Asset Pricing with Slope Factors: Model and Evidence of Outperformance

Siddhartha Chib[1], **Yi Chun Lin**[2], **Kuntara Pukthuanthong**[3], **Xiaming Zeng**[4]

June 2022

[1] Olin School of Business, Washington University in St. Louis, 1 Bookings Drive, St. Louis, MO 63130. E-mail: chib@wustl.edu

[2] Department of Economics, Washington University in St. Louis, 1 Bookings Drive, St. Louis, MO 63130. E-mail: l.yichun@wustl.edu

[3] Department of Finance, Trulaske College of Business, University of Missouri, Columbia MO 65203. Email: pukthuantthongk@missouri.edu

[4] Investment professional. E-mail: zengxiaming@wustl.edu

Abstract

We make a case that characteristics-based long-short factors should be constructed by the slope factor method rather than by sorting methods. This is because sorting does not fully control for the influence of omitted characteristics, rendering them more noisy than slope factors. In contrast, slope factors, by virtue of being pure-play, lead to higher volatility stochastic discount factors, and better potential for pricing the cross-section. We show that an eight slope factor based asset pricing model captures risks that are mostly not priceable by existing models. It outperforms on several dimensions: factor risk premia; OOS Sharpe-ratios of tangency portfolios; and pricing of portfolios, ETFs and stocks.

JEL Classification: G11, G12, G14

Keywords: factor risk premia; firm level characteristics; marginal likelihood; pricing test; stochastic discount factor; risk factors; out-of-sample Sharpe-ratio

1 Introduction

It is commonly believed that factors such as HML, SMB, MOM and the scores of other factors that are constructed as the difference in value-weighted returns of portfolios sorted on size and the lagged characteristic represent the risk associated with that characteristic. Under this assumption, a large literature in finance is concerned with whether the factor is priced in the cross-section of returns. But, what if the factor is not quite what it seems? Suppose that the HML factor constructed in this way is not a value factor, but a factor that incorporates the risks of many other characteristics. We argue that besides complicating interpretation, this dependence on other characteristics increases the correlation amongst the factors and makes each factor more volatile. The higher correlation complicates the search for which of these factors is in the SDF, while the excess noise harms the ability of these factors to price the cross-section.

To understand these points, suppose that we aim to construct factors that represent 50 firm-level characteristics. In the sort based method, which we henceforth refer to as the differential method, for each cross-section consisting of a large number of publicly traded firms, and for each characteristic, we sort the firms on size (market cap), say small and big, and on levels of that lagged characteristic (say three levels based on its quantiles). For each characteristic, this classifies the firms into these six groupings. The essential justification for the differential method is that sorting decouples the returns of firms with high and low levels of that characteristic, for a given size, from the excluded characteristics. Under this decoupling assumption, long-short portfolios made from the extremes of that characteristics represent the returns that accrue to that characteristic. But does sorting decouple returns in this way?

Consider, for example, the five characteristics, size (*mve*), value (*bm*), operating profitability (*operprof*), investment (*agr*), and momentum (*mom12m*) that correspond to the Fama and French

(2018) factors. For monthly data on a large collection of firms from January 1989 to December 2020 ($T = 384$) we sort the excess returns for that month by size (2 levels) and each characteristic (3 levels). We then combine the data that are used to make the long-short portfolio for that factor, for that month, and run a regression of those excess returns on the excluded 48 characteristics. For each month we run five such regressions, one for each of these characteristics. From each regression we tabulate which of those excluded characteristics are significant at the 0.01 level. We do this check at the 0.01, rather than a 0.05 threshold, to reduce the chance of false-positives. We repeat these regressions for each month. The results appear in Table 1 where the numbers in the table indicate the frequency of times (divided by 384) that a particular excluded characteristic (given in the row) is significant in those regressions. From this table we can see that the return data that are used to construct the FF6 factors tend to be correlated with several other characteristics. For example, the returns used to construct the HML factor are correlated with the bid-ask spread (*baspread*), the CAPM beta (*beta*), *mom12m* and *mom1m* characteristics. Thus, the HML differential factor captures the risk not only related to *bm*, but also the risks stemming from the *baspread*, *beta* and *mom12m* characteristics.

Table 1 Regression of excess returns used for the construction of each FF6 characteristics-based double sorted differential factor against the excluded 48 characteristics, by month. The numbers in the table (going down each column) are the number of months divided by the total number of months ($T = 384$) that the excluded characteristic (given in the row) is significant at the 0.01 level. Each column has a NA for the characteristic on which the returns in that column are sorted on. Since the returns are always first sorted on size, the regressions never include the size characteristic. The data are from January 1989 to December 2020.

	returns used for mve	returns used for bm	returns used for operprof	returns used for agr	returns used for mom12m
acc	9.38	7.03	8.33	8.33	9.11
age	7.29	6.25	5.47	4.69	5.73
agr	8.33	7.03	4.43	NA	5.73
baspread	35.42	33.59	34.38	34.11	35.42
beta	35.94	36.46	37.76	36.46	38.02
bm	12.5	NA	14.84	12.76	13.54

Table 1 continued:

	returns (mve)	returns (bm)	returns (operprof)	returns (agr)	returns (mom12m)
cash	10.94	13.8	12.76	12.5	11.72
cashdebt	10.94	10.68	10.94	8.59	11.2
cashpr	3.65	4.43	2.6	2.34	3.39
cfp	13.28	11.72	12.24	10.16	11.72
chatoia	5.21	5.73	4.17	4.43	3.91
chcsho	5.21	5.47	5.21	4.43	5.73
chempia	4.17	4.43	3.91	3.39	5.21
convind	2.86	2.86	3.39	2.86	3.91
depr	6.25	4.43	5.21	5.47	5.21
dy	7.03	7.81	7.55	7.81	8.07
egr	8.59	7.81	4.95	6.51	8.59
ep	15.62	15.1	16.15	15.62	13.02
gma	10.42	8.85	13.8	9.38	9.9
grcapx	4.17	3.39	5.21	4.17	5.21
herf	2.86	3.39	4.17	2.6	3.39
hire	7.03	6.77	5.99	5.21	5.73
idiovol	22.66	22.92	19.53	21.35	19.53
ill	23.18	20.31	21.61	19.27	20.31
indmom	15.62	15.89	15.89	20.31	24.74
invest	6.25	8.33	7.03	5.73	6.77
lev	14.84	17.19	12.5	13.8	15.89
lgr	4.69	4.43	5.21	2.86	3.65
mom12m	27.6	33.33	32.03	31.77	NA
mom1m	39.32	38.54	37.5	39.84	38.8
nincr	2.08	2.34	2.34	2.34	2.86
operprof	3.39	3.39	NA	3.65	2.86
pchgm_pchsale	6.25	5.99	5.73	8.33	5.73
pricedelay	4.43	3.39	2.34	3.39	3.65
ps	3.12	3.12	2.08	2.86	3.12
rd	5.73	6.77	4.43	5.47	4.17
roaq	14.84	14.84	10.42	10.42	13.54
roeq	14.84	8.85	8.33	10.68	11.2
roic	9.11	9.9	10.16	10.16	11.98
salecash	1.04	2.86	2.08	1.82	2.08
saleinv	1.3	1.3	0.78	1.3	1.82
salerec	7.29	6.51	8.07	6.25	8.33
sgr	5.21	7.55	5.47	4.95	5.47
sin	1.3	2.34	2.08	1.3	1.82
sp	10.94	10.68	10.16	10.68	10.94
std_dolvol	17.19	16.41	16.41	18.23	17.97
std_turn	23.96	23.18	21.61	24.74	22.92
tang	5.73	5.47	4.95	5.73	6.25
tb	3.65	1.82	1.82	1.82	2.6

There are three main consequences of this dependence of differential factors on other characteristics. One is that the factors constructed in this way are difficult to interpret. An equally significant consequence is that since each factor depends on several intersecting excluded characteristics, the dependence amongst the factors tends to be high. As a result, suppose we were to construct factors of all the fifty characteristics, with the aim of finding the subset of factors that is involved in pricing. The high correlation amongst the factors makes this discovery more difficult in the context of the monthly sample sizes that are encountered in practice. Finally, and vitally, since the returns that make up these factors also depend on uncontrolled excluded characteristics, the combined variation in all of those omitted characteristics produces factors with excess volatility. A consequence of this excess volatility is that the stochastic discount factor (SDF), linear in these factors becomes less variable. And since prices are determined by the covariance of the SDF with next period payoffs (which are typically very variable), this reduced variability of the SDF jeopardizes the ability of SDFs based on these factors to price the cross-section.[1]

To overcome these problems, we argue that factors from characteristics should be made by the slope factor method that was first described in Fama (1976). Apart from the recent work in Back, Kapadia, and Ostdiek (2013), Back, Kapadia, and Ostdiek (2015) and Fama and French (2020), this method has rarely been used in the literature. We feel that this neglect is entirely undeserved. The slope factor method is a stand-out method that has none of the limitations of the differential construction method. Essentially, in the slope factor method, one runs cross-sectional regressions of excess returns on lagged standardized characteristics. From properties of the OLS estimates, it can be shown that the estimated OLS slopes from these regressions are automatically long-

[1] The same points apply to the rank factor construction method used by Asness, Frazzini, and Pedersen (2019), Kelly, Pruitt, and Su (2019), Chen, Pelger, and Zhu (2020), Freyberger, Neuhierl, and Weber (2020), and Kozak, Nagel, and Santosh (2020). After grouping excess returns by market cap, factors are constructed using the normalized rank of lagged characteristics as weights. To avoid duplication, we do not discuss this approach for constructing factors further in this paper.

short portfolios that give unit weighted exposure to each lagged characteristic and zero weighted exposure to the other lagged characteristics in the regression. Thus, these factors are pure-play representing precisely the risk of the underlying characteristic, purged of the effect of the other characteristics. Because of this pure-play property, the slope factors also tend to be much less mutually correlated than differential factors and, crucially, much less noisy than differential factors (since the effect of other characteristics has been purged out). As a bonus, constructing factors by this method is a simple exercise in regression.

This is also the first paper in which slope factors are constructed from a large pool of initial characteristics, going beyond the five slope factors in Fama and French (2020). In fact, it is only by starting from such a large pool of characteristics that one can take full advantage of the pure-play property. The motivation for this is the same as in any regression analysis. By including more useful controls one gets better estimates of the slopes, hence, improved slope factors. A by-product of constructing improved slope factors is that one can readily observe the difference in noise between the differential and slope factors.

In an effort to realize the potential of slope factors, we provide in this paper the first slope factor based asset pricing model that is built from the ground up. We do not simply take an existing asset pricing model and replace the differential factors by slope factors, because such a model can be easily improved. We call our new model the CLPZ8 slope factor model. The CLPZ8 model consists of the market factor plus seven slope factors, $\mathrm{BASPREAD}_{\mathrm{S}}$, $\mathrm{BETA}_{\mathrm{S}}$, $\mathrm{EGR}_{\mathrm{S}}$, $\mathrm{MOM1M}_{\mathrm{S}}$, $\mathrm{MVE}_{\mathrm{S}}$, RD_{S}, $\mathrm{ROAQ}_{\mathrm{S}}$ (from our pool of fifty) that we infer are in the stochastic discount factor (SDF). We apply a novel Bayesian method, motivated by machine learning precepts, to discover these eight risk factors. We start by being agnostic about which factors are in the stochastic discount factor, just as in Kozak et al. (2020), but instead of using a regularized likelihood procedure to estimate the SDF coefficients conditioned on a plug-in estimate of the fifty one times fifty one factor covariance matrix, we extend their method to learn about the SDF coefficients from a full Bayesian procedure in which the factor risk premium and factor covariance matrix are both assumed unknown. We conduct this analysis under the regularizing prior of Chib, Zeng, and Zhao (2020), a multivariate-student-t likelihood, and a strong 0.99 posterior credibility decision rule to infer which SDF coefficients are non-zero. Then, in the second step, to catch the few false-positives that may have slipped through from the first step, we apply the model scan methodology of Chib and Zeng (2020) on the factors that passed the first step to find the best asset pricing model.

Six of the seven non-market risk factors in the CLPZ8 model have significant risk premia, in stark contrast to the insignificant factor risk premia in existing models. For instance, in the FF6 model, two of the five non-market factors have significant factor risk premia; in the q5 model, two of the four are significant; and in the DHS model, one out of two is significant. One implication of the significant CLPZ8 factor risk premia is that tangency portfolios of the CLPZ8 factors generate substantially larger OOS realized Sharpe-ratios compared to factor portfolios made from the factors of the current models. Beyond this, the CLPZ8 factors, for the most part, are not priceable by the existing models, showing that we have unearthed risks that are relevant for pricing the cross-section that have been overlooked by the existing models. In the reverse direction, we show that the CLPZ8 model is capable of pricing essentially all of the factors in the existing models.

As one would expect, given the understanding from our previous discussion, the SDF that we infer from the data of the CLPZ8 model is more variable than the SDFs of the FF6, q5 and DHS models. This has consequences for pricing which we document. We also provide evidence of outperformance in pricing a large collection of test assets comprised of 1225 portfolios, 1480 ETFs and 6024 stocks.

The rest of the paper is organized as follows. In Section 2 we describe the construction of slope factors and draw contrasts between the slope factors and the differential factors. In Section 3 we provide a new slope factor asset pricing model and the statistical discovery process that led to this model. Section 4 illuminates aspects of this model, posterior distributions of the factor risk premia and of the SDF underlying the model. In Section 5, the pricing performance of the new model is illustrated on a large collection of portfolios, ETFs and stocks. Section 6 concludes.

2 Slope factors

2.1 Data

We collect monthly stock returns data from CRSP. The set of characteristics are those considered in Green, Hand, and Zhang (2017) and Gu, Kelly, and Xiu (2020), and are sourced from Compustat and I/B/E/S. Our data contain information from 14,860 firms on 50 characteristics for the period January 1989 to December 2020.

For our analysis, we begin the sample from January 1989, which is the earliest month for which complete data on our selected characteristics are available in the I/B/E/S data set. Our aim is to be able to estimate cross-sectional regressions for firms that have a complete set of characteristics in the preceding cross-sections. One common approach is to only analyze firms with non-missing values of all fifty characteristics in each cross-section. This approach, however, sacrifices a large portion of the data that may contain valuable information. We apply a simple new non-parametric approach to deal with the missing data that helps to avoid omitting this valuable information. We proceed as follows. In each cross-section, we first classify firms into two groups, small and big, by the median value of firm size, and then within each size group, we further categorize each firm into five industry groups based on its SIC4 code. In this way, each firm is uniquely assigned to one of $2 \times 5 = 10$ groups. Then, if any firm has missing value for a particular characteristic, we replace the missing value with the group mean of that characteristic from firms within the same group. This imputation procedure is based on the assumption that firms of similar size within the same industry have similar characteristics. These steps produce a rich collection of cross-sectional data sets in which the minimum number of firms is 2051 and the maximum number of firms is 5184. We summarize the data, by characteristics, in Table 10.

2.2 OLS estimates are factors

As first stated in Fama (1976), the OLS estimates of the coefficients in cross-sectional regressions of excess returns on standardized lagged characteristics are *pure-play* long-short portfolios. Specifically, these OLS coefficients give unit weighted exposure to each standardized lagged characteristic in the cross-section regression *and* zero weighted exposure to all the other standardized lagged characteristics. Thus, these OLS estimates are characteristic specific long-short portfolios that load entirely on that characteristic. Since this fundamental property is not that well known, we start by providing a simple explanation.

