find significant. (4) is 360 tree portfolios, $\mathbf{m}$_Tree, (5) is nonlinear portfolios and trees portfolios, $\mathbf{m}$_nonlinear and trees, and (6) is all four portfolios, $\mathbf{m}$_Mixed. $\mathbf{m}$ is augmented by conditioning information ($\mathbf{Z}$) including dividend price ratio, default yield spread, risk-free rate, risk-free rate, term spread, CORPR, stock variance, inflation, and net equity expansion available from Amit Goyal's website and Cay by Lettau and Ludvigson. We denote this augmented $\mathbf{m}$ as $\hat{\mathbf{m}}(\mathbf{Z})$. We apply ridge regression to $\mathbf{m}$ and present the conditioning the SDF ($\hat{\mathbf{m}}(\mathbf{Z})$) that has shrinkage term, $\mathrm{s(N)}\mathbf{I}$, of zero and at 100th percentile.

$$\hat{\mathbf{m}}(\mathbf{Z}) = \mathrm{T}(\bar{\mathbf{R}}'\bar{\mathbf{R}} + \mathrm{s(N)}\mathbf{I})^{-1}(\bar{\mathbf{R}}'\mathbf{1}_{\mathrm{NM}})$$

We perform three steps. First we regress each $\hat{\mathbf{m}}(\mathbf{Z})$ on six polynomials of market factors where market factor is market rate of return subtracted by risk-free rate. Second, we use the residual as a dependent variable, apply Elastic Net to select nonlinear factors, and present them in the table. Third, we perform an OLS regression of the residuals of M on the selected nonlinear factors. ***, **, and * present 1%, 5% and 10% significance, respectively. "NA" means there is no nonlinear factors chosen by Elastic Net with lambda of 0.03.

Nonlinear factor	$\boldsymbol{m}$_anomaly		$\boldsymbol{m}$ _stock		$\boldsymbol{m}$_nonlinear		$\boldsymbol{m}$_Tree		$\boldsymbol{m}$_Mixed		$\boldsymbol{m}$_nonlinear and trees	
	Shrinkage level											
	0th	100th	0th	100th	0th	100th	0th	100th	0th	100th	0th	100th
X9						-0.522	0.015	-0.253				
X10								-0.366				
X11							0.78	0.935				
X12						0.805	0.1	-0.915		1.064		1.186
X13				1.976*		0.603		-1.183				
X14							1.869*	1.723				-0.599
X16						-1.8	-0.093	-0.138				-3.157***
X17						-0.177		0.874				
X18						1.013	2.554**	1.732		-1.058	0.011	1.197
X20				0.431			-1.275	-1.447				
X21				-0.671			-1.31	-2.456**				
X22						1.586	0.14	-0.436				0.88

Nonlinear factor	$\boldsymbol{m}$_anomaly		$\boldsymbol{m}$ _stock		$\boldsymbol{m}$_nonlinear		$\boldsymbol{m}$_Tree		$\boldsymbol{m}$_Mixed		$\boldsymbol{m}$_nonlinear and trees	
	Shrinkage level											
	0th	100th	0th	100th	0th	100th	0th	100th	0th	100th	0th	100th
X24				-1.358				0.199				
X26						-2.416**	-1.705	-2.475**				-1.62
X27								-1.028				
X28						-0.597		-0.272				
X29						0.7	1.062	1.375				0.938
X32				-0.694			0.196	0.764		0.694	0.134	-0.024
X33						0.359		-1.095				
X34						-0.814		-0.024				
X35						-0.401		-1.629				
X36						-0.276	1.543	1.510				
X37								0.068				
X38							-1.087	-1.341			-0.248	1.036
X40						0.956	0.110	-1.343		-0.169		
X46						-0.987	-0.410	-0.200				-0.31
X47						1.202	0.137	0.608		0.392	0.973	1.377
X48								-1.645				
X49				0.350				-0.057				
X50						-1.415	-0.281	-0.324		-2.008**	-1.972*	-1.162
X54						-0.08		0.131				
X56								0.059				
X57						-1.733	-0.268	-0.507		-0.859		-1.091
X58				2.201**		0.926		-0.684				
X59						0.42	0.411	0.437				
X61						-0.558	0.253	-0.112				0.118
X63							0.916	1.14				
X64				-0.916			-0.426	-1.009				
X69				-0.121			0.049	-0.151				
X70							1.37	0.337				
X71				-1.342		-0.361	0.424	-0.07				-0.214
X73						-0.206	-1.695	-1.174				
X74				-0.608			1.1	1.308				
X75							0.117	1.081				1.09
X77				3.024***		-3.124***	1.202	1.182				
X78						0.01	-1.595	-1.557				
X79				2.797***		2.889***		0.665				
X81				-1.398		-1.499		0.751				

