Multifactor Models and Their Consistency with the APT

(Review of Asset Pricing Studies, forthcoming)

Ilan Cooper Liang Ma Paulo Maio Dennis Philip

This version: October 2020*

Abstract

We examine the consistency of several prominent multifactor models from the empirical asset pricing literature with the Arbitrage Pricing Theory (APT) framework. We follow the APT-related literature and estimate the common factor structure from a rich cross-section (associated with 42 major CAPM anomalies) by employing the asymptotic principal components method. Our benchmark model contains six statistical factors and clearly dominates, both in economic and statistical terms, most of the empirical multifactor models proposed in the literature by a good margin. These results represent a critical challenge to the current workhorse models in terms of explaining large-scale equity risk premia.

Keywords: asset pricing; linear multifactor models; APT; equity risk factors; statistical factors; stock market anomalies; cross-section of stock returns; asymptotic principal components; spanning regressions

JEL classification: G10; G12

*Cooper, ilan.cooper@bi.no, Department of Finance, Norwegian Business School (BI); Ma, liang.ma@moore.sc.edu, Darla Moore School of Business, University of South Carolina; Maio (corresponding author), paulo.maio@hanken.fi, Department of Finance and Economics, Hanken School of Economics; Philip, dennis.philip@durham.ac.uk, Durham University Business School. We thank an anonymous referee, Kym Ardison, Te-Feng Chen, Thierry Foucault (the editor), Stefano Giglio, Alex Horenstein, Georges Hubner, Konark Saxena, Avanidhar Subrahmanyam, and seminar participants at the University of Mannheim, HEC Liège, Hanken School of Economics, IIM Calcutta Financial Research Workshop, MFA, FMA Asia, ITAM Finance Conference, EFMA, CICF, and FMA for helpful comments. We are grateful to Kenneth French, Yu Yuan, Lu Zhang, and AQR for providing stock market data. Previous versions circulated with the titles "Multifactor models and the APT: Evidence from a broad cross-section of stock returns" and "New factor models and the APT". Any remaining errors are our own.

1 Introduction

For many years, price momentum (Jegadeesh and Titman 1993; Fama and French 1996) and the value premium (Basu 1983; Rosenberg, Reid, and Lanstein 1985; Lakonishok, Shleifer, and Vishny 1994) have been the traditional market anomalies, and hence the focus of attention for newly proposed asset pricing models.[1] However, recent years have seen an explosion of new market anomalies, which correspond to novel patterns in cross-sectional equity risk premia left unexplained by the baseline CAPM of Sharpe (1964) and Lintner (1965). Specifically, Hou, Xue, and Zhang (2015) examine in total around 80 anomalies covering six broad categories: Momentum, value-growth, investment, profitability, intangibles, and trading frictions. They find that nearly half of these anomalies (including those related to trading frictions) are not statistically significant and end up testing their 4-factor model over 35 portfolio sorts. Among the most prominent new patterns in cross-sectional risk premia are a number of investment- and profitability-based anomalies. The investment anomaly can be broadly classified as a pattern in which stocks of firms that invest more exhibit lower average returns than stocks of firms that invest less.[2] The profitability-based anomalies refer to the evidence indicating that more profitable firms earn higher average returns than less profitable firms.[3]

The traditional workhorse in the empirical asset pricing literature—the 3-factor model of Fama and French (1993, 1996)—fails to explain the new market anomalies (see, for example, Fama and French 2015; Hou, Xue, and Zhang 2015). Moreover, the 4-factor model of Carhart (1997) (C4) does a good job in capturing price momentum and other variants of momentum, but struggles in terms of explaining some of the profitability- and investment-based anomalies (see Hou, Xue, and Zhang 2015 for details). In response to this gap, we have witnessed the emergence of new

[1] The value premium refers to the evidence showing that value stocks (stocks with high equity valuation ratios such as book-to-market, earnings-to-price, or cash flow-to-price) outperform growth stocks (low valuation ratios). On the other hand, price momentum corresponds to a cross-sectional pattern where stocks with high prior short-term returns outperform stocks with low prior returns.

[2] The variables that represent corporate investment can be total asset growth (Cooper, Gulen, and Schill 2008), abnormal corporate investment (Titman, Wei, and Xie 2004), investment growth (Xing 2008), inventory growth (Belo and Lin 2011), composite issuance (Daniel and Titman 2006), net stock issues (Pontiff and Woodgate 2008), and different measures of accruals (Sloan 1996; Richardson, Sloan, Soliman, and Tuna 2005; Hafzalla, Lundholm, and Van Winkle 2011).

[3] The profitability measures that have been employed in the literature include return on equity (Haugen and Baker 1996), return on assets (Balakrishnan, Bartov, and Faurel 2010), gross profits-to-assets (Novy-Marx 2013), number of consecutive quarters with earnings increases (Barth, Elliott, and Finn 1999), and failure probability (Campbell, Hilscher, and Szilagyi (2008)).

multifactor models containing (different versions of) investment and profitability factors, in particular the 5-factor model of Fama and French (2015, 2016) (FF5) and the 4-factor model of Hou, Xue, and Zhang (2015) (HXZ4).[4] However, several dimensions of the broad cross-section of stock returns are still not explained by these multifactor models. In particular, the 5-factor model does not account for momentum-based anomalies (including both earnings and industry momentum), while neither of these two models captures several profitability and investment-based (in particular, several forms of accruals) anomalies (see Hou, Xue, and Zhang 2015; Fama and French 2016; Maio and Philip 2018; Cooper and Maio 2019b; Hou, Mo, Xue, and Zhang 2020 for details on the performance of those models for the broad cross-section). In response to this evidence, Fama and French (2018) augment FF5 with the momentum factor into a 6-factor model (FF6), whereas Hou, Mo, Xue, and Zhang (2019, 2020) add an expected growth factor to HXZ4 (HMXZ5).

Given this broad picture, we aim to examine systematically the performance of the current multifactor models proposed in the literature in terms of explaining large-scale cross-sectional risk premia. Our unique approach is to compare the multifactor models against a statistical benchmark model that is motivated by the general framework of the Arbitrage Pricing Theory (APT) of Ross (1976). By doing so, we evaluate the consistency of the existing multifactor models with the APT framework, which has not been well studied in the literature: For the empirical models to be consistent with the APT framework they should not underperform the benchmark statistical model by a great deal. This approach is well justified for two reasons. First, according to the APT, variables that provide a fairly good description of the time-series variation in stock returns should represent risk factors that help to price those same assets. Thus, the APT represents a natural asset pricing benchmark, since several of the most successful multifactor models in the literature, such as those mentioned above, contain factors that are highly correlated with the testing portfolios.[5] Second, the APT is less demanding than other asset pricing frameworks (such as the ICAPM of Merton 1973) in the sense that it relies on relative asset pricing, specifically, given the exogenous common sources of systematic risk (factors), what should be the correct discount rates for equity portfolios. By building the benchmark APT factor model, we also address the research

[4] Feng, Giglio, and Xiu (2020) show that the investment and profitability factors tend to be robust to the presence of other potential factors.