Suppose that the tth cross-section consists of n_t firms that are independently sampled from the population of firms at time t. Let $\boldsymbol{r}_t = (r_{1t}, \dots, r_{n_t,t})$ denote the the sample vector of excess returns and suppose that two firm characteristics, c_1 and c_2 are measured. Let the sample data on these characteristics at time t be denoted by the $n_t \times 1$ vectors, $\boldsymbol{c}_{j,t} = (c_{j,1}, \dots, c_{j,n_t})$, $j = 1, 2$. Assume that $\boldsymbol{c}_{j,t}$ are each standardized by subtracting the respective sample means and dividing by the respective sample standard deviations. In vector-matrix notation, the tth cross-sectional regression is given by

$$\boldsymbol{r}_t = \boldsymbol{X}_t \boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t$$

where $\boldsymbol{X}_t = (\boldsymbol{i}_{n_t}, \boldsymbol{c}_{1,t-1}, \boldsymbol{c}_{2,t-1})$ is a $n_t \times 3$ matrix of sample data on the intercept and the two characteristics and $\boldsymbol{i}_{n_t}$ is a vector of ones. The coefficient vector $\boldsymbol{\beta}_t$ is a 3×1 vector of cross-section specific coefficients and $\boldsymbol{\varepsilon}_t$ is a vector of iid homoskedastic cross-sectional errors. Now consider the OLS estimate of $\boldsymbol{\beta}_t$, namely $\hat{\boldsymbol{\beta}}_t = (\boldsymbol{X}_t'\boldsymbol{X}_t)^{-1}\boldsymbol{X}_t'\boldsymbol{r}_t$. This can be expressed as a linear combination of the excess returns as

$$\hat{\boldsymbol{\beta}}_t = \boldsymbol{W}_t'\boldsymbol{r}_t \tag{1}$$

or as

$$\begin{pmatrix} \hat{\beta}_{0,t} \\ \hat{\beta}_{1,t} \\ \hat{\beta}_{2,t} \end{pmatrix} = \begin{pmatrix} \boldsymbol{w}_0'\boldsymbol{r}_t \\ \boldsymbol{w}_1'\boldsymbol{r}_t \\ \boldsymbol{w}_2'\boldsymbol{r}_t \end{pmatrix}$$

where $\boldsymbol{w}_j'$ is the j^{th} row of $\boldsymbol{W}_t'$. Now from the trivial identity $\boldsymbol{W}_t'\boldsymbol{X}_t = \boldsymbol{I}_3$, where $\boldsymbol{I}_3$ is the 3×3 identity matrix, written out in full as

$$\begin{pmatrix} \boldsymbol{w}_0' \\ \boldsymbol{w}_1' \\ \boldsymbol{w}_2' \end{pmatrix} (\boldsymbol{i}_{n_t}, \boldsymbol{c}_{1,t-1}, \boldsymbol{c}_{2,t-1}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2}$$

we can see, by row-column multiplication, that $\boldsymbol{w}_j$ are proper weights that satisfy the following restrictions,

$$\begin{aligned} &\boldsymbol{w}_0'\boldsymbol{i}_{n_t} = 1;\ \boldsymbol{w}_0'\boldsymbol{c}_{1,t-1} = 0;\ \boldsymbol{w}_0'\boldsymbol{c}_{2,t-1} = 0 \\ &\boldsymbol{w}_1'\boldsymbol{i}_{n_t} = 0;\ \boldsymbol{w}_1'\boldsymbol{c}_{1,t-1} = 1;\ \boldsymbol{w}_1'\boldsymbol{c}_{2,t-1} = 0 \\ &\boldsymbol{w}_2'\boldsymbol{i}_{n_t} = 0;\ \boldsymbol{w}_2'\boldsymbol{c}_{1,t-1} = 0;\ \boldsymbol{w}_2'\boldsymbol{c}_{2,t-1} = 1 \end{aligned} \tag{3}$$

Reading these restrictions row by row, we can now conclude that $\hat{\beta}_{0,t} = \boldsymbol{w}_0'\boldsymbol{r}_t$ is a pure-play long portfolio (its weights $\boldsymbol{w}_0$ sum to one and it gives zero weighted exposure to the other two characteristics); that $\hat{\beta}_{1,t} = \boldsymbol{w}_1'\boldsymbol{r}_t$ is a pure-play long-short portfolio (its weights $\boldsymbol{w}_1$ sum to zero, it gives unit weighted exposure to the first lagged characteristic and zero weighted exposure to the second lagged characteristic); and that $\hat{\beta}_{2,t} = \boldsymbol{w}_2'\boldsymbol{r}_t$ is a pure-play long-short portfolio (its weights $\boldsymbol{w}_2$ sum to zero, it gives zero weighted exposure to the first lagged characteristic and unit weighted exposure to the second lagged characteristic). Thus, $\hat{\beta}_{1,t}$ and $\hat{\beta}_{2,t}$ are characteristic specific long-short portfolios.

It should also be clear from this derivation that regularized estimates, such as the LASSO, or the Bayesian posterior mean with a proper prior, would not satisfy this property.

If we run the preceding cross-sectional regression separately in sequence for $t = 1, 2, ..., T$, then the sequence of OLS estimates $\hat{\beta}_{j,t}$, $t = 1, ..., T$, are a sequence of long-short portfolios that load purely on characteristic c_j, $j = 1, 2$.

2.3 Factors

Given this understanding we now proceed to make these slope factors from cross-sectional regressions with stock premiums on the LHS and 50 lagged characteristics on the RHS

$$r_{it} = \alpha_{it} + \sum_{j=1}^{50} \beta_{j,t} c_{j,it-1} + \varepsilon_{it}, \; i = 1, ...n_t \tag{4}$$

where r_{it} is the excess return of firm i in month t, with t running from January 1989 to December 2020, α_{it} is the intercept, $\beta_{j,t}$ is the slope of the characteristic $c_{j,it-1}$, and ε_{it} is the error term which is assumed to be iid across firms in the cross-section. The total number of firms in the cross-section, n_t, varies across time (see Coqueret (2021) for a theoretical model in which characteristics are a source of return heterogeneity).

As discussed in the derivation above, the RHS variable $c_{j,it-1}$ in these cross-sectional regressions are standardized for each characteristic j within each cross-section. Thus, the sample mean, and standard-deviation, of $c_{j,t-1} = (c_{j,1t-1}, c_{j,2t-1}, ..., c_{j,n_t t-1})$ are zero and one, respectively, for every j and t. Therefore, the lagged variables on the RHS are unit-less, and the slope coefficients on the RHS are in the same units as the stock returns on the LHS.

Let the cross-sectional OLS estimates of $\beta_{j,t}$ be denoted by $\hat{\beta}_{j,t}$, for $j = 1, 2, ...50$. These OLS estimates are the slope factors corresponding to the fifty characteristics at time t. We denote these time t slope factors as $\mathrm{C}_{j,\mathrm{S},t}$, where C_j emphasizes that this is the factor corresponding to characteristic c_j, and S emphasizes that this is the factor constructed by the slope factor method. A sequence of these slope factors is obtained by estimating the cross-sectional regressions for each month t. Finally, we augment the slope factors with the market excess return factor, Mkt. This factor is obtained from the Kenneth French data library.

We provide summary statistics of Mkt and the constructed slope factors in Table 2. The numbers in the table are the monthly returns (in %), along with the median, the standard deviation (sd), and the 0.005 and 0.995 quantiles. Over the span of our data, the market portfolio has a mean monthly return of 0.739% with thick tails running from -13.610% to 11.445%. The monthly return of each of the long-short slope slope factors portfolios is, of course, much smaller. For example, the EGR_S long-short portfolio has a mean monthly return of -0.040% and tails running from -1.596% to 1.366%. In general, the return of these pure-play long-short portfolios is smaller than the return of the corresponding differential factor portfolios because the latter are not pure-play and include the returns stemming from other correlated characteristics. It is important to note the thick tails of these factors, a typical feature of any such data on returns of portfolios constructed from characteristics. In our analysis to discover which factors are in the SDF we deal with this problem by deploying a multivariate-t distribution with small degrees of freedom. This modeling extension to the standard Gaussian assumption is central in seeing through the fog of noise generated by such thick tails.

Table 2 The descriptive statistics of the slope factors
This table presents the descriptive statistics of the slope factors in units of monthly returns (%). The acronym is described in Table 10. Descriptive statistics include the mean, median, standard deviation, and the 0.5%, 99.5% quantiles across the time-series of each slope factor. The data are from January 1989 to December 2020.

	Mean	Median	Std	0.5% quantile	99.5% quantile
Mkt	0.739	1.190	4.362	-13.610	11.445
ACC_S	-0.072	-0.078	0.680	-2.382	1.993
AGE_S	-0.021	-0.006	0.513	-1.572	1.391
AGR_S	-0.147	-0.162	0.798	-2.518	2.400
BASPREAD_S	-0.096	-0.225	1.548	-4.057	6.285
BETA_S	0.029	0.025	1.628	-4.515	5.216
BM_S	0.084	0.070	0.648	-1.708	1.906
CASH_S	0.215	0.171	1.011	-2.475	3.605
CASHDEBT_S	0.020	0.060	0.702	-2.361	1.754
CASHPR_S	-0.016	-0.039	0.400	-1.069	1.180
CFP_S	0.021	0.050	0.778	-2.147	2.116
CHATOIA_S	0.039	0.029	0.444	-1.123	1.159
CHCSHO_S	-0.063	-0.058	0.442	-1.557	1.434
CHEMPIA_S	0.040	0.102	0.571	-1.676	1.540
CONVIND_S	-0.022	-0.035	0.372	-0.933	0.905
DEPR_S	0.007	-0.013	0.470	-1.166	1.327
DY_S	-0.066	-0.102	0.541	-1.177	1.563
EGR_S	-0.040	-0.026	0.551	-1.596	1.366
EP_S	0.077	0.091	0.986	-3.505	2.673
GMA_S	0.077	0.010	0.770	-2.061	2.821
GRCAPX_S	-0.044	-0.048	0.427	-1.240	1.476
HERF_S	-0.047	-0.070	0.427	-1.206	1.614
HIRE_S	-0.040	-0.080	0.685	-1.875	1.871
IDIOVOL_S	-0.081	-0.223	1.259	-3.508	4.091
ILL_S	0.263	0.140	0.817	-1.651	2.898
INDMOM_S	0.236	0.206	0.769	-2.217	2.799
INVEST_S	0.005	-0.022	0.635	-1.625	2.372
LEV_S	-0.010	0.017	0.900	-2.923	2.848
LGR_S	-0.006	0.004	0.597	-1.85	2.163
MOM12M_S	0.066	0.160	1.223	-5.253	2.516
MOM1M_S	-0.609	-0.410	1.275	-6.052	2.038

Table 2 continued: The descriptive statistics of the slope factors

	Mean	Median	Std	0.5% quantile	99.5% quantile
MVE$_S$	-0.281	-0.157	1.314	-5.992	2.694
NINCR$_S$	0.110	0.109	0.344	-0.990	1.000
OPERPROF$_S$	0.022	0.009	0.474	-1.082	1.329
PCHGM_PCHSALE$_S$	0.019	0.053	0.487	-1.511	1.563
PRICEDELAY$_S$	-0.027	-0.022	0.383	-1.227	1.046
PS$_S$	0.029	0.039	0.457	-1.759	1.129
RD$_S$	0.129	0.089	0.467	-0.859	1.708
ROAQ$_S$	0.091	0.201	0.877	-3.257	1.933
ROEQ$_S$	0.089	0.036	0.647	-1.498	2.245
ROIC$_S$	-0.042	0.036	0.837	-4.151	1.920
SALECASH$_S$	-0.047	-0.050	0.341	-0.980	0.800
SALEINV$_S$	0.012	0.006	0.321	-0.836	1.428
SALEREC$_S$	-0.030	-0.049	0.483	-1.171	1.350
SGR$_S$	-0.090	-0.066	0.619	-1.812	1.374
SIN$_S$	0.041	0.022	0.348	-0.945	0.989
SP$_S$	0.019	0.006	0.768	-2.013	3.132
STD_DOLVOL$_S$	-0.106	-0.007	1.053	-3.862	2.323
STD_TURN$_S$	0.043	-0.045	0.867	-1.821	2.530
TANG$_S$	-0.006	-0.009	0.618	-1.811	1.813
TB$_S$	0.021	0.011	0.349	-1.104	1.000

It is important to note that we have constructed our slope factors from a large pool of initial characteristics. This way the resulting factors are as pure-play as we can make them. Adding more characteristics (if they are not relevant) would not be helpful because the factors (the slopes of those characteristics) would be then just noise and would not act as useful controls. One could go in the other direction and start with a limited set of characteristics on the RHS, say just five, as in Fama and French (2020), but then one would get five slope factors whose pure-play property would, by definition, be limited to that set of characteristics. We show that such *limited-pure-play* slope factors are inferior to the slope factors that we have constructed from our broad set of initial characteristics.

2.4 Correlation and volatility differences

Table 3 shows the correlation matrix of five pure-play slope factors, five limited-pure-play slope factors, and five differential factors (made from 3-by-2 sorting) corresponding to the Fama and French (2020) characteristics. From this table, one sees that the pure-play slope factors tend to be less mutually correlated than limited-pure-play slope factors and the differential factors. The highest correlation within the set of the pure-play slope factors is 0.337, that between the slope factor representing the size characteristic, MVE_{S}, and the slope factor representing the momentum characteristic, MOM12M_{S}.

Table 3 The correlation matrix of the five pure-play slope factors, five limited-pure-play slope factors, and five differential factors corresponding to the Fama-French five characteristics

Pure-play slope factors	$\mathrm{MVE}_{\mathrm{S}}$	BM_{S}	$\mathrm{OPERPROF}_{\mathrm{S}}$	$\mathrm{AGR}_{\mathrm{S}}$	$\mathrm{MOM12M}_{\mathrm{S}}$
$\mathrm{MVE}_{\mathrm{S}}$	1				
BM_{S}	0.160	1			
$\mathrm{OPERPROF}_{\mathrm{S}}$	-0.018	0.101	1		
$\mathrm{AGR}_{\mathrm{S}}$	0.016	0.065	0.147	1	
$\mathrm{MOM12M}_{\mathrm{S}}$	0.337	0.054	-0.005	0.067	1
Limited-pure-play slope factors					
$\mathrm{MVE}_{\mathrm{S}}$	1				
BM_{S}	0.523	1			
$\mathrm{OPERPROF}_{\mathrm{S}}$	0.257	0.344	1		
$\mathrm{AGR}_{\mathrm{S}}$	0.023	-0.284	0.002	1	
$\mathrm{MOM12M}_{\mathrm{S}}$	0.060	-0.083	-0.155	-0.003	1
Differential factors	$\mathrm{MVE}_{\mathrm{D}}$	BM_{D}	$\mathrm{OPERPROF}_{\mathrm{D}}$	$\mathrm{AGR}_{\mathrm{D}}$	$\mathrm{MOM12M}_{\mathrm{D}}$
$\mathrm{MVE}_{\mathrm{D}}$	1				
BM_{D}	-0.176	1			
$\mathrm{OPERPROF}_{\mathrm{D}}$	-0.124	-0.006	1		
$\mathrm{AGR}_{\mathrm{D}}$	-0.396	-0.315	0.376	1	
$\mathrm{MOM12M}_{\mathrm{D}}$	-0.409	0.178	-0.132	0.034	1

The pure-play set of slope factors are constructed from cross-sectional regressions that involve the full set of 50 characteristics in Table 10. The limited-pure-play set of slope factors are constructed from cross-sectional regressions that contain the limited set of characteristics, size (*mve*), book-to-market ratio (*bm*), operating profitability ($operprof$), asset growth (*agr*), and momentum ($mom12m$), on the RHS. The differential factors are constructed by double-sorting (3-by-2) high-minus-low method. Sample data are monthly, spanning the period January 1989 - December 2020.