Nonlinear factor	$\boldsymbol{m}$_anomaly		$\boldsymbol{m}$ _stock		$\boldsymbol{m}$_nonlinear		$\boldsymbol{m}$_Tree		$\boldsymbol{m}$_Mixed		$\boldsymbol{m}$_nonlinear and trees	
	Shrinkage level											
	0th	100th	0th	100th	0th	100th	0th	100th	0th	100th	0th	100th
X82						0.609	0.453	1.791*		1.891		0.885
X84						0.321	0.818	1.151				1.658
X85						-1.199		-2.731**				
X86				-0.986		-0.132	-1.689	-1.949*				
X87						-0.492		-0.494				-0.634
X88							-0.951	-1.254				
X90				0.727		-0.227	-0.152	-1.167				
X91							-0.98	-1.383				
X92				0.311				-1.008				
X93						2.331**	-0.349	-0.697				
X94						-0.777	-0.181	0.06		-0.864		-0.993
X95				-1.3		-0.396	1.524	0.893				
X96				-1.224			-0.633	-1.356				
X97				-1.194		-2.635**		-1.258		-2.492		
X98							1.817	1.729				
X99								0.926				
X100						0.078	0.493	0.55				
X101								0.299				
X102								-1.866*				
X104							0.087	-0.632		0.199	-0.399	-0.596
X105						2.662**	1.19	0.794				
X106							-0.377	0.144			-1.036	-0.261
X107				-0.98		2.381**		-1.088				
X108						-1.585	0.098	0.931				
X110						-0.944	0.464	-0.816		-1.321	-1.382	-1.545
X111						1.678	1.985	1.426		1.736	1.975	1.61
X112						-0.795	-1.433	-2.488**				
X113						-0.082		-0.427				
X116						1.476	0.793	1.292				
X117							0.698	0.128				
X118				1.238			-0.808	-0.047				
X119						-0.715	1.588	1.286				
X122				1.277				-0.922				
X123							-1.175	-0.748				
X125						1.051	2.111**	2.338**				
X126						-0.515	0.883	0.673				-0.391

Nonlinear factor	$\boldsymbol{m}$_anomaly			$\boldsymbol{m}$ _stock			$\boldsymbol{m}$_nonlinear			$\boldsymbol{m}$_Tree			$\boldsymbol{m}$_Mixed			$\boldsymbol{m}$_nonlinear and trees	
	Shrinkage level																
	0th	100th		0th	100th		0th	100th		0th	100th		0th	100th		0th	100th
X127					-0.045			-1.01			0.538						
X128											0.679						
X133					-1.231			1.352		1.555	2.618**						
X134								-0.122		0.382	-0.039			1.538		1.145	0.83
X135					1.65			-0.989		-0.468	-0.807						
X137					-0.888					-0.642	-1.116						
X139											-0.093						
X142											0.406						
X143								1.043			0.019						
X145								1.069			0.902						
X146					1.099			-0.643			-1.094						
X147								1.437			0.922						
X148					-0.302			-2.236**		-2.512**	-3.108***						
X150					-3.06***						0.254						
X152								2.16**		1.547	0.522						
X154										-0.915	-0.533						
X155					-1.196			-0.161		-0.534	0.552						
X158					-0.423			0.425			-0.698						
X159								-1.317			0.005						
X161								-0.614		0.362	0.819						
X164										1.403	1.284						
X168					-2.061			-0.667			-0.222						
X170								-1.697			-2.628**						
X171					1.107						0.108						
X172								-0.686			-0.652						
X173										-1.168	-1.422						
X175								0.523			1.78*						
X176											-0.295						
X181								-0.22			0.294						
X183											1.221						
X185											0.825						
X186								0.02		1.228	0.902						
X187								-1.046			-0.218						
X188					-0.107			-0.203		-0.594	-1.625						
X189								-0.205			-0.852						-2.12**
X191					-0.128			0.91			0.469						

Nonlinear factor	$\boldsymbol{m}$_anomaly			$\boldsymbol{m}$ _stock			$\boldsymbol{m}$_nonlinear			$\boldsymbol{m}$_Tree			$\boldsymbol{m}$_Mixed			$\boldsymbol{m}$_nonlinear and trees	
	Shrinkage level																
	0th	100th		0th	100th		0th	100th		0th	100th		0th	100th		0th	100th
X193								-0.884		-0.245	-1.232						-1.258
X195					0.04					-2.023**	-2.27**						
X196								-0.668			-1.517						
X198										1.192	0.267						
X199										0.749	0.502						0.735
X200											0.863						
X202								-0.106		-0.856	-0.33						
X203								0.047		-0.28	-0.459						
X204											-0.366						
X210								0.681		1.051	0.917						
X211					-0.668						-0.228						
X213					0.537						0.234						
X214											0.03						
X215											1.759*						
X216								-1.082			-1.497						
X217					-0.855			0.067			0.47						
X218								-1.156			-1.235						