[5] This is the case of the value-growth factor (HML) in relation to portfolios sorted on valuation ratios, the momentum factor (UMD) against momentum portfolios, and the investment and profitability factors used in Fama and French (2015, 2016) and Hou, Xue, and Zhang (2015) in relation to portfolios sorted on these two variables.

question regarding how many factors we need to successfully describe the broad cross-section of stock returns.[6]

We follow part of the empirical APT literature in terms of estimating common statistical factors by applying asymptotical principal components analysis (PCA) to a large cross-section of excess stock returns (e.g., Connor and Korajczyk (1986, 1988); Goyal, Pérignon, and Villa (2008)). Hence, statistical factors, which explain the covariance matrix of the returns of the testing assets, are also pricing factors that help explaining cross-sectional dispersion in risk premia. We employ 42 anomalies or portfolio sorts, which represent a subset of the anomalies considered in Green, Hand, and Zhang (2017), for a total of 420 decile portfolios. Following Hou, Xue, and Zhang (2015, 2020), these anomalies can be generically classified in strategies related to value-growth, momentum, investment, profitability, intangibles, and trading frictions. The goal is to estimate a benchmark statistical model, which by construction and under the APT intuition, has a large explanatory power for this representative cross-section of stock returns.[7]

The estimation of the principal components indicates nine common factors that are dominant over our sample period (1973 to 2016). These nine factors cumulatively explain around 92% of the common variation in the 420 raw portfolio returns. The first factor acts a level factor and hence is strongly correlated with the market return. The second factor is especially correlated with several anomalies in the profitability and trading frictions categories. The third factor is more correlated with several value-growth anomalies, while the fourth factor mainly captures momentum. The sixth factor is mainly correlated with intangibles, while the eighth factor is correlated with price momentum. Therefore, to a large degree the nine statistical factors capture different dimensions of our cross-section of 42 anomalies.

We conduct cross-sectional asset pricing tests of our APT model by using the 420 equity portfolios as testing assets. The results show that a 6-factor model (denoted by APT6), containing the first,

[6] In his presidential address, Cochrane (2011) raises this question: "Can we again account for N dimensions of expected returns with $K < N$ factor exposures?"

[7] Our APT model estimates static weights for the latent statistical factors. Therefore, any dynamic relationships between the factors and testing assets are not modelled explicitly. In this regard, our empirical framework differs from the recent approaches of Kelly, Pruitt, and Su (2019) and Kim, Korajczyk, and Neuhierl (2020), who utilise firm characteristics as conditioning information for expected returns to introduce time variation in factor loadings and alphas. The other important distinction is that these papers use individual stocks as test assets, while we use a large dimensional panel of portfolios sorted on firm characteristics as testing assets. Our empirical choices of using static statistical factors and employing a large number of portfolios as test assets stems from consistency and comparison purposes: All the empirical multifactor models tested in the paper represent unconditional models (i.e., both factor risk prices and betas are constant over time) and are typically tested on a (large) cross-section of equity portfolios.

second, fourth, sixth, eighth, and ninth principal components as risk factors, explains 51% of the cross-sectional variation in the risk premia associated with the 420 portfolios. Moreover, the corresponding factor risk price estimates are strongly statistically significant. When we impose the constraint that the factor risk price estimates are equal to the factor means, the resulting cross-sectional R^2 ("constrained" R^2) is essentially the same as the OLS counterpart (0.50). Such explanatory ratio uses the intercepts from the time-series regressions (alphas), or equivalently, the pricing errors from a "constrained" cross-sectional regression in which the risk price estimates are equal to the factor means. This estimate is also strongly statistically significant, based on the inference associated with a bootstrap simulation.

We conduct an alternative asset pricing test with 252 portfolios associated with the three extreme deciles (on each leg) for each market anomaly. This stems from the fact that most cross-sectional dispersion in risk premia is concentrated on the extreme deciles within each portfolio group. The results show a slightly larger fit for the statistical model, as indicated by the OLS and constrained explanatory ratios of around 0.58.

The central analysis in the paper is to compare some of the most popular multifactor models existent in the literature (denoted by empirical models) against our APT model in terms of pricing the 420 portfolios. That is, the statistical model represents a benchmark to evaluate the performance of the proposed factor models. The empirical models we examine include the above-mentioned C4, HXZ4, HMXZ5, FF5, and FF6 models. We also estimate the 6-factor model proposed by Barillas and Shanken (2018) (BS6, which roughly combines the factors from HXZ4 and FF5), and the 4-factor model of Stambaugh and Yuan (2017) (SY4, which contains two composite factors related to firms' performance and management). By running spanning regressions of each statistical factor (in our benchmark model) onto the equity long-short factors (associated with each of the seven empirical models) we find there is a substantial amount of information in the statistical factors that is not explained by these seven factor models.

The asset pricing tests show that the performance of the seven empirical models lags behind the fit of the 6-factor APT, with differences in cross-sectional R^2 between 28 percentage points (comparison against HMXZ5) and 91 percentage points (comparison with BS6). When using the shorter and more interesting cross-section associated with the extreme portfolios, the gaps in explanatory ratio relative to the benchmark model vary between 25 percentage points (HMXZ5) and 86 percentage points (BS6). Critically, the APT model dominates most of the empirical models in statistical terms (at the 5% level). The sole exception is HMXZ5, in which case the gaps in cross-sectional R^2 estimates relative to the benchmark model are not always significant at the 5% level. However, this last finding is not robust to changes in the empirical design.

Our main findings are robust to employing a shorter time-series, excluding some anomalies from the testing assets, employing alternative statistical inference, using alternative model evaluation metrics, or estimating the models with alternative testing portfolios. We also find that the benchmark statistical model dominates the empirical models by conducting an "out-of-sample" analysis over the time-series dimension.

The overall conclusions from this paper are simple, but important. Several of the current empirical workhorses employed in the asset pricing literature fail to be good empirical proxies for the APT. That is, they deviate significantly, both in economic and statistical terms, from a benchmark statistical model that is designed in such a way (APT intuition) to explain well a rich cross-section of equity risk premia. Therefore, assuming that explaining the broad cross-section of CAPM anomalies is the main goal of a successful empirical multifactor model, our results suggest that most of the models proposed in the literature fail considerably on such dimension. The only possible (but weak) exception to this pattern is the 5-factor model of Hou, Mo, Xue, and Zhang (2019), which is not always dominated statistically by the reference model at conventional significance levels.

Our main goal in the paper is to use the statistical model as a reference point to evaluate the performance of the empirical models, rather than proposing yet a new multifactor model to the long list already existent in the literature. We follow this approach for two main reasons. First, the design of our benchmark statistical model is deliberately in-sample, that is, the statistical factors that explain the covariance matrix of returns of the testing assets are employed to price the risk premia of the very same assets. This implies that a priori the pricing performance of our model may not be generalizable to other testing assets. For example, it is possible that our model does not do a good job in pricing bond risk premia or individual stocks. If we are interested a priori in pricing other sources of risk premia (e.g., individual stocks), the benchmark statistical model should be constructed from the realized returns of those very same assets. In that sense, our analysis represents a "new empirical method" to evaluate current multifactor models rather than a "new multifactor model", and such exercise is critically sensitive to the cross-section chosen in the first place. Second, an important ingredient of every linear factor model is the economic plausibility of the cross-sectional dispersion in the factor loadings (which ultimately drives the pricing performance of the model). The economic interpretation of the patterns in factor loadings becomes much less clear if we employ statistical factors, which are (by construction) related with several segments (anomalies or characteristics) of the cross-section. Using equity factors constructed from a single characteristic, which is the norm pursued in the empirical asset pricing literature, provides a clearer and sharper interpretation of the factor loadings. All in all, a major implication from the paper is that we need "better" single-characteristic-based factors than those currently included in most empirical multifactor models proposed in the literature.