As can be seen from the table, the correlation among the limited-pure-play slope factors and that among the differential factors are, in general, higher than that of the pure-play slope factors. This shows the importance of constructing slope factors from a relevant, broad set of characteristics.

As mentioned above, because slope factors are purged off the influence of other characteristics, while the differential factors are not, the factors from the two methods for the same characteristic are quite different. For example, consider the SMB factor in FF6 and its slope factor counterpart, MVE_S. The time series of this pair of factors is shown in the top panel of Figure 1. The bottom panel of the same figure shows another companion pair, that of the MOM factor in FF6 and the slope factor of 12-month momentum MOM12M_S. While the movement in the series follow roughly the same pattern, the correlation is not perfect. In fact the correlation between the two series is 0.53 in the top panel and 0.75 in the bottom panel. This imperfect correlation between the pairs shows that the differential factors measure something different than the stated underlying characteristic (see the discussion surrounding Table 1). In effect, the differential factors are measuring the effect of several other characteristics, not just the characteristic on which that factor was sorted on.

The second feature that jumps out is the difference in variation. The two FF6 factors have standard deviations (sds) of 3.04 and 4.71, respectively, while the corresponding sds of the slope factors are 1.31 and 1.22, respectively, both substantially smaller. The difference in volatility is particularly acute in the period around 1998-2002. The excess volatility of the differential factors is also due to the *uncontrolled* dependence of each factor on other characteristics. During the period when the volatility of the differential factors is high, the regressions discussed in Table 1 show that the returns that are used to make the SMB and MOM factors depend strongly on more characteristics than in other months of the sample. The excess variation in the differential factors is

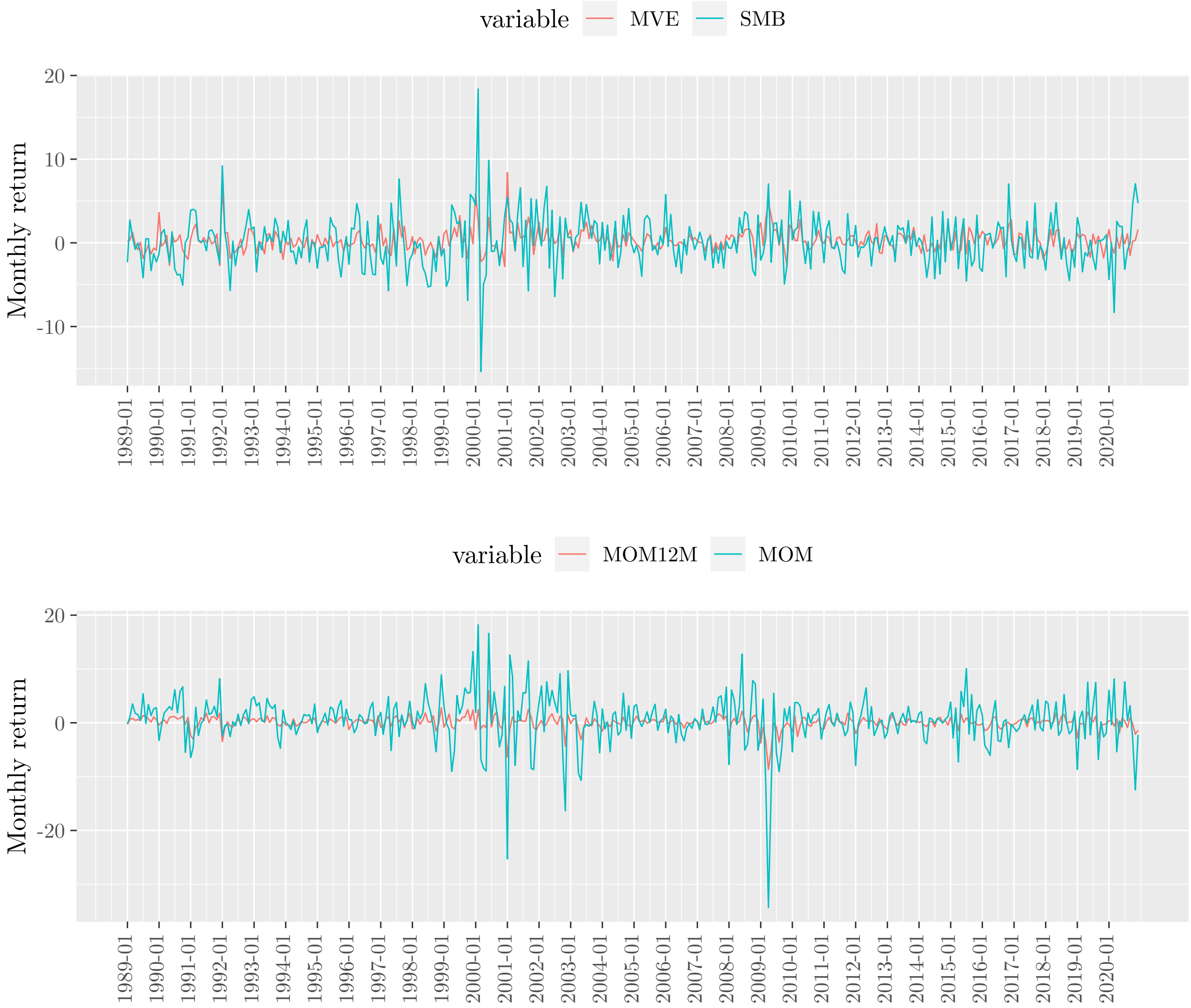

Figure 1 The upper figure shows the slope factor of size, MVE$_S$, and the SMB factor in FF6 (in order to make the former conform to what the latter represents, minus of the former factor is plotted); the lower figure shows the slope factor of 12-month momentum, MOM12M$_S$, and the MOM factor in FF6. The differential factors are more noisy due to the fact that the differential factors are not pure-play. The returns on which these factors are constructed also depend on excluded characteristics, but, since these are uncontrolled, the combined variation in all of those omitted characteristics produces excess volatility.

thus reflecting the combined variation in all of those (uncontrolled for) correlated characteristics. Stated yet another way, this excess noise stems from dependence on a time-varying unknown number of other characteristics that cannot even be fully identified. Ultimately, as will show below, this affects the ability of these factors to price the cross-section.

3 Asset pricing with slope factors

3.1 Motivation

It is our contention that asset pricing with slope factors has the potential to substantially enhance our understanding of the cross-section of expected returns. To see why this might be the case, let $\boldsymbol{x}$: $k \times 1$, denote the long-short risk factors, the factors that are in the SDF. Also let $\mathrm{E}[\boldsymbol{x}] = \boldsymbol{\mu}_x$ and $\mathrm{Var}[\boldsymbol{x}] = \boldsymbol{\Omega}_x$. Now consider the SDF that underpins linear asset pricing models

$$m = 1 - \boldsymbol{\lambda}_x' \boldsymbol{\Omega}_x^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_x), \tag{5}$$

where the first term is $\mathrm{E}[m]$ and $\boldsymbol{\lambda}_x$ is the factor risk premium. Under the SDF pricing restriction

$$\mathrm{E}[m\boldsymbol{x}] = 0 \tag{6}$$

one gets that $\boldsymbol{\lambda}_x = \boldsymbol{\mu}_x$ which identifies $\boldsymbol{\lambda}_x$. Now suppose that $\boldsymbol{r}$ are the *excess* returns on a vector of traded assets in the cross-section. Again, given the SDF, these returns are subject to the restrictions

$$\mathrm{E}[m\boldsymbol{r}] = 0 \tag{7}$$

which on substituting for m and taking expectations leads to unconditional risk premium of $\boldsymbol{r}$

$$\mathrm{E}[\boldsymbol{r}] = \boldsymbol{\Omega}_{rx}\boldsymbol{\Omega}_{x}{}^{-1}\boldsymbol{\lambda}_{x} \tag{8}$$

where $\boldsymbol{\Omega}_{rx}$ is the covariance between $\boldsymbol{r}$ and $\boldsymbol{x}$. Now think of this as the true risk premium and now suppose that holding $\boldsymbol{\Omega}_{rx}$ and $\boldsymbol{\lambda}_{x}$ constant, $\boldsymbol{\Omega}_{x}$ increases (in the sense that the new covariance matrix minus the original covariance matrix is positive semi-definite). Then from the fact that the inverse of the original covariance minus the inverse of the new, higher covariance matrix is positive semi-definite, it follows that the new risk premium is smaller than the original (true risk premium), if not element-by-element, at least in the usual Euclidean vector norm. Thus, a larger $\boldsymbol{\Omega}_{x}$ leads to a smaller risk premium relative to the true risk premium. Thus, when the conditional version of the risk premium model

$$\boldsymbol{r} = \boldsymbol{B}\boldsymbol{x} + \boldsymbol{u} \tag{9}$$

is estimated, one gets a biased estimate of $\boldsymbol{B}$ and, hence, a biased estimate of the risk premium, which shows up as mispricing.

The other problem with larger variability of factors is that it hinders discovery of the true risk factors from the data and increases the possibility that the risk factors identified from the data are incorrect, or put another way, increases the possibility that the SDF is misspecified. Due to this, once again, one would expect that there would be an increase in mispricing of the cross-section.

3.2 Slope factor model discovery

To realize the potential of slope factors, we turn to the question of how to screen the market plus fifty factors for inclusion in an asset pricing model (equivalently, for inclusion in the SDF). It is not possible to directly apply the recent approach of Chib and Zeng (2020) because the number of models that would need to be compared as part of that approach would be $2^{51} - 1$, which would be too large to handle. Instead, we discover our new asset pricing model by a methodology that combines the methodologies in Kozak et al. (2020) and Chib and Zeng (2020). This combined methodology is novel in that it does joint full Bayesian inference on all parameters (thus advancing on the former paper) and handles a large initial pool of factors (thus advancing on the latter).

We employ a three-step discovery process that is motivated by machine learning precepts. At the outset, we start by being agnostic about which factors are in the SDF, but we discipline our search by injecting the implied pricing restrictions (that the mean is equal to the factor risk premium), regularizing prior information, and a high threshold for inclusion in the SDF. We then use a Bayesian method to derive the marginal posterior distributions of the SDF coefficients. We use these posterior distributions to determine which factor risks are strongly supported by the data. Then, in the second step, to mitigate any problem of false-positives, we apply the model scan methodology of Chib and Zeng (2020) to this subset of potential risk factors to find the best asset pricing model. Finally, in the third step, we reconstruct the factors by purging them of the influence of the included important characteristics, and not also of the characteristics that were found to be not important.

As we show below, the model we discover is different from others in the literature. This is partly because we have searched broadly for useful factors, in the manner of machine learning methods. It is also partly because, as shown above, slope factors are just different from differential factors. What may be true for a particular differential factor (that is present or not present in the SDF) need not be true for the corresponding slope factor.

To keep the focus on the bigger picture (that concerning the usefulness of this new asset pricing model), we only briefly sketch the methodology. An implementation of the approach is available in a user-friendly R package. The user just has to supply the data. The package takes that data, calculates the prior from a training sample, implements the steps below and outputs the best model.

Step 1

Let $\boldsymbol{f}_t$ consist of the market factor plus the fifty constructed slope factors. An agnostic starting point is to suppose that $\boldsymbol{f}_t$ is in the SDF, estimate a model for $\boldsymbol{f}_t$ under the pricing restrictions $\mathrm{E}(\boldsymbol{f}_t) \triangleq \boldsymbol{\mu}_f = \boldsymbol{\lambda}_f$, the factor risk premia, and let the data refine the initial assumption about the risk factors. With this in mind, we begin our analysis by estimating the model

$$\underset{(51\times1)}{\boldsymbol{f}_t} = \underset{(51\times1)}{\boldsymbol{\lambda}_f} + \underset{(51\times1)}{\boldsymbol{\varepsilon}_t}$$

where we have underset the dimensions of the variables below the variables. Given what we know about factors, we make the (safe) assumption that the errors are independently distributed over time. To protect against thick tails, however, (see the discussion surrounding Table 2) we adopt the assumption that the distribution of the errors is multivariate student-t with ν degrees of freedom, mean 0 and covariance matrix $\boldsymbol{\Omega}_f$. We fix the value of ν to six (the best value of ν from comparing models with ν ranging over the values from three to seven), which leads to a distribution with much thicker tails than the Gaussian. The key parameter of interest in this model is the SDF loading vector

$$\boldsymbol{b}_f = \boldsymbol{\Omega}_f^{-1}\boldsymbol{\lambda}_f$$

Note that we use the convention of subscripting the parameters according to variable on the LHS of the model.

One can use the latter definition to express the factor risk premia as $\boldsymbol{\lambda}_f = \boldsymbol{\Omega}_f \boldsymbol{b}_f$ and then express the factors as $\boldsymbol{f}_t = \boldsymbol{\Omega}_f \boldsymbol{b}_f + \varepsilon_t$, as in Kozak et al. (2020), where this model is estimated by regularized maximum likelihood, but under a Gaussian assumption on the error and under the assumption that $\boldsymbol{\Omega}_f$ is known. This parameterization is problematic, however, if $\boldsymbol{\Omega}_f$ is unknown since in that case the covariance matrix that must be estimated appears in both the mean and the distribution of the noise. Nonetheless, it is important to see this connection between the two parameterizations.

Let $\text{data}_T = (\boldsymbol{f}_1, ..., \boldsymbol{f}_T)$ denote the sample data and let $p(\text{data}_T | \boldsymbol{\lambda}_f, \boldsymbol{\Omega}_f))$ denote the likelihood function (a product of of multivariate-t densities). Also let $\pi(\boldsymbol{\lambda}_f, \boldsymbol{\Omega}_f)$ denote the prior. We adopt the prior of Chib et al. (2020) and fix the hyperparameters (the parameters of the prior) in a training sample, a pragmatic, black-box approach that is highly recommended in situations with a large number of hyperparameters. Up to a normalizing constant, the posterior density, $\pi(\boldsymbol{\lambda}_f, \boldsymbol{\Omega}_f | \text{data}_T)$, given by the product of the likelihood and the prior, is not a recognizable density, but is amenable to straightforward simulation by MCMC methods (Chib and Greenberg, 1996). The output of the simulation is a sample of draws,

$$\{(\boldsymbol{\lambda}_f^{(1)}, \boldsymbol{\Omega}_f^{(1)}), ..., (\boldsymbol{\lambda}_f^{(M)}, \boldsymbol{\Omega}_f^{(M)})\} \tag{10}$$

from the posterior distribution, where, as a default, M is say 20,000. These are the draws that are retained following an ignored burn-in of 1000 cycles. The next step is the point-wise transformation of the the draws into $\boldsymbol{b}_f$, resulting in the sample

$$\{\boldsymbol{\Omega}_f^{(1)-1} \boldsymbol{\lambda}_f^{(1)}, ..., \boldsymbol{\Omega}_f^{(M)-1} \boldsymbol{\lambda}_f^{(M)}\} \tag{11}$$

From Chib (2001), for example, we know that this is automatically a sample from $\pi(\boldsymbol{b}_f | \text{data}_T)$, and can be used to learn about $\boldsymbol{b}_f$.