Internet Appendix A: Implementation of the agnostic method

The regression analysis:

From the equation in the model setup, when we have data with finite sample (small T),

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\otimes\mathbf{F}(\mathbf{Z_{t-1}}) = \frac{1}{T}\sum_{t=1}^{T} m_t(\mathbf{Z}_{t-1})\widetilde{\mathbf{R}}_t + \boldsymbol{\varepsilon}$$

The term $\boldsymbol{\varepsilon}$, an MN by one vector, represents the pricing error in the finite sample. This is standard regression, with augmented returns ($\widetilde{\mathbf{R}}_t$) from time 1 to time T as the independent variables, and $\frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\otimes\mathbf{F}(\mathbf{Z_{t-1}})$ as the dependent variable. The coefficients to be estimated are $m_t(\mathbf{Z_{t-1}})$ for t=1 to T.

Note that if we do not use the economic state variables, the regression above reduces to the following equation (we will call it the reduced form regression):

$$\mathbf{1} = \frac{1}{T}\sum_{t=1}^{T} m_t\mathbf{R}_t + \boldsymbol{\varepsilon}$$

In this case, we still estimate the coefficients m_t. If the number of assets (N) is higher than the number of periods (T), the coefficients can be identified.

The detailed implementation is as follows. Regress: $\sum_{t=1}^{T} \mathbf{1}\otimes\mathbf{F}(\mathbf{Z_{t-1}})$ on $\mathbf{R_t}\otimes\mathbf{F}(\mathbf{Z_{t-1}})$. Specifically, we run the following regression:

$$\frac{1}{T}\sum_{t=1}^{T} f\left(Z_{t-1}^{k}\right) = \frac{1}{T}\left(R_{i1}f\left(Z_{0}^{k}\right)m_1 + \cdots + R_{it}f\left(Z_{t-1}^{k}\right)m_t \cdots + R_{iT}f\left(Z_{T-1}^{k}\right)m_T\right) + \varepsilon_{ift}.$$

Here R_{it} represents the return for stock i at time t, Z_{t-1}^{k} is the conditioning information k at time t-1 (we assume that there are K conditioning information and together with the constant term, which is denoted by Z^0; thus the dimension of $\mathbf{Z_{t-1}}$ is (k+1) by 1), and f is the function for conditioning information. Given that the conditioning information, functional form and returns are known from the data, our goal is to estimation coefficients m_1 through m_T.

Suppose we start with the identity function only, i.e., $f\left(Z_{t-1}^{k}\right) = Z_{t-1}^{k}$. The regression can reduce to

$$\frac{1}{T}\sum_{t=1}^{T} Z_{t-1}^{k} = \frac{1}{T}\left(R_{i1}Z_{0}^{k}m_1 + \cdots + R_{it}Z_{t-1}^{k}m_t \cdots + R_{iT}Z_{T-1}^{k}m_T\right) + \varepsilon_{ift}.$$

Hence, the regression can be written as

$$\mathbf{Y} = \mathbf{Xm} + \mathbf{E},$$

where $\mathbf{Y} = [1, \cdots, 1, \frac{1}{T}\sum_{t=1}^{T} Z_{t-1}^{1}, \cdots \frac{1}{T}\sum_{t=1}^{T} Z_{t-1}^{1}, \cdots, \frac{1}{T}\sum_{t=1}^{T} Z_{t-1}^{K}, \cdots \frac{1}{T}\sum_{t=1}^{T} Z_{t-1}^{K}]'$, which is an N(K+1) by 1 vector.