This paper is related to the growing literature that focuses on evaluating and comparing asset pricing models containing only traded factors by using a relatively rich cross-section of portfolio risk premia. Examples include Fama and French (2015, 2016), Hou, Xue, and Zhang (2015), Maio and Santa-Clara (2017), Cooper and Maio (2019b), and Hou, Mo, Xue, and Zhang (2020). Our key innovation relative to these studies is that, apart from assessing the fit of each model for large-scale portfolio risk premia, we evaluate the difference in performance against a benchmark statistical model that is designed to have a large fit for cross-sectional equity risk premia. This is important because a given empirical model can have a seemingly good fit for a given cross-section (as indicated by a small average pricing error or a high cross-sectional R^2), while still considerably underperforming the reference model. Critically, following the prescription of Lewellen, Nagel, and Shanken (2010), our empirical analysis relies on forcing the models to price simultaneously the full cross-section containing the 42 CAPM anomalies, consistent with the procedure used in Hou, Xue, and Zhang (2015), Maio and Santa-Clara (2017), Cooper and Maio (2019b), and Hou, Mo, Xue, and Zhang (2020). This represents a substantially more challenging cross-sectional asset pricing test than merely forcing a given model to price each anomaly (e.g., book-to-market portfolios) on a stand-alone basis, as in Fama and French (2015, 2016).

Our work also has implications for a recent strand of the literature that conducts comparison of asset pricing models without relying on cross-sectional risk premia as testing assets (e.g., Barillas and Shanken 2017, 2018; Fama and French 2018; Hou, Mo, Xue, and Zhang 2019). The underlying principle is that the comparison between two competing models reduces to the degree by which each factor on a given model is spanned by the factors in the other model, hence the testing assets become irrelevant for model comparison. However, ranking factor models exclusively by the magnitudes and statistical significance of the estimated intercepts (alphas) from spanning regressions might mask the fact that those models do not price the cross-section of testing assets in the first place. This may render the comparison meaningless, as pricing the cross-section of asset risk premia should be the primary goal for any candidate asset pricing model. Our results suggest that this might be the case for most workhorses proposed in the literature if we aim to price a sufficiently rich cross-section of equity risk premia. Therefore, to properly evaluate and compare models, it is important to provide evidence from cross-sectional tests (including the deviations relative to a benchmark statistical model) in addition to the time-series spanning tests.

This study is also related to the recent works of Giglio and Xiu (2019) and Kozak, Nagel, and Santosh (2018), who also construct common statistical factors from the cross-section of realized stock returns. Among other aspects, our paper differs from those studies in two key dimensions. First, we use the model containing the PCA factors as a reference point to show that the current factor models (used in the literature) are not sufficiently successful in terms of pricing a large number of market anomalies. Secondly, in comparison with Kozak, Nagel, and Santosh (2018), we use a considerably larger cross-section of portfolio returns to evaluate factor models. Our study is also related to the recent work of Pukthuanthong, Roll, and Subrahmanyam (2019) in the sense that in their paper the selection of risk factors (that price assets) is motivated by information obtained from the covariance matrix of the returns of the testing assets.

2 Theoretical Background

The Arbitrage Pricing Theory (APT) is first developed by Ross (1976). Chamberlain and Rothschild (1983) present a generalized version of the APT that embeds an approximate factor structure, and suggest the use of principal component analysis in empirical tests of factor models motivated by the APT. In this section, we provide a simple derivation of the APT to motivate the empirical analysis conducted in the following sections. Our presentation largely follows Back (2017) (Chapter 6) and Cochrane (2005) (Chapter 9).

Consider the following equation for any of the N risky assets indexed by $i = 1, ..., N$,

$$R^e_{i,t+1} = a_i + \beta_{i,1}\widetilde{G}_{1,t+1} + ... + \beta_{i,K}\widetilde{G}_{K,t+1} + \varepsilon_{i,t+1}, \tag{1}$$

where $R^e_{i,t+1} = R_{i,t+1} - R_{f,t+1}$ denotes the return on asset i in excess of the risk-free rate, $\varepsilon_{i,t+1}$ is the residual return, and $\widetilde{G}_{j,t+1} \equiv G_{j,t+1} - \mathrm{E}(G_{j,t+1}), j = 1, ..., K$ represents each of the demeaned common K factors. Since the factors are demeaned, it follows that $\mathrm{E}(R^e_{i,t+1}) = a_i$.

Assume that there is a stochastic discount factor (SDF), M_{t+1}, that prices assets in this economy. By multiplying both sides of the regression above by M_{t+1}, taking unconditional expectations, and using both $\mathrm{E}(M_{t+1}R^e_{i,t+1}) = 0$ and $\mathrm{E}(R^e_{i,t+1}) = \alpha_i$, we obtain:

$$\mathrm{E}(R^e_{i,t+1}) = -\beta_{i,1}\frac{\mathrm{E}(M_{t+1}\widetilde{G}_{1,t+1})}{\mathrm{E}(M_{t+1})} - ... - \beta_{i,K}\frac{\mathrm{E}(M_{t+1}\widetilde{G}_{K,t+1})}{\mathrm{E}(M_{t+1})} - \frac{\mathrm{E}(M_{t+1}\varepsilon_{i,t+1})}{\mathrm{E}(M_{t+1})}. \tag{2}$$

Consistent with our empirical analysis, we assume there is a risk-free asset. Hence, we have $\mathrm{E}(R_{f,t+1}) = 1/\,\mathrm{E}(M_{t+1})$, which leads to the following expected return-beta equation,

$$\mathrm{E}(R^e_{i,t+1}) = \beta_{i,1}\lambda_1 + ... + \beta_{i,K}\lambda_K + \delta_i, \tag{3}$$

where

$$\lambda_j = -\,\mathrm{E}(R_{f,t+1})\,\mathrm{E}(M_{t+1}\widetilde{G}_{j,t+1}), j = 1, ..., K,$$

represents the risk price for factor j, and

$$\delta_i \equiv -\frac{\mathrm{E}(M_{t+1}\varepsilon_{i,t+1})}{\mathrm{E}(M_{t+1})},$$

denotes the "pricing error" for asset i.[8]

Now assume that idiosyncratic risk is small for all assets, $\mathrm{Var}(\varepsilon_{i,t+1}) \approx 0$. This implies that $\mathrm{E}(M_{t+1}\varepsilon_{i,t+1}) = \mathrm{Cov}(M_{t+1}, \varepsilon_{i,t+1}) \approx 0$, which in turn implies $\delta_i \approx 0$: In the limit, a very small