We have found that we can do the preceding estimation efficiently. This is partly due to the regularizing effects of the prior. The minimal correlation amongst the slope factors also helps. This also means that our analysis on fifty one factors can be scaled up (if more characteristics were available). In contrast, with differential factors a similar analysis, even with fifty one factors, is more complex because of the higher correlation amongst the factors. Thus, in that case discovery of the best model is not that straightforward, especially with the sample sizes one encounters in practice.

The idea now is to infer which elements of $\boldsymbol{b}_f$ are not zero. For this, we use the sample of draws on $\boldsymbol{b}_f$ to construct the posterior quantiles of each component of $\boldsymbol{b}_f$. We then declare a particular component of $\boldsymbol{b}_f$ to be non-zero if its Bayesian 0.99 credibility interval excludes 0. It is important to set a high-bar at this stage to limit false-positives. In simulations with data generated to resemble the actual data, we have found that the 0.99 threshold works well. In Figure 2, we report the marginal 99% credibility intervals for each of the fifty one SDF coefficients. The intervals that exclude zero are indicated in red.

One can observe that the data evidence is quite clear about which coefficients are non-zero. This information helps us infer that, at the 99% posterior credibility level, the SDF coefficients of the following 15 factors,

$$\begin{aligned}\boldsymbol{f}^* = \{&Mkt, \mathrm{AGR_S}, \mathrm{BASPREAD_S}, \mathrm{BETA_S}, \mathrm{CFP_S}, \mathrm{EGR_S}, \mathrm{EP_S}, \mathrm{ILL_S}, \\ &\mathrm{MOM1M_S}, \mathrm{MVE_S}, \mathrm{RD_S}, \mathrm{ROAQ_S}, \mathrm{SGR_S}, \mathrm{STD\_DOLVOL_S}, \mathrm{STD\_TURN_S}\} \qquad (12)\end{aligned}$$

are non-zero.

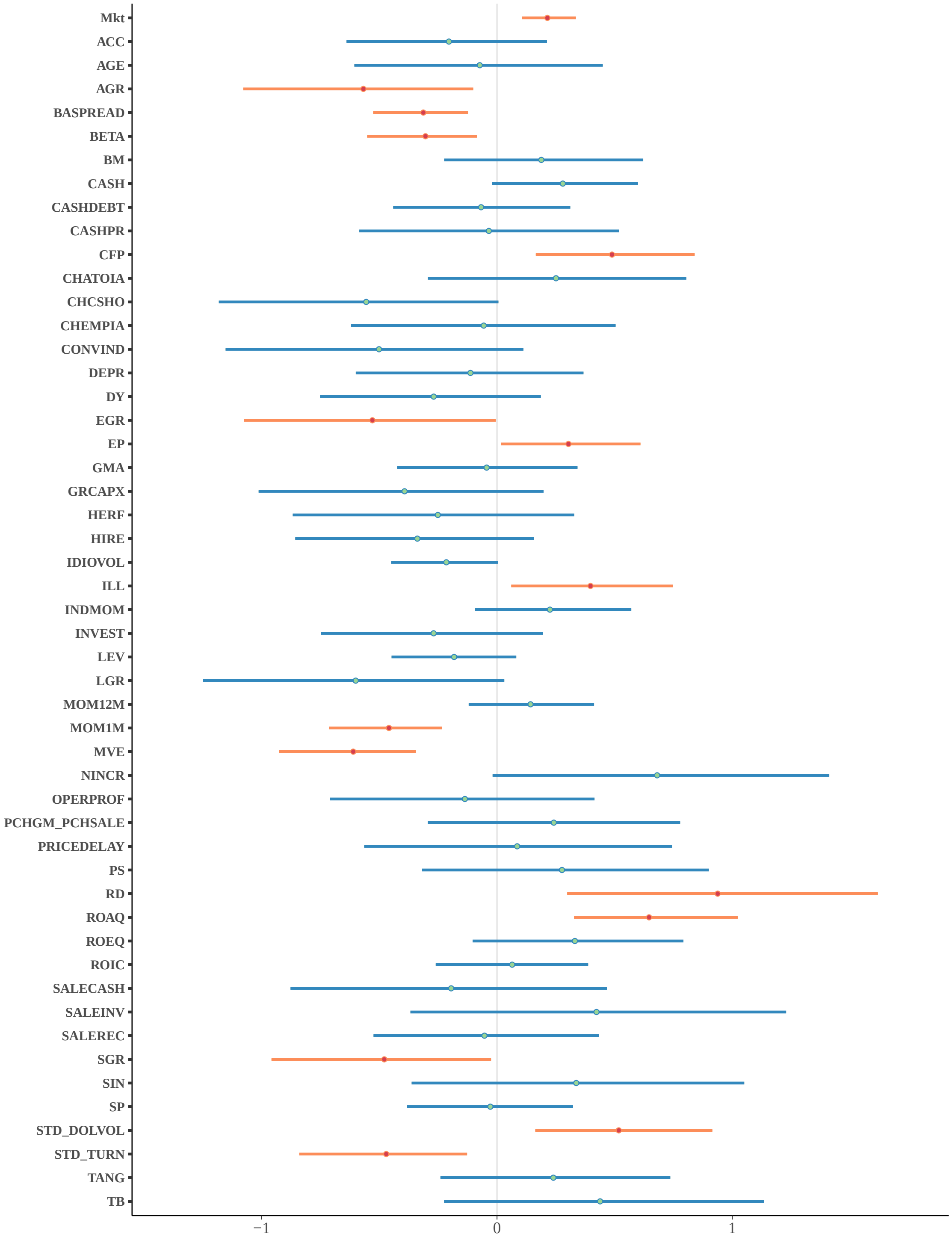

Figure 2 Step 1: Posterior 99% credibility intervals of the SDF coefficients.
This figure presents posterior 99% credibility intervals of the SDF coefficients of *Mkt* and fifty slope factors obtained from estimating the model $\boldsymbol{f}_t = \boldsymbol{\lambda}_f + \varepsilon_t$ on monthly data from January 1989 to December 2020. A normal-inverse Wishart prior was used and the MCMC simulation was run for 20,000 iterations beyond a burn-in of 1000. The posterior sample of $\boldsymbol{b}_f$ was obtained by applying the transformation $\boldsymbol{\Omega}_f^{-1}\boldsymbol{\lambda}_f$ to each sampled parameter. The credibility intervals colored in red exclude zero. See Table 10 for acronyms.

It is premature to stop at this point and think of these factors as the risk factors. Drawing guidance from our simulations, we have found that the output of Step 1 can contain a few false-positives, not many, because the high 0.99 threshold discourages false-positives. On the other hand, the problem of false-negatives is almost not a problem because the 0.99 threshold is tight, but generous enough, to allow all reasonable contending factors to come through. Therefore, to enhance our discovery of risk-factors, we employ another step that is aimed at catching and eliminating any false-positives.

Step 2

The idea behind this step is to exhaustively examine all subsets of $\boldsymbol{f}_t^*$, with each subset a possible set of risk factors. This is the strategy developed in Chib and Zeng (2020) where it is called model scanning. In this methodology, $\boldsymbol{f}_t^*$ is split into a group $\boldsymbol{x}_t^*$ of factors that are in the SDF and a complementary set of factors $\boldsymbol{w}_t^*$ that are priced by $\boldsymbol{x}^*$, but are not in the SDF. The implied model takes the form

$$\boldsymbol{x}_t^* = \boldsymbol{\lambda}_x + \boldsymbol{\varepsilon}_t, \tag{13}$$

$$\boldsymbol{w}_t^* = \boldsymbol{\Gamma}\boldsymbol{x}_t^* + \boldsymbol{u}_t, \tag{14}$$

where the assumption is that the joint distribution of $(\boldsymbol{x}_t^*, \boldsymbol{w}_t^*)$ is multivariate student-t with $\nu = 6$ degrees of freedom. Once again, in light of the thick tails of the factors, and to robustify the procedure, it is important to conduct this analysis under a multivariate student-t assumption. There are 32,767 possible ways of forming these splits. We estimate the model under each split, calculate its marginal likelihood by the method of Chib (1995) (the marginal likelihood is the integral of the likelihood over the prior of the parameters), and rank all the models according to the size of these marginal likelihoods. Because of the importance of this step, we have included a parallelized implementation of this scanning procedure in the R package alluded to above.

We report the top ten models from the model scan in Table 4. Models appear in columns and the factors in the rows. An entry of one in row i and column j of the table indicates that factor i is in M_j, the jth top ranked model.

One can process the information in this table in different ways. One way is to take the factors in the top model as the best risk factors. According to the semiparametric asymptotic theory in Chib, Shin, and Simoni (2018), the highest marginal likelihood model is the best model in the limit. In finite samples, however, it is better to examine the top model *and* the models that are close to the top model, and to select factors that are common to all those models. We define a model as being close to the top model if its posterior odds, relative to the top model, is at least 1:7. The point is that if a model has a posterior odds, relative to the top model, that is even smaller than 1:7, say 1:8, then that model is just too distant to be considered an equal of the top model.

One can now, by elementary algebra, translate the 1:7 posterior odds ratio to the marginal likelihood scale. From the definition of odds, a posterior odds of at least 1:7 means that the posterior probability of the last included model, relative to the top model, has to be at least 0.125. In other words, a model makes the cut if it has a posterior probability of 0.125 (or higher) in the binary comparison with the top model. Equivalently, models that have a lower probability than 0.125 in this binary comparison are just too weak. Now we can convert this threshold probability back into the log-marginal likelihood scale. Let $\text{data}_T^* = (\boldsymbol{f}_1^*, ..., \boldsymbol{f}_T^*)$ denote the sample data on the fifteen factors, and

$$m(\text{data}_T^* | M_j)$$

denote the marginal likelihood of the sample data under M_j, the jth ranked model. Then, from

Bayes theorem, under equal prior probabilities on M_1 and M_j, the posterior probability of M_j is equal to

$$\Pr(M_j|\text{data}_T^*) = \frac{m(\text{data}_T^*|M_j)}{m(\text{data}_T^*|M_j) + m(\text{data}_T^*|M_1)} \tag{15}$$

$$= \frac{1}{1+\exp(d_{1j})} \tag{16}$$

where

$$d_{1j} = \log m(\text{data}_T^*|M_1) - \log m(\text{data}_T^*|M_j)$$

is the difference in log marginal likelihoods of the jth ranked model relative to the top ranked model M_1. On substituting 0.125 on the LHS and solving for d_{1j} we get the simple rule that a model is close to the top model if $d_{1j} \leq 1.95$, or, equivalently, on substituting for the definition of d_{1j}, that

$$\log m(\text{data}_T^*|M_j) \geq \log m(\text{data}_T^*|M_1) - 1.95$$

We now apply this log-marginal likelihood distance measure to the output given in Table 4. The log-marginal likelihood of the top model is -5752.16, which on subtracting 1.95 gives the lower limit of -5754.11 for inclusion in the set of nearby models. Reading across the last row, the last model to make the cut is, therefore, M_8. The idea next is to isolate the factors that are common to the top model and the nearby models. These are the factors that are stable across the top model and the nearby models. As can been seen from the table, there are eight such factors. Denoting them as $\boldsymbol{x}$, these are

$$\boldsymbol{x} = \{Mkt, \text{BASPREAD}_\text{S}, \text{BETA}_\text{S}, \text{EGR}_\text{S}, \text{MOM1M}_\text{S}, \text{MVE}_\text{S}, \text{RD}_\text{S}, \text{ROAQ}_\text{S}\} \tag{17}$$

Table 4 Step 2: Top 10 models from a model scan of 32,767 models
This table shows the top 10 models from a model scan applied to the fifteen factors from Step 1 that have non-zero SDF coefficients at the 0.99 level. Each model in the scan is of the type given in equations (13) and (14). The models are indicated by the columns M_1 to M_{10}. A one in row i and column j of the table indicates that factor i is in the jth top model M_j. Models are ranked by log-marginal likelihoods computed by the method of Chib (1995). The set of nearby models to the top model includes models up to the eighth. These are within 1.95 on the log-marginal likelihood scale to the top model.

Model	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}
Mkt	1	1	1	1	1	1	1	1	1	1
AGR$_S$	1	1	1	1	1	0	1	1	1	1
BASPREAD$_S$	1	1	1	1	1	1	1	1	1	1
BETA$_S$	1	1	1	1	1	1	1	1	1	1
CFP$_S$	1	1	1	0	1	1	0	1	1	0
EGR$_S$	1	1	1	1	1	1	1	1	1	1
EP$_S$	1	1	1	1	0	1	1	1	0	0
ILL$_S$	0	0	0	0	0	0	0	0	0	0
MOM1M$_S$	1	1	1	1	1	1	1	1	1	1
MVE$_S$	1	1	1	1	1	1	1	1	1	1
RD$_S$	1	1	1	1	1	1	1	1	1	1
ROAQ$_S$	1	1	1	1	1	1	1	1	1	1
SGR$_S$	1	1	1	1	1	1	1	0	1	1
STD_DOLVOL$_S$	1	0	1	1	0	0	0	0	1	0
STD_TURN$_S$	0	1	1	0	1	1	1	1	0	1
Log-marg	-5752.16	-5752.32	-5752.95	-5753.72	-5753.72	-5753.87	-5753.98	-5754.05	-5754.20	-5754.32

Thus, according to our discovery procedure, these risk-factors encode the SDF. Even though we started with a high-dimensional model of the SDF, the outcome of the discovery process produces a low-dimensional SDF. Henceforth, we refer to these risk factors and the implied linear asset pricing model, as the CLPZ8 slope factor model.

Extensive simulations have verified that our approach to discovery is remarkably effective. This is backed up by support from the pricing performance of the CLPZ8 model. For instance, the pricing performance of these eight factors, relative to the thirteen factors in the top model, is not materially different. In other words, using the thirteen factors versus the eight does not change the pricing performance of the factors materially. This speaks to the fact that our approach of constructing nearby models to the top model from the model scan, followed by considering factors that are common to those models, is effective. It produces a parsimonious model that prices as well as the top model, a win-win outcome.

Step 3

We conclude the discovery process by reconstructing $\boldsymbol{x}$. We do this by re-running cross-sectional regressions of excess returns on the LHS, but now with only the underlying seven lagged standardized characteristics on the right hand side. Specifically, we run the OLS regressions on excess returns r_{it}

$$\begin{aligned} r_{it} = \alpha_{it} + \beta_{1t}\text{BASPREAD}_{it-1} + \beta_{2t}\text{BETA}_{it-1} + \\ \beta_{3t}\text{EGR}_{it-1} + \beta_{4t}\text{MOM1M}_{it-1} + \beta_{5t}\text{MVE}_{it-1} + \beta_{6t}\text{RD}_{it-1} + \beta_{7t}\text{ROAQ}_{it-1} + \varepsilon_{it} \end{aligned} \quad (18)$$

for each month $t = 1, 2, ..., T$ in our sample. The estimates of the slopes from these regressions are the final characteristic based factors of our model, ie.,

$$\begin{aligned} (\text{BASPREAD}_{\text{S},t}, \text{BETA}_{\text{S},t}, \text{EGR}_{\text{S},t}, \text{MOM1M}_{\text{S},t}, \text{MVE}_{\text{S},t}, \text{RD}_{\text{S},t}, \text{ROAQ}_{\text{S},t}) = (\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t}, ..., \hat{\beta}_{7t}) \\ t = 1, ..., T \end{aligned} \quad (19)$$

The purpose of this reconstruction is to ensure that these seven factors are purged of the influence of the *included* important characteristics, and not also of the characteristics that have been deemed unimportant. This final refitting is a *de-biasing* procedure that is common in regression in other contexts. Our experiments show that this final step improves the pricing performance of the factors. As a side point, quite important for applications, is how easy it is to construct the factors that go into our model. Nonetheless, to broaden accessibility, we have a website that provides these factors.