$$\mathbf{X} = \begin{bmatrix} \frac{1}{T}(R_{11}) & \frac{1}{T}(R_{12}) & \cdots & \cdots & \frac{1}{T}(R_{1t}) & \cdots & \cdots & \cdots & \frac{1}{T}(R_{1T}) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{T}(R_{N1}) & \frac{1}{T}(R_{N2}) & \cdots & \cdots & \frac{1}{T}(R_{Nt}) & \cdots & \cdots & \cdots & \frac{1}{T}(R_{NT}) \\ \frac{1}{T}(R_{11}Z_0^1) & \frac{1}{T}(R_{12}Z_1^1) & \cdots & \cdots & \frac{1}{T}(R_{1t}Z_{t-1}^1) & \cdots & \cdots & \cdots & \frac{1}{T}(R_{1T}Z_{T-1}^1) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{T}(R_{N1}Z_0^1) & \frac{1}{T}(R_{N2}Z_1^1) & \cdots & \cdots & \frac{1}{T}(R_{Nt}Z_{t-1}^1) & \cdots & \cdots & \cdots & \frac{1}{T}(R_{NT}Z_{T-1}^1) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{T}(R_{N1}Z_0^K) & \frac{1}{T}(R_{N2}Z_1^K) & \cdots & \cdots & \frac{1}{T}(R_{Nt}Z_{t-1}^K) & \cdots & \cdots & \cdots & \frac{1}{T}(R_{NT}Z_{T-1}^K) \end{bmatrix}$$

Here the matrix **X** is N(K+1) by T. If N(K+1)>T, we can identify the coefficients $\mathbf{m} = [m_1, \cdots, m_T]$. And **E** is the estimation error.

Internet Appendix B: Justification of three empirical applications

In this section, we show that three applications do ***not*** depend on the bias of the SDF.

Justification for application 1:

Following Kim (2017), the estimated the SDF converges to $\mathrm{T}(\sigma^2\mathbf{I})^{-1}\mathbf{F}'((\boldsymbol{\beta}'\boldsymbol{\beta})^{-1} + \mathbf{F}(\sigma^2\mathbf{I})^{-1}\mathbf{F}')(\boldsymbol{\beta}'\boldsymbol{\beta})^{-1}(\boldsymbol{\beta}'\mathbf{1})$, as N converges to infinity. If a candidate factor, $\mathrm{F_e}$, is correlated with $\mathbf{F}$; it can be written as $\mathrm{F_e} = \mathrm{a} + \mathrm{b}\mathbf{F} + \varepsilon$. We want to show that the $\mathrm{F_e}$ is correlated with the SDF almost surely. In case that they are uncorrelated,

$$\mathrm{cov}(\mathrm{F_e}', \mathrm{T}(\sigma^2\mathbf{I})^{-1}\mathbf{F}'((\boldsymbol{\beta}'\boldsymbol{\beta})^{-1} + \mathbf{F}(\sigma^2\mathbf{I})^{-1}\mathbf{F}')(\boldsymbol{\beta}'\boldsymbol{\beta})^{-1}(\boldsymbol{\beta}'\mathbf{1})) = 0.$$

Define $\mathrm{T}((\boldsymbol{\beta}'\boldsymbol{\beta})^{-1} + \mathbf{F}(\sigma^2\mathbf{I})^{-1}\mathbf{F}')(\boldsymbol{\beta}'\boldsymbol{\beta})^{-1}(\boldsymbol{\beta}'\mathbf{1}) \equiv \mathbf{A}$. Then the equation can be written as $\mathrm{cov}(\mathrm{F_e}', (\sigma^2\mathbf{I})^{-1}\mathbf{F}'\mathbf{A}) = 0$.

Plugin $\mathrm{F_e}$,

$$\mathrm{cov}(\mathrm{F_e}', (\sigma^2\mathbf{I})^{-1}\mathbf{F}'\mathbf{A}) = \mathrm{cov}\big((\mathrm{a} + \mathrm{b}\mathbf{F}' + \varepsilon), (\sigma^2\mathbf{I})^{-1}\mathbf{F}'\mathbf{A}\big) = \mathrm{acov}(1', (\sigma^2\mathbf{I})^{-1}\mathbf{F}'\mathbf{A}) + \mathrm{bcov}(\mathrm{F}', (\sigma^2\mathbf{I})^{-1}\mathbf{F}'\mathbf{A}) = 0.$$

Given that $\mathrm{cov}(1', (\sigma^2\mathbf{I})^{-1}\mathbf{F}'\mathbf{A})$ and $\mathrm{cov}(\mathbf{F}', (\sigma^2\mathbf{I})^{-1}\mathbf{F}'\mathbf{A})$ are fixed, the solution (a, b) of the above equation spans a K-dimensional hyperplane of a K+1 dimensional space (which is the space of all $\mathrm{a} + \mathrm{b}\mathbf{F}$). From the above analysis, the space of $\mathrm{F_e}$ that can be uncorrelated with the SDF has a zero Euclidean measure, or $\mathrm{F_e}$ is correlated with the SDF almost surely.