[8]If we further assume that all the factors are excess returns, $\mathrm{E}(M_{t+1}G_{j,t+1}) = 0$, the risk price for factor j is equal to the corresponding factor mean:

$$\lambda_j = -\,\mathrm{E}(R_{f,t+1})\,\mathrm{E}[M_{t+1}(G_{j,t+1} - \mathrm{E}(G_{j,t+1}))] = \mathrm{E}(R_{f,t+1})\,\mathrm{E}(M_{t+1})\,\mathrm{E}(G_{j,t+1}) = \mathrm{E}(G_{j,t+1}).$$

value of $\text{Var}(\varepsilon_{i,t+1})$ means that $\varepsilon_{i,t+1}$ is not a random variable (see Cochrane 2005; Back 2017). Consequently, we have an approximate linear K-factor model for asset i:

$$\text{E}(R^e_{i,t+1}) \approx \beta_{i,1}\lambda_1 + ... + \beta_{i,K}\lambda_K. \tag{4}$$

However, the assumption of small or negligible idiosyncratic risk is not realistic for individual assets (e.g., stocks). Therefore, the APT approximation relies on the construction of well-diversified portfolios. By defining a portfolio p with excess return given by $R^e_{p,t+1} = \sum_{i=1}^{N} \omega_i R^e_{i,t+1}$, and using Equation (3), we obtain,

$$\text{E}(R^e_{p,t+1}) = \beta_{p,1}\lambda_1 + ... + \beta_{p,K}\lambda_K + \delta_p, \tag{5}$$

where $\beta_{p,j} \equiv \sum_{i=1}^{N} \omega_i \beta_{i,j}, j = 1, ..., K$ denote the factor betas for portfolio p; $\varepsilon_{p,t+1} \equiv \sum_{i=1}^{N} \omega_i \varepsilon_{i,t+1}$ represents the residual return for portfolio p; and $\delta_p \equiv \sum_{i=1}^{N} \omega_i \delta_i = -\text{E}(M_{t+1}\varepsilon_{p,t+1})/\text{E}(M_{t+1})$ denotes the pricing error for portfolio p.

The traditional derivation of the APT relies on the assumption that the covariance matrix of the residual returns is diagonal,

$$\text{E}(\varepsilon_{i,t+1}\varepsilon_{l,t+1}) = 0,$$

for any two different assets i and l. Furthermore, we have the critical assumption that the variance of the residual returns of individual assets is bounded:

$$\max_{i=1,...,N} \text{Var}(\varepsilon_{i,t+1}) \leq \sigma^2.$$

These two conditions imply that the variance of the residual return of a well diversified portfolio (e.g., $\omega_i = 1/N$) is given by

$$\text{Var}(\varepsilon_{p,t+1}) = \frac{1}{N^2}\sum_{i=1}^{N} \text{Var}(\varepsilon_{i,t+1}) \leq \frac{\sigma^2}{N}.$$

This expression approaches zero for a large number of assets. Consequently, the pricing error (δ_p) is approximately close to zero and the linear multifactor model holds as a good approximation in terms of explaining the risk premia of portfolio p:

$$\mathrm{E}(R^e_{p,t+1}) \approx \beta_{p,1}\lambda_1 + ... + \beta_{p,K}\lambda_K. \tag{6}$$

By relying on asymptotic arguments, it turns out that the linear factor model also applies (approximately) to individual assets. Under the conditions of a diagonal covariance matrix of residual returns combined with bounded $\mathrm{Var}(\varepsilon_{i,t+1})$, it follows that

$$\lim_{N\to\infty} \sum_{i=1}^{N} \delta_i^2 < \infty. \tag{7}$$

This condition in turn implies that most assets have small pricing errors in magnitude, that is, for any real number $\delta > 0$, there are only a finite number of assets with $|\delta_i| \geq \delta$.[9] Hence, within the general framework of the APT, it follows that statistical factors that explain well the covariance matrix of returns should also be pricing factors that explain well cross-sectional risk premia for both portfolios and (many of the) individual assets.

However, the critical assumption stated above of a strict factor structure, which generates exactly orthogonal residual returns across assets, is unrealistic from an empirical viewpoint. This invalidates that the pricing errors for diversified portfolios on the one hand, or for many individual assets on the other hand, are approximately zero. To overcome such limitation, Chamberlain and Rothschild (1983) provide an alternative derivation of the APT under weaker conditions. In their framework, the covariance matrix of the residual returns does not need to be diagonal with bounded diagonal elements. Instead, sufficient conditions for obtaining the APT approximation are that (i) there is an approximate factor structure (in which weak correlations of the residual returns are allowed) and (ii) the maximum eigenvalue of the covariance matrix of the residual returns is bounded as the number of assets increases.

A couple of observations about the empirical implications of the APT are pertinent. First, the discussion above implies that the APT approximation is more plausible if we use well-diversified equity portfolios, rather than individual stocks, as testing assets. The reason is that the individual

[9] See Reisman (1988) and Back (2017) for details on the derivation of this result. An equivalent condition, $\lim_{N\to\infty}(1/N)\sum_{i=1}^{N}\delta_i^2 = 0$, can be derived by relying on the absence of asymptotic arbitrage opportunities: Construct a large zero-cost portfolio with zero variance and impose the restriction that both the realized and expected portfolios returns are zero (see Ingersoll (1987) and Pennacchi (2008) for details).

stocks have larger idiosyncratic risk than most equity portfolios.[10]

Secondly, the exposition above shows that the factors in the APT can be either traded or non-traded. Yet, given the restriction of an approximate factor structure, it follows that successful empirical applications of the APT should contain factors that represent excess stock returns (zero-cost portfolios). The reason is that those factors are typically more correlated with the excess returns on the testing assets (equity portfolios) than non-traded factors such as macro variables (e.g., CPI inflation, industrial production growth, bond yields, short-term interest rates), and thus should do a better job in explaining the covariance matrix of the residual returns.[11] Critically, if the factors are excess returns, the risk price estimates cannot be freely estimated by a cross-sectional regression and should be equal to the corresponding factor means.[12]

3 Statistical Factors

In this section, we estimate the statistical factors that summarize the information from the broad cross-section of stock returns.

3.1 Data

The portfolio return data used in the estimation of the common statistical factors are associated with the most relevant market or CAPM anomalies, which represent patterns in cross-sectional stock returns that are not explained by the baseline CAPM. We employ 42 anomalies or portfolio sorts, which represent a subset of the anomalies considered in Green, Hand, and Zhang (2017), for a total of 420 portfolios. Table 1 contains the list and description of the anomalies included in this

[10]Using well-diversified equity portfolios (associated with firm characteristics) as testing assets has been the most popular practice in the cross-sectional asset practice literature (see Fama and French 2015, 2016; Hou, Xue, and Zhang 2015; Maio and Santa-Clara 2017; Cooper and Maio 2019a,b; Hou, Mo, Xue, and Zhang 2020 as recent examples). In related work, Kirby (2020) employs "managed" equity portfolios. For asset pricing tests with individual stocks, see, for example, Kim and Skoulakis (2018), Pukthuanthong, Roll, and Subrahmanyam (2019), and Ang, Liu, and Schwarz (2020). For examples of asset pricing tests conducted with other asset classes, see Lettau, Maggiori, and Weber (2014) and Delikouras and Kostakis (2019).