This website which is linked to the web-page of the first author will be maintained and revised going forward as new data become available.

It should be noted that although our discovery process is data-intensive, in the manner of machine learning methods, the discovery process itself is grounded in finance. The decision rule in Step 1 examines the posterior distributions of the SDF coefficients. The search in Step 2 is grounded in the question of which subset of the factors best belongs in the SDF and is the best in pricing the excluded factors. We believe that without this combination of data-intensive strategies of machine learning methods with finance the search would lead to sub-optimal discoveries.

4 Understanding the CLPZ8 model

We now explain some aspects of this new model. First, we describe the economic dimensions captured by the CLPZ8 slope risk factors. Second, we derive the posterior distributions of the factor risk premia. Third, we examine the OOS realized returns on factor portfolios composed of the CLPZ8 factors and compare those returns to factor portfolios made from the factors of the current models. Finally, we infer the implied posterior distribution of the SDF for each month in the sample and point to implications for pricing.

4.1 Economic dimension

The distinct economic dimensions captured by the factors in the CLPZ8 slope factor model: EGR_S captures investment risk; BASPREAD_S, BETA_S, and MVE_S capture trading frictions; MOM1M_S incorporates momentum; RD_S is related to intangibles, and ROAQ_S captures profitability. The characteristics underlying these factors have been discussed in the literature as predictors of future expected returns. In some cases, differential factors of these characteristics have been constructed. Regardless, this is the first time that these characteristics have been conceived in terms of slope factors, and the first time that they have put together in one asset pricing model. It bears emphasizing that the slope factor versions capture the partial correlation (correlation purged off the influence of other characteristics) between a given characteristic and returns and these can be quite different than what is captured by differential factors (essentially correlation between a given characteristic and returns (without controlling for the effect of other characteritics on returns, except size). Because of this fundamental difference, even the risk premia of a slope factor can have a different sign than the risk premia of the corresponding differential factor. Thus, it is important to interpret the slope factor correctly, as a long-short portfolio that is pure-play (purged of the effect of other characteristics) in contrast to differential factors which are long-short portfolios without the latter vital property.

Now we mention the first paper that discussed the characteristic underlying our factors.

- BASPREAD, which is based on the bid-ask spread, in Amihud and Mendelson (1986).
- BETA in Fama and MacBeth (1973).
- EGR, or annual percent change in book value of equity, in Richardson, Sloan, Soliman, and Tuna (2005).
- MOM1M, the one-month momentum, in Jegadeesh and Titman (1993).
- MVE, or the size factor, in Banz (1981)
- RD, which is based on R&D spending, in Eberhart, Maxwell, and Siddique (2004).
- ROAQ, which is related to returns on assets, in Balakrishnan, Bartov, and Faurel (2010).

4.2 Factor risk premia

A puzzling result noted in the literature is that the factor risk premia of factors in several of the standard models are not significant. Interestingly, as we now show, this is not the case for the CLPZ8 factors.

Commonly, the factor risk premia are estimated by the Fama-Macbeth procedure. Alternatively, one can estimate the model

$$\boldsymbol{x}_t = \boldsymbol{\lambda}_x + \varepsilon_t, \quad t = 1, 2, ..., T \tag{20}$$

where $\boldsymbol{x}_t$ is the vector of factors in the asset pricing model. In this way, one can learn about the $\boldsymbol{\lambda}_x$ (the vector of risk premia) without involving any test assets. We estimate this model assuming that $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \text{MVT}(0, \Omega_x, \nu = 6)$, a multivariate student-t distribution with $\nu = 6$ degrees of freedom. Note that this is same model that we estimated in Step 1, but now the LHS consists only of the factors that are in the SDF. Of course, the estimation algorithm is the same as before. Given sample data on the market plus characteristics-based factors, $\text{data}_T = (\boldsymbol{x}_1, ..., \boldsymbol{x}_T)$, we use MCMC to sample the joint posterior distribution $\pi(\boldsymbol{\lambda}_{\boldsymbol{x}}, \boldsymbol{\Omega}_x | \text{data}_T)$ to get the draws

$$\{(\boldsymbol{\lambda}_x^{(1)}, \boldsymbol{\Omega}_x^{(1)}), ..., (\boldsymbol{\lambda}_x^{(M)}, \boldsymbol{\Omega}_x^{(M)})\} \tag{21}$$

where the draws on $\boldsymbol{\lambda}_x$, by MCMC theory, are automatically from the marginal posterior distribution of the factor risk premia. Inference on each component of $\boldsymbol{\lambda}_x$ can now be based on component-by-component histograms and smoothed kernel density estimates constructed from these draws.

These component-by-component marginal posterior distributions are reported in Figure 3.

The posterior 0.025 and 0.975 quantiles are marked by blue crosses and the posterior means by triangles. From looking at the (0.025,0.975) quantile intervals, six of the seven non-market risk factors have significant factor risk premia, a rather dramatic result. Of the six with significant risk premia, $\mathrm{RD_S}$ and $\mathrm{ROAQ_S}$ have significant positive risk premia. Given that slope factors are pure-play long-short portfolios based on the underlying characteristics, the positive factor risk premia suggests that a strategy of going long on stocks with high R&D expenses and return on assets, and going short on stocks with low values of these characteristics can be expected to earn a positive return. Slope factors that have significant negative factor risk premia are $\mathrm{BASPREAD_S}$, $\mathrm{EGR_S}, \mathrm{MOM1M_S}$ and $\mathrm{MVE_S}$. The negative risk premia of these factors suggests that a strategy of going short on stocks with high values of the underlying characteristics of these factors, and going long with low values of the underlying characteristics, can be expected to earn a positive return.

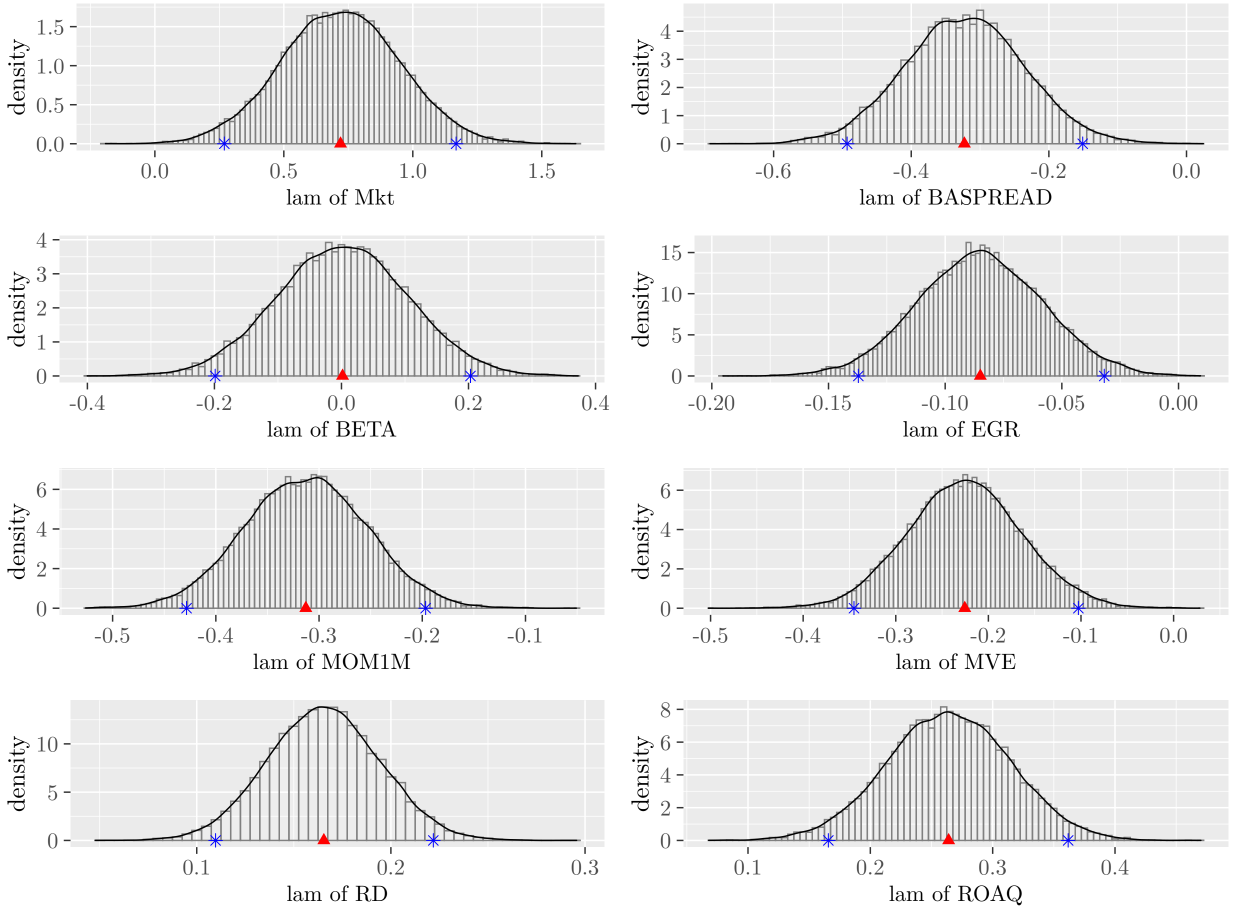

Figure 3 Marginal posterior distributions of the factor risk premia $\boldsymbol{\lambda}_x$ in the CLPZ8 model. These are based on 20,000 MCMC draws, following a burn-in of 1000 draws, from fitting the model $\boldsymbol{x}_t = \boldsymbol{\lambda}_x + \boldsymbol{\varepsilon}_t$, where $\boldsymbol{x}_t$ is the vector consisting of the eight risk factors, $\{Mkt_t, \text{BASPREAD}_{\text{S},t}, \text{BETA}_{\text{S},t}, \text{EGR}_{\text{S},t}, \text{MOM1M}_{\text{S},t}, \text{MVE}_{\text{S},t}, \text{RD}_{\text{S},t}, \text{ROAQ}_{\text{S},t}\}$, and the error is multivariate student-t with $\nu = 6$ degrees of freedom and covariance matrix $\boldsymbol{\Omega}_x$. The Chib and Zeng (2020) training sample-based Gaussian-inverse Wishart prior is used for the parameters. The posterior 0.025 and 0.975 quantiles are marked by blue crosses and the posterior means by triangles.

If we apply the same procedure to the factors of existing models, we get a different, contrasting view. For instance, in the FF6 model of Fama and French (2018), two of the five non-market factors have significant factor risk premia; in the q5 model of Hou, Mo, Xue, and Zhang (2021) two out of the four are significant; and in the DHS Daniel, Hirshleifer, and Sun (2020) model, one out of two is significant. We show these results in Figures 4, 5 and 6. Thus, it is not a given that factors generate significant risk premia. Six out of seven in the case of the CLPZ8 model is a standout.

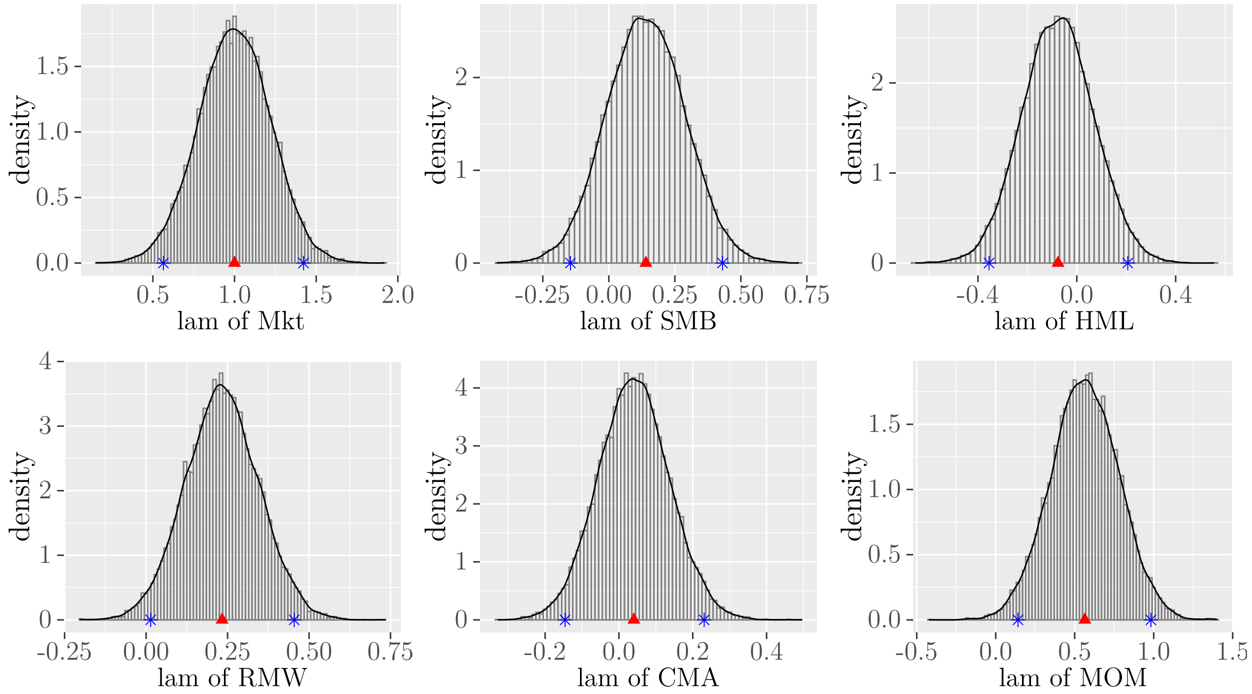

Figure 4 Marginal posterior distributions of the factor risk premia in the FF6 model. These are based on 20,000 MCMC draws, following a burn-in of 1000 draws, from fitting the model $\boldsymbol{x}_{\mathrm{FF6},t} = \boldsymbol{\lambda}_{\mathrm{FF6}} + \boldsymbol{\varepsilon}_{\mathrm{FF6},t}$, where $\boldsymbol{x}_{\mathrm{FF6},t}$ is the vector consisting of the six FF6 factors, and the error is multivariate student-t. The posterior 0.025 and 0.975 quantiles are marked by blue crosses and the posterior means by triangles.