Justification for application 2:

We can write the covariance between the SDF and return R as:

$$\mathrm{cov}(\widehat{\mathrm{m}}, \mathbf{R}') = \mathrm{cov}(\mathrm{T}(\sigma^2\mathbf{I})^{-1}\mathbf{F}'((\boldsymbol{\beta}'\boldsymbol{\beta})^{-1} + \mathbf{F}(\sigma^2\mathbf{I})^{-1}\mathbf{F}')(\boldsymbol{\beta}'\boldsymbol{\beta})^{-1}(\boldsymbol{\beta}'\mathbf{1}), \mathbf{F}'\boldsymbol{\beta}' + \boldsymbol{\varepsilon}') = \mathrm{Tcov}((\sigma^2\mathbf{I})^{-1}\mathbf{F}'\mathbf{A}, \mathbf{F}')\boldsymbol{\beta}'.$$

Define $\mathrm{Tcov}((\sigma^2\mathbf{I})^{-1}\mathbf{F}'\mathbf{A}, \mathbf{F}') \equiv \mathbf{L}$, which is a constant vector of dimension 1 by k. Then $\mathrm{cov}(\widehat{\mathrm{m}}, \mathbf{R}')$ is a linear combination of beta ($\boldsymbol{\beta}$). Following IPCA setup, the factor loadings are a linear combination of firm characteristics, i.e., $\boldsymbol{\beta} = \mathbf{C}\boldsymbol{\gamma} + \vartheta$, where $\mathbf{C}$ is the firm characteristics, $\boldsymbol{\gamma}$ is the coefficient and ϑ is the estimation error. Replacing beta in the above equation, $\mathrm{cov}(\widehat{\mathrm{m}}, \mathbf{R}') = \mathrm{L}(\boldsymbol{\gamma}'\mathbf{C}' + \vartheta')$. Hence, following the similar argument in application 1, if a characteristics γ is correlated with the true $\mathrm{cov}(\mathrm{m}, \mathbf{R}')$, the correlation between $\mathrm{cov}(\widehat{\mathrm{m}}, \mathbf{R}')$ and γ should be nonzero almost surely.

Justification for application 3:

Assume that the market is integrated. Then the factors used to construct SDFs from different assets should be the same. Write the covariance between SDFs as:

$$\mathrm{cov}(\widehat{\mathrm{m}}_1, \widehat{\mathrm{m}}_2) = \mathrm{cov}((\sigma^2\mathbf{I}_1)^{-1}\mathbf{F}'\mathbf{A}_1, (\sigma^2\mathbf{I}_2)^{-1}\mathbf{F}'\mathbf{A}_2).$$

Following the same argument from application 1, we can show that the covariance should be nonzero almost surely if $\mathrm{cov}(\mathrm{m}_1, \mathrm{m}_2) \neq 0$ (the covariance between true SDFs is nonzero).

Internet Appendix C: Regularization methods

We apply three regularization techniques to alleviate overfitting issue and testing the whether the SDF is correlated with known factors. Regularization involves adding some noise to the objective function of the model before optimizing it. That is, we add a penalty on the different parameters of the model. By adding this penalty to reduce freedom of the model, we can reduce fitting of the noise to the training data and make it more generalized. With regularization, we want to minimize loss and complexity, which is a penalty term

$$\min(\mathrm{Loss}(\mathrm{data}|\mathrm{Model}) + \mathrm{complexity}(\mathrm{model}))$$

There are three regularization methodologies:

1. Lasso

$$\sum_{i=1}^{n}\left(y_i - \sum_{j-1}^{p} x_{ij}\,\beta_j\right)^2 + \boldsymbol{\alpha}\sum_{j=1}^{p}|\beta_j|$$

Lasso adds "absolute value" of coefficient as penalty term to the loss function. $\boldsymbol{\alpha}$ is the regularization parameter that we provide as an input to the model. Increase in $\boldsymbol{\alpha}$ results in reduced overfitting. Ridge regression

$$\sum_{i=1}^{n}\left(y_i - \sum_{j-1}^{p} x_{ij}\,\beta_j\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2$$

Lambda is also called regularization rate. We multiply the regularization term by lambda (scalar) that tunes the overall impact of regularization. Increasing the lambda value strengthens the regularization effect and vice versa.

It is important to choose the value of Lambda. If Lambda is very large, it will add too much weight and lead to under-fitting. The L2 regularization technique works well to avoid the over-fitting problem.

2. Elastic Net

$$\boldsymbol{\sum_{i=1}^{n}\left(y_i - \sum_{j-1}^{p} x_{ij}\,\beta_j\right)^2 + \alpha\sum_{j=1}^{p}|\beta_j| + \lambda(1-\alpha)\sum_{j=1}^{p}\beta_j^2}$$

Lambda is the shared penalization parameter. Alpha is used to set the ratio between Lasso and Ridge regularization.