[11]Actually, in some cases these large correlations are (nearly) mechanical, such as the case of HML (against portfolios sorted on the book-to-market ratio) or the case of UMD (in relation to momentum portfolios).

[12]Another implication from the APT is that such framework is mainly about relative asset pricing: Given the factors, what should be the correct prices (i.e., expected returns) of the other assets in the economy. However, the APT does not provide an economic explanation for the risk premium associated with each original source of systematic risk (the factors). Alternative asset pricing frameworks, which provide a theory of the factor risk premiums, include the Consumption CAPM (Breeden 1979), the Intertemporal CAPM (Merton 1973), and the baseline CAPM (Sharpe 1964; Lintner 1965).

paper. Following Hou, Xue, and Zhang (2015, 2020), these anomalies can be generically classified in strategies related to value-growth, momentum, investment, profitability, intangibles, and trading frictions.[13]

For all anomalies, we form value-weighted decile portfolios with NYSE breakpoints and rebalance these decile portfolios monthly. For most of the anomalies, we follow the same procedure of portfolio construction as in Green, Hand, and Zhang (2017). The exception applies to five anomalies that use quarterly earnings/sales information, including earnings announcement return (*ear*); return on assets (*roaq*); return on equity (*roeq*); and revenue surprise (*rsup*). For these anomalies, we use earnings/sales data in Compustat quarterly files in the months immediately after the most recent public earnings announcement dates (Compustat item RDQ) when forming portfolio sorts of stocks.[14] Furthermore, for a firm to be included in the portfolio sorts, we require the end of the fiscal quarter corresponding to the most recently announced earnings/sales to be within six months prior to the portfolio formation, to exclude stale earnings/sales information. This procedure is consistent with Hou, Xue, and Zhang (2015, 2020). In comparison to the 102 portfolio groups employed in Green, Hand, and Zhang (2017), we start with 72 anomalies that have return data available since 1973. Of these 72 anomalies, we exclude 28 anomalies in which the corresponding "high-minus-low" return spreads produce insignificant (at the 10% level) CAPM alphas.[15] We exclude two additional groups—market beta squared and stock return volatility—for which the corresponding high-minus-low return spreads have correlations above 90% (in magnitude) relative to other anomalies. This leads to a total of 42 groups in the end. To construct portfolio excess returns, we subtract the 1-month Treasury bill rate, available from Kenneth French's website. The sample period is 1973:01 to 2016:12.

The descriptive statistics for high-minus-low return spreads (between the last and first deciles among each portfolio class) are presented in the Online Appendix. The anomaly with the largest spread in average returns is 12-month momentum ($mom12m$), with a premium above 1% per month. The return spreads associated with book-to-market (bm), change in 6-month momentum ($chmom$),

[13] We follow most of the cross-sectional asset pricing literature in working with all the deciles associated with a given anomaly rather than just focusing on the extreme first and last deciles (e.g., Fama and French 2015, 2016; Hou, Xue, and Zhang 2015; Hou, Mo, Xue, and Zhang 2020).

[14] As discussed in Jegadeesh and Livnat (2006), sales are generally announced with earnings during quarterly earnings announcements.

[15] Yan and Zheng (2017) show that many CAPM anomalies cannot be attributed to random chance.

ear, earnings-to-price ratio (*ep*), and sales-to-price ratio (*sp*) are also economically important, with (absolute) means above 0.60% per month. The anomalies with lower average returns are market beta (*beta*), current ratio (*currat*), idiosyncratic volatility (*idiovol*), % change in current ratio (*pchcurrat*), quick ratio (*quick*), sales-to-cash ratio (*salecash*), sales-to-inventory ratio (*saleinv*), and share turnover (*turn*), all with average return spreads below 0.20% (in absolute value). Beta is the anomaly with the most volatile spread in returns (standard deviation above 8% per month), followed by bid-ask spread (*baspread*), *idiovol*, and $mom12m$, all three spreads with volatilities above 7%. At the other end of the spectrum, there are several anomalies with volatilities of return spreads below 3% per month: Industry-adjusted change in asset turnover (*chatoia*), growth in long-term debt (*lgr*), % change in capital expenditures (*pchcapx*), *pchcurrat*, % change in depreciation (*pchdepr*), % change in sales minus % change in inventory (*pchsale_pchinvt*), and % change in sales-to-inventory ratio (*pchsaleinv*).

3.2 Estimation

To estimate the pervasive factors spanning the common factor space of the broad cross-section of stock returns, we use the approximate factor model framework developed by Connor and Korajczyk (1986, 1988), which has been successfully implemented to uncover the cross-correlations present in large macroeconomic or financial panels (see Ludvigson and Ng 2007, 2009, 2010; Goyal, Pérignon, and Villa 2008; Connor, Korajczyk, and Uhlaner 2015; Maio and Philip 2015, among others). This method is denoted by asymptotic principal components.[16] Given the exposition in the previous section, the multifactor model containing the statistical factors is a valid approximation to the linear combination of the true risk factors embedded in our cross-section of stock returns. Indeed, Connor and Korajczyk (1986, 1988) show that the statistical factors estimated under this approach converge to the true (unobserved) risk factors as the number of assets diverges. More recently, Connor, Korajczyk, and Uhlaner (2015) show that the statistical factors associated with this method are identical (up to a rotation) to the factors estimated by iterating the two-step cross-sectional regression method (used to estimate factor models), irrespective of the initial set of factors specified by the researcher.

[16] Roll and Ross (1980) propose a related approach. Chen, Connor, and Korajczyk (2018) provide a simulation exercise based on individual stock returns.

Consider that equity portfolio excess returns are driven by a finite number of K static unobservable factors,

$$R_{i,t}^{e} = \mathbf{f}_t' \boldsymbol{\beta}_i + \varepsilon_{i,t}, \tag{8}$$

where $R_{i,t}^{e}$ is the portfolio $(i = 1, ..., N)$ excess return at time $t(= 1, ..., T)$; $\mathbf{f}_t$ is the K-dimensional vector of latent common factors for all excess returns at t; $\boldsymbol{\beta}_i$ is the K-dimensional vector of factor loadings for the excess return on asset i; and $\varepsilon_{i,t}$ stands for the idiosyncratic $i.i.d.$ errors, which are allowed to have limited correlation among returns.[17]

This model captures the main sources of variations and covariations among the N portfolio returns with a set of K common factors ($K << N$). The framework is estimated using asymptotic principal components analysis (PCA), which involves an eigen decomposition of the sample covariance matrix. The $(K \times T)$ pervasive factors' matrix $\widehat{\mathbf{F}}$ contains the K eigenvectors corresponding to the first K largest eigenvalues of the $T \times T$ matrix, $\mathbf{R}\mathbf{R}'/(NT)$, where $\mathbf{R}$ is a $(T \times N)$ data matrix of excess returns. The normalization $\widehat{\mathbf{F}}\widehat{\mathbf{F}}' = \mathbf{I}_K$ is imposed, where $\mathbf{I}_K$ is the K-dimensional identity matrix, since $\mathbf{F}$ and the factor loadings matrix are not separately identifiable. Bai and Ng (2002) show that for large N and large T panels, this methodology can effectively distinguish noise from signal and summarize information into a small number of estimated common factors.