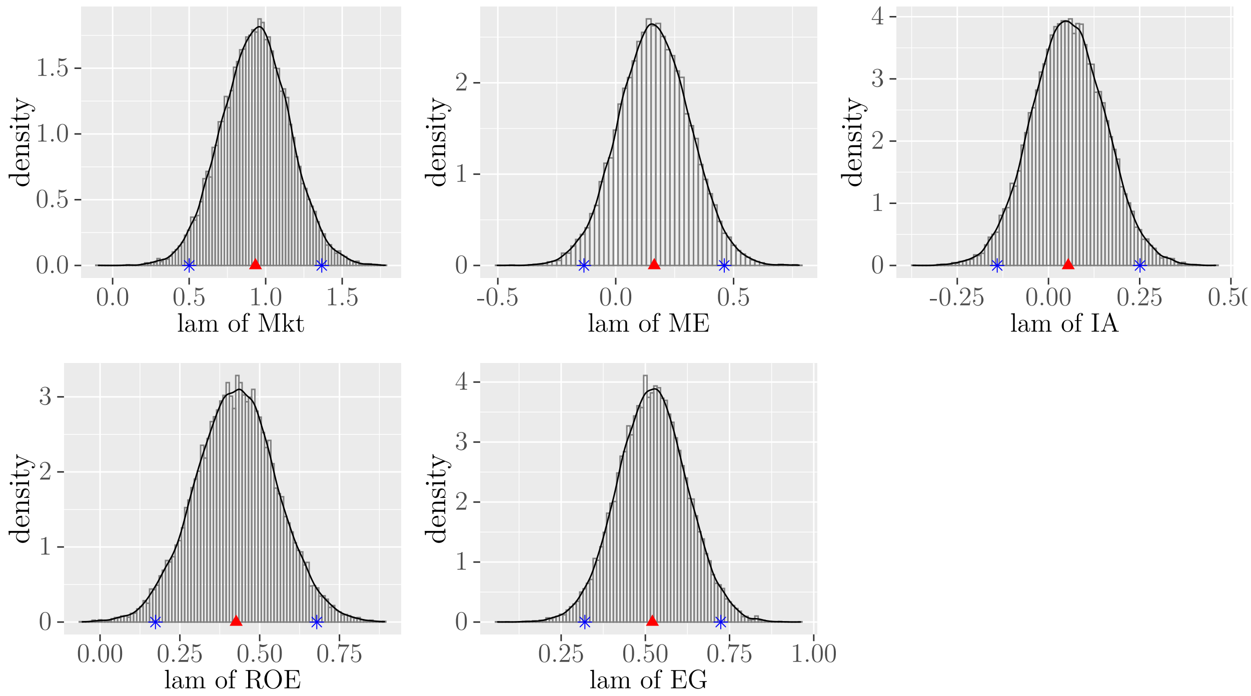

Figure 5 Marginal posterior distributions of the factor risk premia in the q5 model. These are based on 20,000 MCMC draws, following a burn-in of 1000 draws, from fitting the model $\boldsymbol{x}_{\mathrm{q5},t} = \boldsymbol{\lambda}_{\mathrm{q5}} + \boldsymbol{\varepsilon}_{\mathrm{q5},t}$, where $\boldsymbol{x}_{\mathrm{q5},t}$ is the vector consisting of the five q5 factors, and the error is multivariate student-t. The posterior 0.025 and 0.975 quantiles are marked by blue crosses and the posterior means by triangles.

It is also worth highlighting that our slope factor, $\mathrm{MVE_S}$, is able to revive the risk premium associated with the size characteristic, which is believed to have weakened after the 1980s. Indeed, neither SMB nor ME has a significant risk premium as can be seen from the posterior distributions of these factors in Figures 4 and 5. But the latter are based on differential factors for size, which as we have argued in this paper, incorporate the effect of other characteristics on returns that are correlated with size, and do not tell the whole (or even correct) story. Purged of those confounding effects, our slope factor $\mathrm{MVE_S}$ provides a more appropriate rendering of the risk premium associated with size. Along these lines, Asness, Frazzini, Israel, Moskowitz, and Pedersen (2018) also find that a significant size premium re-emerges after controlling for quality exposures, but their control for quality was ex-post, taking the standard SMB factor and regressing it on various quality factors, while in our case the purging of the effect of confounding characteristics is done at the time each factor is constructed (ie., ex-ante), automatically for a large set of characteristics without needing any user intervention. This speaks to the fact that slope factors are really useful objects for asset pricing. By taking full advantage of all that they offer, we have been able to develop the CLPZ8 model in which six of the seven characteristics based factors have significant risk premia, compelling evidence in favor of slope factors in general and the CLPZ8 model in particular.

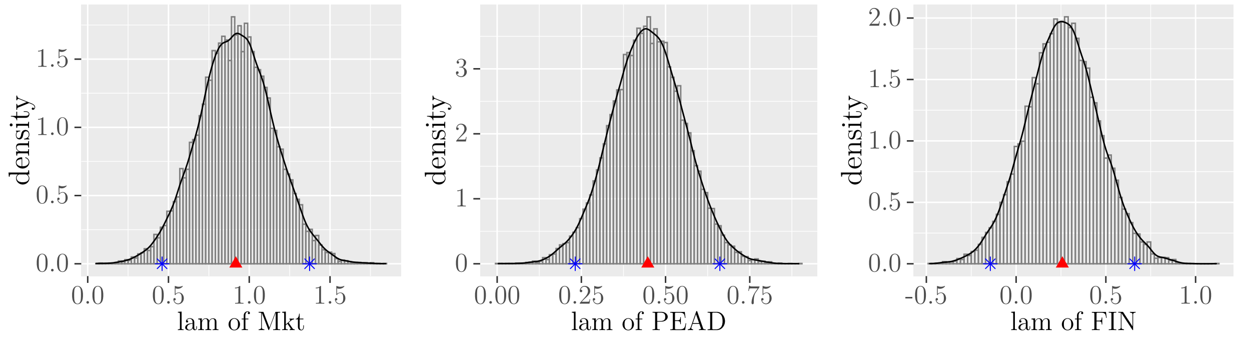

Figure 6 Marginal posterior distributions of the factor risk premia in the DHS model. These are based on 20,000 MCMC draws, following a burn-in of 1000 draws, from fitting the model $\boldsymbol{x}_{\mathrm{DHS},t} = \boldsymbol{\lambda}_{\mathrm{DHS}} + \boldsymbol{\varepsilon}_{\mathrm{DHS},t}$, where $\boldsymbol{x}_{\mathrm{DHS},t}$ is the vector consisting of the three DHS factors, and the error is multivariate student-t. The posterior 0.025 and 0.975 quantiles are marked by blue crosses and the posterior means by triangles.

4.3 Maximum Sharpe-ratio factor portfolio

The preceding discussion on the factor premia of the CLPZ8 factors (and that of the factor premia in the other models) suggests that factor portfolios composed of the CLPZ8 factors are likely to generate higher realized Sharpe-ratios (SRs) compared to factor portfolios made from the factors of the other three models.

Consider the following situation. Suppose that we are at a particular time t and we are given data on the CLPZ8 factors up to that point, data_t. We can use these data on the factors to find the maximum SR factor portfolio. By the well known result, the weight vector of the portfolio is given by

$$\boldsymbol{w}_t = \frac{\hat{\boldsymbol{\Omega}}_{x,t}^{-1}\hat{\boldsymbol{\lambda}}_{x,t}}{\boldsymbol{i}'\hat{\boldsymbol{\Omega}}_{x,t}^{-1}\hat{\boldsymbol{\lambda}}_{x,t}} \tag{22}$$

where $\hat{\boldsymbol{\lambda}}_{x,t}$ and $\hat{\boldsymbol{\Omega}}_{x,t}$ are the sample estimates of the parameters given data_t, and $\boldsymbol{i}$ is the sum vector (vector of ones). We now go long or short on the corresponding factor according to the weights in $\boldsymbol{w}_t$. We hold this factor portfolio for a certain number of periods into the future, say until time $t+s$. Let the actual, ie., realized returns of the factors from $t+1$ to $t+s$ be denoted by $\boldsymbol{X}_{t+1:t+s}$, a matrix of dimension $8\times s$. Then, the out-of-sample (OOS) realized return (rr) of the factor portfolio constructed at time t is

$$\text{rr}_{t+1:t+s} = \boldsymbol{X}'_{t+1:t+s}\boldsymbol{w}_t \tag{23}$$

and the realized Sharpe-ratio OOS is

$$\text{SR}_{\text{OOS},t+1:t+s} = \frac{\text{mean}(\boldsymbol{X}'_{t+1:t+s}\boldsymbol{w}_t)}{\text{sd}(\boldsymbol{X}'_{t+1:t+s}\boldsymbol{w}_t)} \tag{24}$$

We can then move t to $t+s$ and make a new factor portfolio with data_{t+s}, the data up to time $t+s$ and hold that portfolio for the next s periods. We can repeat this computation several times to see if the results are robust to the choice of t. Then we can repeat the entire computation for other sets of factors, say those in the FF6, q5 and DHS models to evaluate the difference in the realized Sharpe-ratios out of sample across factor models.

Table 5 Out-of-sample realized Sharpe-ratios: OOS realized Sharpe-ratios of buy and hold tangency portfolios constructed from the CLPZ8, FF6, q5 and DHS factors
This table presents the OOS realized Sharpe-ratios of tangency portfolios constructed from the CLPZ8, FF6, q5 and DHS factors. Last 60 means that the weights are computed for the data from January 1989 to Dec 2015 and the OOS returns and SR are calculated over the last 60 months spanning Jan 2016 to Dec 2020; last 48 months means that the weights are calculated given data from Jan 1989 to Dec 2016 and the OOS realized returns and SR are over the period Jan 2017 to Dec 2020. Each row expands the estimation window by 12 months and shrinks the OOS evaluation window by 12 months. The realized returns on which the SR is calculated are annualized percentages.

	CLPZ8	FF6	q5	DHS
last 60	0.75	0.29	0.39	0.33
last 48	0.82	0.26	0.44	0.35
last 36	0.80	0.25	0.36	0.33
last 24	1.01	0.37	0.25	0.38
last 12	1.09	0.29	0.27	0.18

We report results from these calculations in Table 5 for five different OOS periods, letting the start date of the first OOS period be January 2016 (the first row of the table) which is then moved up by 12 months in each succeeding row. The end date of the OOS period is December 2020 in each row. Estimation of the weights is on an expanding window, starting in each case from January 1989 and ending the month prior to the start of the OOS period. We can see from the results that the realized SR of the maximum SR portfolio constructed from the CLPZ8 factors is roughly three times larger than those from the FF6, q5 and DHS factors, likely due to the fact that seven out of eight factors in the CLPZ8 model earn a significant factor risk premia.

4.4 Volatility of SDF

We now probe an aspect of the CLPZ8 model (and of the other models) by asking a question that is rarely asked in practice. What is the SDF that underpins the CLPZ8 model? More specifically, what does the data say about the SDF of the CLPZ8 model, and what does that tell us about this model relative to the other three models? To answer this question, we calculate the posterior distribution of the underlying SDF of the CLPZ8 model for every month t in the sample, and we do this for each of the other three models. While it is possible to examine these posterior distributions in many ways, we focus on the inferred volatility of the underlying SDF as this has direct implications for pricing, as we discuss in the next section.

To learn about the SDF, given the data, we proceed as follows. Consider the CLPZ8 model, for example. The SDF that underpins this model at time t is

$$m_t = 1 - \boldsymbol{\lambda}_{\boldsymbol{x}}' \boldsymbol{\Omega}_{\boldsymbol{x}}^{-1} (\boldsymbol{x_t} - \boldsymbol{\mu_x}), \quad t = 1, ..., T$$

We can think of this object, given the data on the factors, as a function of the parameters $\boldsymbol{\lambda_x}$ and $\boldsymbol{\Omega_x}$ ($\boldsymbol{\mu_x} = \boldsymbol{\lambda_x}$). This means that by the usual Bayesian/MCMC arguments, we can get the posterior distribution of m_t by evaluating m_t point wise as

$$m_t^{(g)} = 1 - \boldsymbol{\lambda}_{\boldsymbol{x}}^{(g)\prime} \boldsymbol{\Omega}_{\boldsymbol{x}}^{(g)-1} (\boldsymbol{x_t} - \boldsymbol{\mu}_x^{(g)}), \quad g = 1, 2, ..., M, \quad t = 1, ..., T$$

where

$$(\boldsymbol{\lambda}_x^{(g)}, \boldsymbol{\Omega}_x^{(g)}) \sim \pi(\boldsymbol{\lambda}_x, \boldsymbol{\Omega}_x | \text{data}_T), \quad g = 1, 2, ..., M$$

are the draws on the parameters from estimating the model $\boldsymbol{x}_t = \boldsymbol{\lambda}_x + \varepsilon_t$, $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \text{MVT}(0, \Omega_x, \nu)$, $t = 1, 2, ..., T$ With these M samples on m_t, $t = 1, ..., T$, we can get the posterior mean of m_t and the posterior sd of m_t. We can make a time series plot of these posterior means and, if necessary, superimpose plus-minus 2×sd bands. We then repeat exactly the same computations for the FF6, q5 and DHS models.

We plot the results in Figure 7. Each panel of this plot contains the posterior mean of m_t underlying the FF6, q5 and DHS models. For easy comparison, the posterior mean of m_t underlying the CLPZ8 model is included in each panel. The higher variability of the latter is quite evident. We also take the mean of the m_t over months and the mean of the monthly sds to capture the over variability. These are summarized in Table 6. It is apparent that the underlying SDF of the CLPZ8 model has much higher variability while the variability of the SDF of the other three models is roughly the same. This is perhaps the first asset pricing model with such a starkly different SDF volatility. This difference in variability has consequences for the pricing of the cross-section, as we now discuss.

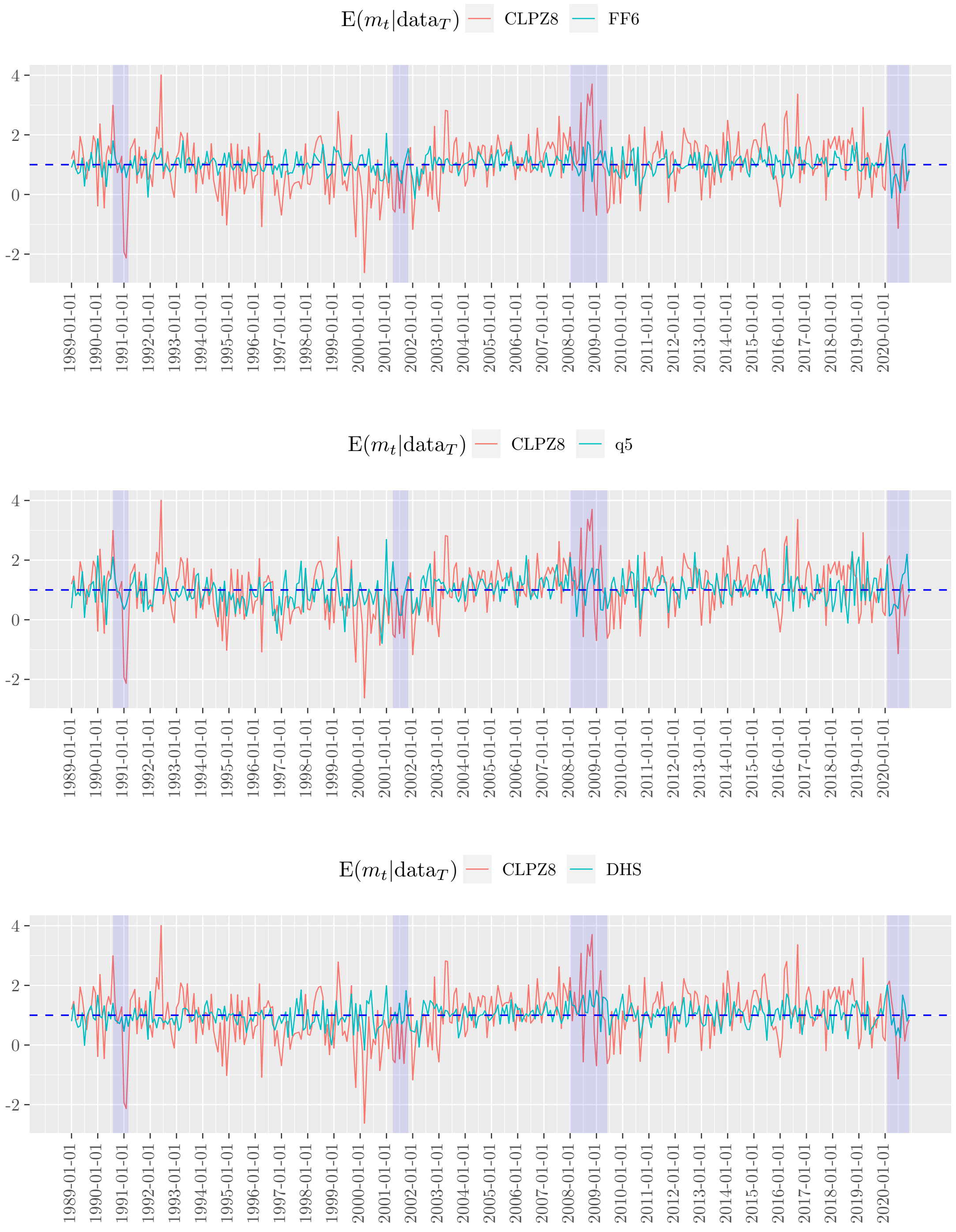

Figure 7 Posterior mean of $m_t = 1 - \boldsymbol{\lambda}_{\boldsymbol{x}}' \boldsymbol{\Omega}_{\boldsymbol{x}}^{-1}(\boldsymbol{x_t} - \boldsymbol{\mu_x})$, by month, underlying each asset pricing model. This figure presents $\mathrm{E}(m_t|\mathrm{data}_T)$, by month, where data_T is the sample data that runs from January 1989 to December 2020. The posterior mean of the m_t underlying the CLPZ8 model is included in red in each plot to show the difference in variability. The shaded area indicate the NBER classified recession periods: August 1990 - March 1991, April 2001 - November 2001, January 2008 - June 2009, and February 2020 - December 2020.