To determine the value of K, which is the number of common factors, we use the IC_2 information criterion suggested by Bai and Ng (2002). We minimize over K the following criterion,

$$\ln(V_K) + K\left(\frac{N+T}{NT}\right)\ln(\min\{N,T\}), \tag{9}$$

where $V_K = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(R_{i,t}^{e} - \sum_{j=1}^{K}\widehat{F}_{j,t}\widehat{\beta}_{ij}\right)^2$, with $\widehat{F}_{j,t}$ denoting the j^{th} normalized factor estimate at time t. We consider a maximum set of 40 factors when estimating the optimal K; however the test results are robust to the choice of maximum.

Based on the IC_2 information criterion, we observe the optimal value of K to be ten for our full sample period. However, we also evaluate the stability of the IC_2 criteria across different sampling periods. Specifically, we estimate the factors for each sample that starts from years 1973 to 2012. Untabulated results reveal that there are nine common factors dominant over the excess return's space. Hence, we choose to work with the first nine statistical factors as a first reference

[17] From now on, i refers to an arbitrary equity portfolio.

point.[18] Table 2 reports summary statistics of the nine estimated factors. We can see that none of the factors is persistent, as shown by the first-order autocorrelation coefficients around or below 0.15 (in magnitude). This stems from the fact that stock returns do not usually exhibit significant serial correlation. The nine factors cumulatively explain around 92% of the total variations in the 420 portfolio returns, with the first factor explaining the largest proportion of the cross-sectional variation in returns (around 85%).

To understand the correlations of the estimated common factors with the raw portfolio returns, we conduct simple regressions of the 42 return spreads indicated above on each of the statistical factors,

$$R_{l,10,t} - R_{l,1,t} = \varpi_{l,j} + \beta_{l,j}\widehat{F}_{j,t} + \varepsilon_{l,t}, j = 1, ..., 9, \tag{10}$$

where $R_{l,10,t} - R_{l,1,t}$ denotes the spread high-minus-low associated with anomaly $l, l = 1, ..., 42$.

Table 3 presents the R^2 estimates associated with these simple regressions. These estimates represent the square of the pairwise correlations between the returns and each of the factors. We can see that the first two factors are especially correlated with several anomalies in the trading frictions category, including *baspread*, *beta*, *idiovol*, *maxret*, and *turn*, in all cases with explanatory ratios around or above 0.40. $\widehat{F}_2$ is also correlated with *ep*, *quick*, and several profitability anomalies (*roaq* and *roeq*), while the first factor also has a large correlation with *currat*. The third factor is correlated with several value-growth anomalies (*bm* and *sp*), price momentum (*mom*12*m*), cash productivity (*cashpr*), and dollar trading volume (*dolvol*). The fourth factor mainly captures momentum, as indicated by the explanatory ratios around or above 40% for the *indmom*, *mom*6*m*, and *mom*12*m* return spreads. $\widehat{F}_6$ is mainly correlated with intangibles (*salecash* and *salerec*), while $\widehat{F}_8$ is more correlated with the change in momentum (*chmom*). The remaining statistical factors show a number of smaller correlations with several anomalies, and hence their economic meaning is less clear.

Overall, the results from Table 3 suggest that to a large degree the nine common factors capture different subsets of the 42 market anomalies. This is consistent with the role of these factors in terms of successfully describing the covariance matrix of returns associated with this cross-section

[18] For robustness, we also implement the ER and GR criterion functions proposed by Ahn and Horenstein (2013). The tests results indicate the presence of six significant common return factors. However, to avoid undue omission of important factors explaining the variation of the cross-section of stock returns, we opt to use the first nine PCA factors as an initial reference point for our analysis in the next section.

of 420 equity portfolios.

4 Asset Pricing Tests

In this section, we test our "APT model" containing statistical factors over the broad cross-section of stock returns.

4.1 Methodology

To test our statistical model (as well as the factor models covered in the next section) in the cross-section of average stock returns, we use several empirical methods. The first method consists of the two-step time-series/cross-sectional regression procedure employed in Black, Jensen, and Scholes (1972), Jagannathan and Wang (1998), and Brennan, Wang, and Xia (2004), among others, which is widely used in the literature. In the first step, the factor betas are estimated from the time-series (multivariate) regressions for each of the testing portfolios,

$$\mathbf{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{f}_t + \boldsymbol{\varepsilon}_t, \tag{11}$$

where $\mathbf{r}_t \equiv (R^e_{1,t}, ..., R^e_{N,t})'$ is a vector of excess portfolio returns; $\boldsymbol{\alpha}$ is a vector of intercepts; $\boldsymbol{\beta}(N \times K)$ is a matrix of K factor loadings for the N test assets; $\mathbf{f}_t(K \times 1)$ is a vector of factor realizations; and $\boldsymbol{\varepsilon}_t(N \times 1)$ is the vector of return disturbances.[19]

In the second step, the K-factor model is estimated by an OLS cross-sectional regression,

$$\overline{\mathbf{r}} = \widehat{\boldsymbol{\beta}}\boldsymbol{\lambda} + \boldsymbol{\pi}, \tag{12}$$

where $\overline{\mathbf{r}}(N \times 1)$ is a vector of average excess returns, $\overline{\mathbf{r}} \equiv (1/T)\sum_{t=1}^{T} \mathbf{r}_t = (\overline{R^e_1}, ..., \overline{R^e_N})'$; $\boldsymbol{\lambda}(K \times 1)$ is a vector of risk prices; $\boldsymbol{\pi}(N \times 1)$ is the vector of pricing errors; and $\widehat{\boldsymbol{\beta}}(N \times K)$ denotes the matrix with the estimated factor loadings.[20]

[19] Here $\mathbf{f}$ denotes the vector of realizations on the pricing factors in a generic K-factor model, which includes not only our benchmark statistical model, but also the factor models covered in the next section.

[20] We do not use a GLS cross-section regression in the second step to estimate the factor risk prices (see Kandel and Stambaugh 1995; Shanken and Zhou 2007; Lewellen, Nagel, and Shanken 2010 for applications of this method). The reason is that we are particularly interested in pricing the original equity portfolios (which have an economic interest) rather than an efficient (minimum variance) combination of these portfolios (see Cochrane 2005; Ludvigson 2013 for a detailed discussion).

The t-statistics associated with the factor risk price estimates are based on Shanken's standard errors (Shanken 1992), which incorporate a correction for the estimation error in the factor loadings. We do not include an intercept in the cross-sectional regression, since we want to impose the economic restrictions associated with each factor model. If the statistical model is correctly specified, the intercept in the cross-sectional regression should be equal to zero.[21]

To gauge the fit of each model, we compute the cross-sectional OLS coefficient of determination,

$$R_{OLS}^2 = 1 - \frac{\mathrm{Var}_N(\widehat{\pi}_i)}{\mathrm{Var}_N(\overline{R_i^e})}, \tag{13}$$

where $\mathrm{Var}_N(\widehat{\pi}_i)$ stands for the cross-sectional variance of the pricing errors and $\mathrm{Var}_N(\overline{R_i^e})$ denotes the cross-sectional variance of the raw risk premia. R_{OLS}^2 represents the fraction of the cross-sectional variance of average excess returns (on the testing assets) explained by the factor loadings associated with a given model. Since we do not include an intercept in the cross-sectional regression, this R^2 measure can assume negative values. A negative explanatory ratio means that the regression including the factor loadings (associated with a given model) as regressors performs worse than a trivial regression containing just an intercept (see Campbell and Vuolteenaho 2004).[22] To evaluate the statistical significance of R_{OLS}^2, we use empirical p-values obtained from a bootstrap simulation (see Maio and Santa-Clara (2017) and Guo and Maio (2020) for details).

A related cross-sectional R^2 metric is given by

$$\rho^2 = 1 - \frac{\mathrm{Var}_N(\widehat{\pi}_i)}{S_N(\overline{R_i^e})}, \tag{14}$$

where $S_N(\cdot)$ stands for the cross-sectional second-moment.[23] Contrary to the benchmark R^2 measure (R_{OLS}^2), $\widehat{\rho}^2$ always lies between zero and one. However, the new measure is less informative about the explanatory power of a model for the cross-sectional dispersion in risk premia. Indeed, a model

[21] Another important reason for not including the intercept in the cross-sectional regressions is to preserve consistency with the time-series regression approach. This last method applies to models where all factors are excess returns and represents the focus of our empirical analysis, as discussed below. Nonetheless, as a robustness check, we conduct an alternative cross-sectional regression containing an unrestricted zero-beta rate. The results are discussed in the online appendix.

[22] Similar R^2 measures are used in Jagannathan and Wang (1996), Lettau and Ludvigson (2001), Maio (2013b), Lioui and Maio (2014), and Lettau, Maggiori, and Weber (2014), among others.

[23] Kan, Robotti, and Shanken (2013) employ this measure in the analysis with excess returns and restricted zero-beta rate (see their Section 3).

can have a large value of $\widehat{\rho}^2$ just by fitting well the cross-sectional mean despite not explaining cross-sectional dispersion in risk premia (indicated by low or negative estimates of R^2_{OLS}). Hence, the two measures are not directly comparable.

We compute the difference in ρ^2 between the benchmark APT model (containing several statistical factors) and the single-factor model containing only the first principal component, $S\rho^2 = \rho^2(APT6) - \rho^2(APT1)$. This allows us to assess the additional explanatory power of the higher-order statistical factors relative to the first principal component when it comes to price the 42 anomalies. To assess the statistical significance of both ρ^2 and $S\rho^2$, we rely on the asymptotic distribution derived in Kan, Robotti, and Shanken (2013).[24]

An alternative method to estimate and evaluate asset pricing models is the popular time-series regression approach (see Fama and French 1993, 1996, 2015; Hou, Xue, and Zhang 2015, and many others). This methodology is adequate when all the factors in a model represent excess stock returns as it is the case with the statistical model (see Cochrane 2005). Indeed, the statistical factors estimated in Section 3 can be interpreted as excess returns since they represent linear combinations of the excess returns on the raw 420 equity portfolios. Consequently, the implied risk price estimates are forced to be exactly equal to the respective factor means. Therefore, this method avoids the common criticism of implausible risk price estimates within the two-step regression approach (see Lewellen and Nagel 2006; Lewellen, Nagel, and Shanken 2010), making it a more correct procedure to test our statistical model (as well as the factor models tested in the following sections).

Under the time-series approach, the intercepts from the time-series regressions represent the pricing errors ($\boldsymbol{\alpha}$) associated with the factor model, as all factors are traded. To evaluate the fit of the model, we compute the mean absolute alpha across the testing portfolios:

$$\text{MAA} = \frac{1}{N}\sum_{i=1}^{N} |\widehat{\alpha}_i|. \tag{15}$$

To assess the statistical significance of the pricing errors (alphas), we compute the GRS statistic of Gibbons, Ross, and Shanken (1989), which tests the null hypothesis that the alphas are jointly equal to zero. We also compute the number of alphas that are statistically significant (at the 5% level) on an individual basis. To gauge the individual statistical significance, we employ GMM-based

[24]We thank the referee for suggesting this analysis.

t-ratios (White 1980).

One limitation of MAA is that it does not relate the magnitudes of the pricing errors with the magnitudes of the raw portfolio risk premia that we seek to explain in the first place. For example, a given model may produce an average pricing error that is apparently large, but is actually small in comparison with the scale of the raw risk premia that we are trying to explain. This is especially important in our case, as we have joint asset pricing tests involving many different anomalies, and thus, with different magnitudes of risk premia. To overcome such limitation, we compute the "constrained" cross-sectional R^2 proposed in Maio and Santa-Clara (2017) and Cooper and Maio (2019b),[25]

$$R_C^2 = 1 - \frac{\mathrm{Var}_N(\widehat{\alpha}_i)}{\mathrm{Var}_N(\overline{R_i^e})}. \tag{16}$$

This metric is similar to the cross-sectional OLS R^2, but is based on the pricing errors (intercepts) from the time-series regressions. Hence, this explanatory ratio is only valid when all the factors in the model are excess returns. As shown in Cooper and Maio (2019b) (see also Lewellen, Nagel, and Shanken 2010; Kan, Robotti, and Shanken 2013), equivalently, the pricing errors can be obtained from the following constrained cross-sectional regression,

$$\overline{R_i^e} = \widehat{\beta}_{i,1}\overline{F_1} + ... + \widehat{\beta}_{i,K}\overline{F_K} + \widehat{\alpha}_i, \tag{17}$$

where $\overline{F_j}$ represents the time-series mean of risk factor $j, j = 1, ..., K$ and $\widehat{\beta}_{i,j}$ denotes the corresponding factor loading for asset i. If the risk price estimates obtained under the OLS two-step regression approach roughly coincide with the factor means, it turns out that these two measures (R_C^2 and R_{OLS}^2) are approximately equivalent. However, in general, R_C^2 represents a more correct metric to evaluate the fit of a model (containing only traded factors) than R_{OLS}^2, as the risk price estimates (obtained from the cross-sectional regression) my differ substantially from the corresponding factor means.

[25] A related measure is given by

$$R_{C^*}^2 = 1 - \frac{\widehat{\boldsymbol{\alpha}}'\widehat{\boldsymbol{\alpha}}}{\mathrm{Var}_N(\overline{R_i^e})}.$$

This measure accounts for the possibility that R_C^2 might be inflated due to small values of $\mathrm{Var}_N(\hat{\alpha}_i)$, even when the magnitudes of the pricing errors ($\hat{\alpha}_i$) are large. Unreported results show that the estimates of $R_{C^*}^2$ are quite similar to the corresponding estimates for R_C^2.