Table 6 Overall mean and sd of the SDF underlying each asset pricing model, January 1989 - December 2020

	CLPZ8	FF6	q5	DHS
mean	1.002	1.017	0.996	0.995
sd	0.898	0.333	0.486	0.363

5 Pricing performance

We now consider an all important question: Does the CLPZ8 model, and the slope factors on which it is based, improve on the pricing of the cross-section? To price the cross-section, a more variable SDF is better. For intuition about this, suppose that the SDF were constant. In that case, pricing based on such an SDF would be vacuous. This intuition carries over in comparing SDFs. One with limited variability should price worse than an SDF with more variability.

This is really just an implication of the Hansen-Jagannathan bound on the volatility of the SDF which says that

$$\frac{\mathrm{sd}(m)}{\mathrm{E}(m)} \geq \frac{\mathrm{E}(r_i)}{\mathrm{sd}(r_i)}, \quad i = 1, 2, ..., n, \tag{25}$$

where r_i is the excess return of the ith asset, $i = 1, 2, ..., n$, and, since each asset in the cross-section generates a different lower bound, the LHS has to exceed the maximum of the Sharpe-ratios of all n assets in the cross-section. Assuming that $\mathrm{E}(m) = 1$, this means that the volatility of the SDF has to be larger than the maximum of the Sharpe-ratios in the cross-section. Thus, more variable SDFs are more likely to satisfy this bound (and therefore do better pricing) than less variable SDFs (which, therefore, would do worse pricing). As stated by Cochrane (2009, pg. 93), "...we need very volatile discount factors with a mean near one to understand stock returns."

This implication is not easy to check out if the volatility of the SDFs of different models is about the same, as is the case with the FF6, q5 and DHS models, but we now have a model that has a much higher SDF volatility. Thus, on theoretical grounds alone we would expect the CLPZ8 model to do better pricing than the three other models. The empirical analysis supports this theoretical contention. To be as comprehensive as possible, we consider the pricing question on a large collection of assets. We split the analysis into two parts. First, we ask if the CLPZ8 slope factor model can price the factors in the FF6, q5 and DHS models. If it can, then that means that the CLPZ8 model subsumes these other models. As part of this question we also examine the question in the reverse direction. Can the FF6, q5 and DHS models price the factors in the CLPZ8 model?

In the second part of our analysis we consider a large collection of test assets, consisting of 1225 portfolios, 1480 ETFs and 6024 stocks. On this large collection of test assets we provide the pricing performance of the CLPZ8 slope factor model in relation to the FF6, q5 and DHS models. Our results, which we document in this section, show that the CLPZ8 model provides a sharper rendering of the risks that are embedded in the cross-section of returns than any of these existing models, in line with the intuition coming from the difference in variability of the SDFs.

5.1 Pricing methodology

We begin by casting the familiar pricing question into Bayesian terms. Consider the CLPZ8 model (exactly the same formulation applies to other models with the CLPZ8 factors replaced by the factors in those models). Let the *excess* return of a factor or a test asset i be denoted by r_i. Then, the question of whether this excess return is priced amounts to a comparison of the two models

$$\begin{aligned} \mathrm{M}_0: \quad & r_{it} = \beta_{1i}\mathrm{Mkt}_t + \beta_{2i}\mathrm{BASPREAD}_{\mathrm{S},t} + \beta_{3i}\mathrm{BETA}_{\mathrm{S},t} + \\ & \beta_{4i}\mathrm{EGR}_{\mathrm{S},t} + \beta_{5i}\mathrm{MOM1M}_{\mathrm{S},t} + \beta_{6i}\mathrm{MVE}_{\mathrm{S},t} + \beta_{7i}\mathrm{RD}_{\mathrm{S},t} + \beta_{8i}\mathrm{ROAQ}_{\mathrm{S},t} + \varepsilon_{it}, \quad t \leq T \end{aligned} \tag{26}$$

versus the model

$$\begin{aligned} \mathrm{M}_1: \quad & r_{it} = \alpha + \beta_{1i}\mathrm{Mkt}_t + \beta_{2i}\mathrm{BASPREAD}_{\mathrm{S},t} + \beta_{3i}\mathrm{BETA}_{\mathrm{S},t} + \\ & \beta_{4i}\mathrm{EGR}_{\mathrm{S},t} + \beta_{5i}\mathrm{MOM1M}_{\mathrm{S},t} + \beta_{6i}\mathrm{MVE}_{\mathrm{S},t} + \beta_{7i}\mathrm{RD}_{\mathrm{S},t} + \beta_{8i}\mathrm{ROAQ}_{\mathrm{S},t} + \varepsilon_{it}, \quad t \leq T \end{aligned} \tag{27}$$

In the non-Bayesian approach, one estimates M_1 and concludes that the r_i is not priced if the test of the null of correct pricing, i.e., $\alpha = 0$ is rejected. In this approach, one can reject the null, but not accept the null that the asset is priced. For example, in a t-test, if the value of the t-statistic is less than the threshold, one cannot conclude that the premium is priced.

In the Bayesian formulation, however, which goes back to Jeffreys (1961), the two models are treated symmetrically. One estimates both models to obtain evidence about pricing versus non-pricing. This evidence is summarized by the posterior probability $\Pr(\mathrm{M}_0|\text{data})$ of (say) M_0. Then, according to Jeffreys scale if the posterior odds of M_0 vs M_1 exceeds 3.15:1 then we conclude that the data evidence strongly prefers M_0 and we claim that the asset is priced. In fact, if one sees even a posterior odds in favor of pricing of at least 2:1, that is compelling evidence of pricing because the models differ by just one parameter. By the algebra discussed above in connection with nearby models, the 2:1 and 3.15:1 rules imply that one can claim correct pricing if d_{01} (the difference in log-marginal likelihoods of M_0 and M_1) exceeds 0.69 and 1.15, respectively, and mispricing otherwise.

We estimate these two models for every factor and every test asset. We assume that the noise term ε_i is Gaussian and employ a normal-inverse Gamma prior on the parameters of these models. The hyperparameters of these prior distributions are fixed by a training sample approach. This training sample consists of the first 15% of the sample.[2]

5.2 Does CLPZ8 price existing risk factors?

Before going forward, we examine if the CLPZ8 factors can price the risk factors of the FF6, q5 and DHS models (we ignore the market factor from this question because it is common to all the models). To answer this question, we apply the pricing methodology just described, putting each FF6, q5 and DHS factor on the LHS of the regression and the CLPZ8 factors on the RHS. We then fit the models without and with an intercept and calculate d_{01}. In addition, we also estimate the model with an intercept by OLS and calculate $\hat{\alpha}$, the OLS estimate of the intercept, and the associated absolute value of the t-statistic.

The results are given in Table 7. If one looks at the last column of the table, which reports d_{01}, one sees that d_{01} exceeds 0.69 for four of the five FF6 factors, all the q5 factors, except r_{EG}, and for FIN in the DHS model. Reading across, one sees that whenever d_{01} is greater than 0.69, the $|t|$ statistic in that row is small, and when $d_{01} < 0.69$, the $|t|$ statistic in that row is large. Since one gets the same conclusion from either perspective, henceforth, we will limit ourselves to looking at the evidence summarized in the d_{01} column. Thus, the evidence shows that eight out of the eleven factors can be priced by the CLPZ8 model, strong necessary evidence in favor of the CLPZ8 model.

[2] If the sample size of a certain stock is small, less than 100 (as in the case of some small stocks), we use the first 40% of the data as a training sample. If the available sample size is between 100 and 150, we use the first 30% of the sample as a training sample. When the sample size exceeds 150, the most common case, we deploy the first 15% of the data to form the training sample prior.

Table 7 Pricing test: FF6, q5 and DHS factors as test assets on the LHS and the CLPZ8 factors on the RHS
This table presents the OLS pricing error $\hat{\alpha}$ and the absolute value of the associated t-statistics in regressions with each FF6, q5 and DHS risk factors on the LHS and the CLPZ8 factors on the RHS. The last column has d_{01}, the log-marginal likelihood difference of the same regressions without and with an intercept. If $d_{01} > 0.69$ (indicated in bold) then the conclusion is that the posterior odds in *favor* of pricing is at least 2:1. The data are from January 1989 to December 2020.

	CLPZ8		
	$\hat{\alpha}$	$\lvert t_{\alpha}\rvert$	d_{01}
SMB	0.09	0.57	**1.10**
HML	-0.24	1.37	0.29
RMW	-0.07	0.57	**1.49**
CMA	0.16	1.27	**2.02**
MOM	0.53	1.98	**0.90**
r_{ME}	-0.01	0.06	**1.25**
r_{IA}	0.17	1.37	**2.06**
r_{ROE}	0.05	0.37	**1.70**
r_{EG}	0.45	3.98	-2.22
PEAD	0.59	4.58	-4.18
FIN	0.20	1.18	**1.94**

We also ask the pricing question in the reverse direction. How many of the CLPZ8 factors can be priced by the FF6, q5 and DHS models? The results appear in Table 8 where the results with the FF6 factors as the independent variables are in the first three columns, those for the q5 factors as the independent variables are in the middle three columns, and those for the DHS factors as the independent variables are in the last three columns.

Table 8 Pricing test: CLPZ8 factors as test assets on the LHS and (separately) the FF6, q5 and DHS factors on the RHS

This table presents the OLS pricing error $\hat{\alpha}$ and the absolute value of the associated t-statistics in regressions with each CLPZ8 factors on the LHS and the factors of the FF6 (first three columns), q5 (the next three) and DHS (last three) models on the RHS. The columns labeled d_{01} contains the log-marginal likelihood difference of the same regressions without and with an intercept. If $d_{01} > 0.69$ (indicated in bold), then the conclusion is that the posterior odds in *favor* of pricing is at least 2:1. The data are from January 1989 to December 2020.

	FF6			q5			DHS		
	$\hat{\alpha}$	$\|t\|$	d_{01}	$\hat{\alpha}$	$\|t\|$	d_{01}	$\hat{\alpha}$	$\|t\|$	d_{01}
BASPREAD$_S$	0.02	0.25	**2.48**	0.08	0.86	**2.43**	0.04	0.45	**4.21**
BETA$_S$	0.01	0.15	**1.66**	0.01	0.15	**1.58**	0.10	1.49	**2.49**
EGR$_S$	-0.12	4.38	-5.83	-0.10	3.39	-2.03	-0.12	4.04	-4.19
MOM1M$_S$	-0.54	8.20	-19.44	-0.53	7.05	-14.04	-0.59	8.45	-20.32
MVE$_S$	-0.35	5.56	-11.38	-0.37	5.35	-11.62	-0.38	5.13	-7.41
RD$_S$	0.22	7.58	-23.07	0.18	5.23	-10.48	0.24	7.84	-25.07
ROAQ$_S$	0.16	3.33	-0.88	0.18	3.34	-2.42	0.15	2.96	-1.81

Going down the d_{01} columns, one sees that $d_{01} > 0.69$ only for BASPREAD$_S$ and BETA$_S$. That is, each model is only able to price two of the seven CLPZ8 risk factors, and that the two factors that can be priced by each model are the same. This is already rather remarkable evidence that the CLPZ8 factors are capturing risks that are completely missed by these existing models. Ultimately, this has to do with the fact that the factors in the FF6, q5 and DHS models are differential factors, that were deemed risk factors by looking at differential factors, and that the universe of differential factors, as we have shown above, give a coarse and incomplete understanding of the risks stemming from the different characteristics. Thus, perhaps it is not surprising that the end-product would leave risks on the table undetected. Not only is the CLPZ8 model based on slope factors, but it is a model that was inferred from a large pool of such factors, and by spreading the net wide, in the manner of machine learning methods, it captures risks that were not on the horizon of these other models.

5.3 CLPZ8 and pricing of portfolios, ETFs and stocks

Another dimension to pricing is how well the CLPZ8 model can price portfolios, ETFs and stocks. Once more, we benchmark the performance relative to the FF6, q5 and DHS models. For a broad pricing test on portfolios, we consider testing assets that are portfolios constructed by us from our sample data. In particular, for each of our 49 characteristics (excluding size), we construct 5×5 sorts on size and characteristic, leading to 1225 ($= 25 \times 49$) value-weighted portfolios. We then apply our pricing test to each of these LHS testing assets with, in turn, each set of risk factors on the RHS. The results, given in the first two columns of Table 9, show that the CLPZ8 factor model prices more of these portfolios than the FF6, q5 and DHS factor models. In particular, under the at least 2:1 posterior odds threshold in favor of pricing criteria, the CLPZ8 model prices 1069 of these portfolios, while the FF6, q5 and DHS models price 467, 467 and 440 of these portfolios, respectively, a substantial difference in pricing ability. Under the even more demanding criteria, of at least 3:1 posterior odds threshold in favor of pricing, the CLPZ8 prices 919 of these portfolios, while the other three models price 337, 306 and 291 of these portfolios, respectively. The evidence of outperformance is clear.

Table 9 Pricing results on test assets: portfolios, ETFs and stocks

	Portfolios		ETFs		Stocks	
	priced:not priced					
model ↓	2:1	3:1	2:1	3:1	2:1	3:1
CLPZ8	1069	919	976	698	4642	3447
FF6	467	337	870	561	4331	2891
q5	467	306	958	633	4411	3046
DHS	440	291	851	532	4115	2679

Number of priced portfolios, ETFs and stocks for each of the CLPZ8, FF6, q5 and DHS models. LHS assets are the test assets; 1225 portfolios from 5×5 sorts on size and 49 characteristics from our sample data; 1480 ETFs obtained from CRSP (sharecode 73) that have at least 60 months of observation between January 1989 - December 2020; 6024 common stocks obtained from CRSP (sharecode 10 and 11) that have least 60 months of observations within January 1989 - December 2020, financial firms, firms with negative book equity and stocks with P/S lower than $5 are excluded. The number of priced assets is reported for posterior odds of priced vs not priced of 2:1 and 3:1.

In the second part of this pricing evaluation, we consider as test assets a large collection of equity ETFs. These assets are diversified portfolios that are traded, liquid, transparent, and cheap to trade. As a result, the premium of these assets is likely to be determined by exposure to the common non-diversifiable sources of risk manifested in the risk factors. With this in mind, we collect data on 1480 ETFs from CRSP (sharecode 73) spanning the period February 1993 (the earliest month for which data on ETFs is available) to December 2020. Within this time-span, each of 1480 ETFs has at least least 60 months of data. Our pricing methodology applied to these test assets shows that under the at least 2:1 posterior odds in favor of pricing criteria, the CLPZ8 model is able to price 976 ETFs, again substantially more than the existing asset pricing models. We report the results in the middle two columns of Table 9.

Finally, we consider the pricing of a sample of 6024 stocks. While, in general, stocks can be difficult to price because of the excess noise relative to portfolios, but they offer a demanding check for the pricing ability of any asset pricing model. We collect the excess returns data on our stock from CRSP (sharecode 10 and 11). The sample period is from January 1989 to December 2020. Once again, we ensure that we have least 60 months of observations within this time frame on any given stock. Financial firms and firms with negative book equity are excluded. Stocks with prices per share lower than $5 are also excluded. The results on pricing of these data are summarized in the last two columns of Table 9. As can be seen from the table, the pricing performance of the CLPZ8 factor model strictly dominates that of the other three models. Specifically, under the 2:1 criteria, the CLPZ8 factor model prices 4642 of these stocks, but the FF6, q5 and DHS models manage to price 4331, 4411, 4115 stocks, respectively, again substantially fewer than the new model.

6 Conclusion

In this paper we have demonstrated the benefits that accrue from using slope factors in asset pricing, provided a new asset pricing model with slope factors, and shown that it outperforms existing asset pricing models in all the dimensions that are used to evaluate asset pricing models. Slope factors are long-short portfolios that directly connect to the underlying characteristics, purged of the influence of other characteristics. In addition, constructing these slope factors in a simple exercise in regression. Furthermore, the slope factors tend to be less noisy than differential factors because of the pure-play long-short property. This, as we have shown, leads to a more variable SDF, which in turn leads to an improved ability to price the cross-section. Thus, our case for slope factors in asset pricing rests on a combination of theoretical arguments (less noisy factors and a more variable SDF), and extensive empirical evidence involving a large collection of test assets, including factors of current models.

To complete our case for asset pricing with slope factors we provide in this paper the first asset pricing model with slope factors, the CLPZ8 model, that is built from the ground-up starting with a large representative group of fifty firm-level characteristics. We used a data intensive procedure, motivated by machine learning precepts, to infer which factors are in the SDF. The result is a parsimonious model that consists of the market factor plus seven slope factors, $\mathrm{BASPREAD_S}$, $\mathrm{BETA_S}$, $\mathrm{EGR_S}$, $\mathrm{MOM1M_S}$, $\mathrm{MVE_S}$, $\mathrm{RD_S}$, $\mathrm{ROAQ_S}$, of which seven earn a significant risk premium (in significant contrast to the case of the factors in existing models). We show that existing models are not able to price five of the seven characteristics based factors in this model, showing that the CLPZ8 model contains risks that are relevant for asset pricing, but have been overlooked by existing models.

We show that the CLPZ8 model outperforms the existing models in terms of the significance of the risk premia of the factors; OOS Sharpe-ratios of tangency portfolios; pricing of the factors in existing models; and pricing on a large collection of portfolios, ETF and stocks. The evidence supports the new model on each of these dimensions and reinforces our contention that slope factors should be the basis for tackling and unearthing the risks that are embedded in the cross-section.

Data, software and code to reproduce the results in this paper are available on request.

Table 10 The descriptive statistics of the 50 characteristics
This table presents acronym, full names, definition, and descriptive statistics of characteristics generated by the code from Green et al. (2017). The min, mean, max, median, and standard deviation are for the characteristics across firms and months. The data are from January 1989 to December 2020.

Acronym	Firm characteristics	Definition	Min	Mean	Max	Median	Std
acc	Working capital accruals	Annual income before extraordinary items (ib) minus operating cash flows (oancf) divided by average total assets (at); if oancf is missing then set to change in act - change in che - change in lct + change in dlc + change in txp-dp	-1.022	-0.045	0.5	-0.055	0.109
age	# years since first Compustat coverage	Number of years since first Compustat coverage	1	13	58	16.653	12.94
agr	Asset growth	Annual percent change in total assets (at)	-0.685	0.067	6.062	0.148	0.419
baspread	Bid-ask spread	Monthly average of daily bid-ask spread divided by average of daily spread	0	0.035	0.901	0.048	0.051
beta	Beta	Estimated market beta from weekly returns and equal weighted market returns for 3 years ending month t-1 with at least 52 weeks of returns	-0.742	0.992	3.937	1.068	0.666
bm	Book-to-market	Book value of equity (ceq) divided by end of fiscal-year-end market capitalization	-2.346	0.531	7.644	0.651	0.605
cash	Cash holdings	Cash and cash equivalents divided by average total assets	-0.079	0.077	0.978	0.161	0.204
cashdebt	Cash flow to debt	Earnings before depreciation and extraordinary items (ib+dp) divided by avg. total liabilities (lt)	-99.683	0.111	2.176	-0.015	1.245
cashpr	Cash productivity	Fiscal-year-end market capitalization plus long-term debt (dltt) minus total assets (at) divided by cash and equivalents (che)	-520.623	-0.072	600.277	-1.224	55.49
cfp	Cash flow to price ratio	Operating cash flows divided by fiscal-year-end market capitalization	-2.797	0.075	2.623	0.074	0.235
chatoia	Industry-adjusted change in asset turnover	2-digit SIC - fiscal-year mean-adjusted change in sales (sale) divided by average total assets (at)	-1.429	0.001	1.194	-0.003	0.216
chcsho	Chane in shares outstanding	Annual percent change in shares outstanding (csho)	-0.891	0.008	2.576	0.1	0.298
chmpia	Industry-adjusted change in profit margin	Industry-adjusted change in number of employees	-24.162	-0.077	3.502	-0.151	0.796
convind	Convertible debt indicator	An indicator equal to 1 if company has convertible debt obligations	0	0	1	0.114	0.318
depr	Depreciation/PP&E	Depreciation divided by PP&E	-0.984	0.188	6.703	0.307	0.432

Table 10 The descriptive statistics of the 50 characteristics

Acronym	Firm characteristics	Definition	Min	Mean	Max	Median	Std
dy	Dividend to price	Total dividends (dvt) divided by market capitalization at fiscal-year-end	-6.122	0	0.35	0.014	0.033
egr	Growth in common shareholder equity	Annual percent change in book value of equity (ceq)	-3.837	0.072	8.286	0.134	0.699
ep	Earnings to price	Annual income before extraordinary items (ib) divided by end of fiscal-year market cap	-7.523	0.043	0.437	-0.038	0.345
gma	Gross profitability	Revenues (revt) minus cost of goods sold (cogs) divided by lagged total assets (at)	-0.961	0.297	1.778	0.345	0.335
grcapx	Growth in capital expenditures	Percent change in capital expenditures from year t-2 to year t	-13.886	0.225	61.947	0.91	3.294
herf	Industry sales concentration	2-digit SIC - fiscal-year sales concentration (sum of squared percent of sales in industry for each company)	0.009	0.044	1	0.07	0.077
hire	Employee growth rate	Percent change in number of employees (emp)	-0.711	0.027	3.973	0.088	0.323
idiovol	Idiosyncratic return volatility	Standard deviation of residuals of weekly returns on weekly equal weighted market returns for 3 years prior to month end	0	0.055	0.279	0.064	0.037
ill	Illiquidity	Average of daily (absolute return / dollar volume)	0	0	0.001	0	0
indmom	Industry momentum	Equal weighted average industry 12-month returns	-0.761	0.112	3.641	0.138	0.284
invest	Capital expenditures and inventory	Annual change in gross property, plant, and equipment (ppegt) + annual change in inventories (invt) all scaled by lagged total assets (at)	-0.507	0.032	1.385	0.061	0.153
lev	Leverage	Total liabilities (lt) divided by fiscal-year-end market capitalization	0	0.622	77.752	2.211	4.674
lgr	Growth in long-term debt	Annual percent change in total liabilities (lt)	-0.758	0.069	9.612	0.229	0.727
mom12m	12-month momentum	11-month cumulative returns ending one month before month end	-0.957	0.056	11.952	0.125	0.582
mom1m	1-month momentum	1-month cumulative return	-0.721	0.002	2.167	0.011	0.153
mve	Size	Natural log of market capitalization at end of month t-1	2.357	12.313	19.018	12.414	2.257
nincr	Number of earnings increases	Number of consecutive quarters (up to eight quarters) with an increase in earnings (ibq) over same quarter in the prior year	0	1	8	0.989	1.326
operprof	Operating profitability	Revenue minus cost of goods sold - SG&A expense - interest expense divided by lagged common shareholders' equity	-8.828	0.614	13.119	0.783	1.201
pchgm_pchsale	% change in gross margin - % change in sales	Percent change in gross margin (sale-cogs) minus percent change in sales (sale)	-12.26	-0.004	4.761	-0.07	0.843

Table 10 The descriptive statistics of the 50 characteristics

Acronym	Firm characteristics	Definition	Min	Mean	Max	Median	Std
pricedelay	Price delay	The proportion of variation in weekly returns for 36 months ending in month explained by 4 lags of weekly market returns incremental to contemporaneous market return	-15.849	0.068	15.597	0.153	1.076
ps	Financial statement score	Sum of 9 indicator variables to form fundamental health score	0	5	9	4.621	1.655
rd	R&D	An indicator variable equal to 1 if R&D expense as a percentage of total assets has an increase greater than 5%	0	0	1	0.141	0.343
roaq	Return on assets	Income before extraordinary items (ibq) divided by one quarter lagged total assets (atq)	-0.59	0.005	0.159	-0.004	0.055
roeq	Quarterly return on equity	Earnings before extraordinary items divided by lagged common shareholders' equity	-2.28	0.021	1.766	-0.001	0.154
roic	Return on invested capital	Annual earnings before interest and taxes (ebit) minus nonoperating income (nopi) divided by non-cash enterprise value (ceq+lt-che)	-23.554	0.059	1.005	-0.143	1.267
salecash	Sales to cash	Annual sales divided by cash and cash equivalents	-300.275	7.889	2503.483	58.36	190.378
saleinv	Sales to inventory	Annual sales divided by total inventory	-35.442	12.16	1031.216	34.889	72.929
salerec	Sales to receivables	Annual sales divided by accounts receivable	-21796	5.949	210.006	11.472	70.916
sgr	Sales growth	Annual percent change in sales (sale)	-0.936	0.086	8.5	0.171	0.522
sin	Sin stocks	An indicator variable equal to 1 if a company's primary industry classification is in smoke or tobacco, beer or alcohol, or gaming	0	0	1	0.01	0.099
sp	Sales to price	Annual revenue (sale) divided by fiscal-year-end market capitalization	-4.131	0.87	37.551	1.778	2.882
std_dolvol	Volatility of liquidity (dollar trading volume)	Monthly standard deviation of daily dollar trading volume	0	0.708	2.783	0.813	0.427
std_turn	Volatility of liquidity (share turnover)	Monthly standard deviation of daily share turnover	0	2.342	736.352	4.833	11.773
tang	Debt capacity/firm tangibility	Cash holdings + 0.715 * receivables +0.547 * inventory + 0.535 * PPE/ totl assets	0	0.525	0.982	0.52	0.162
tb	Tax income to book income	Tax income, calculated from current tax expense divided by maximum federal tax rate, divided by income before extraordinary items	-27.344	-0.048	15.362	-0.096	1.685

Bibliography

Yakov Amihud and Haim Mendelson. Asset pricing and the bid-ask spread. *Journal of Financial Economics*, 17(2):223–249, 1986.

Clifford Asness, Andrea Frazzini, Ronen Israel, Tobias J Moskowitz, and Lasse H Pedersen. Size matters, if you control your junk. *Journal of Financial Economics*, 129(3):479–509, 2018.

Clifford S Asness, Andrea Frazzini, and Lasse Heje Pedersen. Quality minus junk. *Review of Accounting Studies*, 24(1):34–112, 2019.

Kerry Back, Nishad Kapadia, and Barbara Ostdiek. Slopes as factors: Characteristic pure plays. *Available at SSRN 2295993*, 2013.

Kerry Back, Nishad Kapadia, and Barbara Ostdiek. Testing factor models on characteristic and covariance pure plays. *Available at SSRN 2621696*, 2015.

Karthik Balakrishnan, Eli Bartov, and Lucile Faurel. Post loss/profit announcement drift. *Journal of Accounting and Economics*, 50(1):20–41, 2010.

Rolf W Banz. The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1):3–18, 1981.

Luyang Chen, Markus Pelger, and Jason Zhu. Deep learning in asset pricing. *Available at SSRN 3350138*, 2020.

Siddhartha Chib. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90(432):1313–1321, 1995.

Siddhartha Chib. Markov chain Monte Carlo methods: Computation and inference. *Handbook of Econometrics*, 5:3569–3649, 2001.

Siddhartha Chib and Edward Greenberg. Markov chain Monte Carlo simulation methods in econometrics. *Econometric theory*, 12(3):409–431, 1996.

Siddhartha Chib and Xiaming Zeng. Which factors are risk factors in asset pricing? A model scan framework. *Journal of Business & Economic Statistics*, 38(4):771–783, 2020.

Siddhartha Chib, Minchul Shin, and Anna Simoni. Bayesian estimation and comparison of moment condition models. *Journal of the American Statistical Association*, 113(524):1656–1668, 2018.

Siddhartha Chib, Xiaming Zeng, and Lingxiao Zhao. On comparing asset pricing models. *The Journal of Finance*, 75(1):551–577, 2020.

John H Cochrane. *Asset pricing: Revised edition*. Princeton university press, 2009.

Guillaume Coqueret. Characteristics-driven returns in equilibrium. *Available at SSRN 3941195*, 2021.

Kent Daniel, David Hirshleifer, and Lin Sun. Short- and long-horizon behavioral factors. *The Review of Financial Studies*, 33(4):1673–1736, 2020.

Allan C Eberhart, William F Maxwell, and Akhtar R Siddique. An examination of long-term abnormal stock returns and operating performance following R&D increases. *The Journal of Finance*, 59(2):623–650, 2004.

Eugene F Fama. *Foundations Of Finance*. Basic books, 1976.

Eugene F Fama and Kenneth R French. Choosing factors. *Journal of Financial Economics*, 128 (2):234–252, 2018.

Eugene F Fama and Kenneth R French. Comparing cross-section and time-series factor models. *The Review of Financial Studies*, 33(5):1891–1926, 2020.

Eugene F Fama and James D MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636, 1973.

Joachim Freyberger, Andreas Neuhierl, and Michael Weber. Dissecting characteristics nonparametrically. *The Review of Financial Studies*, 33(5):2326–2377, 2020.

Jeremiah Green, John RM Hand, and X Frank Zhang. The characteristics that provide independent information about average US monthly stock returns. *The Review of Financial Studies*, 30(12): 4389–4436, 2017.

Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273, 2020.

Kewei Hou, Haitao Mo, Chen Xue, and Lu Zhang. An augmented q-factor model with expected growth. *Review of Finance*, 25(1):1–41, 2021.

Narasimhan Jegadeesh and Sheridan Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1):65–91, 1993.

Bryan T Kelly, Seth Pruitt, and Yinan Su. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3):501–524, 2019.

Serhiy Kozak, Stefan Nagel, and Shrihari Santosh. Shrinking the cross-section. *Journal of Financial Economics*, 135(2):271–292, 2020.

Scott A Richardson, Richard G Sloan, Mark T Soliman, and Irem Tuna. Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics*, 39(3):437–485, 2005.