We also compute the difference in R^2_C between the benchmark APT model and the single-factor model, $SR^2 = R^2_C(APT6) - R^2_C(APT1)$. To assess the statistical significance of both R^2_C and SR^2, we compute empirical p-values obtained from a bootstrap simulation. The empirical p-values correspond to the fractions of artificial samples in which the pseudo statistics are higher than the corresponding sample estimates. In the simulation, we impose the condition that the factors are independent from the returns ("useless factors" as in Kan and Zhang (1999)), but preserve the correlations among the factors in a given model. Full details of the simulation are presented in the Internet Appendix.

4.2 Results

The results for the asset pricing tests of the 9-factor statistical model, as well as a nested model (containing a subset of the nine factors), based on the OLS cross-sectional regression approach are displayed in Table 4 (Panel A). The testing assets are the 420 equity portfolios. The 9-factor APT model containing the first nine PCA factors explains about 51% of the cross-sectional variation in the equity risk premia among the 420 portfolios. This represents a large fit given the large dimension of the cross-section and the high number of anomalies considered, some of them being negatively correlated. To put these results in perspective, the single-factor model containing only the first principal component produces an R^2 estimate of -58%. Since the first factor represents basically a market factor (as shown in the next section), these results are consistent with previous evidence showing that the baseline CAPM has a negative fit when it comes to explaining the market anomalies considered in the paper. Indeed, untabulated results show that the CAPM produces an explanatory ratio of -59%, which represents basically the same fit as that associated with the single-factor APT model.[26]

Turning to the risk price estimates, it turns out that most of these estimates are statistically significant at the 1% or 5% level. The exceptions are λ_3, λ_5, and λ_7, in which cases the risk price estimates are not significant even at the 10% level.[27] In light of this evidence, we estimate a

[26] When tested on portfolios such as those used in this paper, the CAPM typically produces negative R^2_{OLS} estimates (see Campbell and Vuolteenaho 2004; Yogo 2006; Maio 2013a; Maio and Santa-Clara 2017, among others). This means that the model performs worse than a trivial model that predicts constant average returns in the cross-section of equity portfolios.

[27] This suggests that those three statistical factors are mainly related with equity portfolios that have small risk premia.

restricted version of the statistical model (denoted by APT6), which excludes the third ($\widehat{F}_3$), fifth ($\widehat{F}_5$), and seventh ($\widehat{F}_7$) principal components (PC):[28]

$$\mathrm{E}(R^e_{i,t+1}) = \beta_{i,1}\lambda_1 + \beta_{i,2}\lambda_2 + \beta_{i,4}\lambda_4 + \beta_{i,6}\lambda_6 + \beta_{i,8}\lambda_8 + \beta_{i,9}\lambda_9. \tag{18}$$

The fit of the 6-factor model is basically the same as that associated with the 9-factor model (51% with significance also at the 1% level), which suggests that the three factors enumerated above do not add explanatory power to a model that already contains the other six statistical factors. In other words, the two models are equivalent when it comes to pricing these testing assets. The estimate of ρ^2 for APT6 is identical to that of APT9 (0.97), compared to 0.91 for the model containing the first PC, and both estimates are statistically significant at the 1% level (p-values close to zero), based on the asymptotic inference provided by Kan, Robotti, and Shanken (2013).[29] We can also see that the risk price estimates within APT6 are unchanged by excluding the three factors mentioned above, which stems from the fact that the statistical factors are uncorrelated by construction.

In Panel B of Table 4, we report the estimation results for a smaller cross-section of portfolio returns. This contains the extreme three deciles (on each leg) within each group of decile portfolios for a total of 252 (42×6) portfolios. The reason for using this restricted cross-section hinges on the fact that most of the cross-sectional dispersion in portfolio risk premia is concentrated on the extreme deciles within each anomaly. Hence, by excluding the middle deciles we obtain a more powerful asset pricing test. Moreover, we obtain a partial decoupling between the portfolios used in the construction of the statistical factors and the portfolios employed as testing assets. The results show that the fit of the 6-factor model is marginally above the corresponding explanatory ratio in the estimation with all portfolios, as indicated by the R^2_{OLS} of 0.58. As in the benchmark case, such an estimate is the same as the explanatory ratio associated with the 9-factor model, indicating that excluding the third, fifth, and seventh principal components has no impact on the model's performance. The estimate of ρ^2 for APT6 is 0.97, compared to 0.88 for APT1, with both estimates

[28]This dimension of the model is close to the five statistical factors proposed in Ahn, Horenstein, and Wang (2018) and Kelly, Pruitt, and Su (2019).

[29]The discrepancy in the estimates of R^2_{OLS} and ρ^2 in the case of APT1 stems from the single-factor model being able to match the average cross-sectional risk premium (provided by the first PC), while not being able to capture any cross-sectional variation in risk premia.

being strongly significant (1% level). Such difference in fit between the two models ($S\rho^2$) around 9 percentage points is statistically significant at the 1% level (p-value of zero).[30] We also observe that the risk price estimates in APT6 are quite similar to those obtained under the benchmark asset pricing test. The main difference occurs for the estimates of λ_9 (and to a lower degree for the estimates of λ_6), which assume slightly larger magnitudes relative to the estimation with all portfolios.

Next, we evaluate the APT model by using the time-series approach. The results reported in Table 5 (Panel A) show that the mean absolute alpha for the 6-factor model is 0.08%, compared to 0.13% for the single-factor model (APT1). More importantly, our 6-factor model produces an R_C^2 of 50%, which is largely significant based on the empirical p-value (1% level). In comparison, APT1 produces a negative explanatory ratio (-0.58), which indicates that it performs worse than a trivial model containing only an intercept. The corresponding spread in R_C^2 among these two models (1.09) is strongly different than zero in statistical terms (p-value of zero). Despite the substantially larger explanatory power of APT6, untabulated results indicate that the 6-factor model is rejected (at the 1% level) by the GRS test (with a p-value about zero), in the same vein as APT1. This stems from the large cross-section employed (420 portfolios), which makes it very difficult to accept the null hypothesis that *all* pricing errors are equal to zero. This issue is reinforced by the problems in inverting a "large" covariance matrix of the residual returns, which even causes the formal rejection of a model with "small" pricing errors.[31] To get a more robust idea of the statistical significance of the pricing errors, we compute the number of portfolio groups in which the model is not rejected (at the 5% level) by the GRS test. This metric is used, for example, in Hou, Xue, and Zhang (2015), Cooper and Maio (2019b), and Hou, Mo, Xue, and Zhang (2020). We find that APT6 passes the specification test for 32 (out of 42) of the anomaly groups, which represents more than twice the number obtained for APT1 (13). Additionally, we find that the 6-factor model produces significant alphas (at the 5% level) for 42 portfolios (10% of all the testing portfolios used), which represents about a third of the number of significant alphas generated by APT1. This illustrates another dimension of the outperformance of APT6 in relation to the single-factor model.

[30] The estimation of $S\rho^2$ is conducted exclusively for the asset pricing tests associated with the extreme deciles. The reason is that when computing the p-values for this statistic in the main asset pricing estimation (with 420 portfolios) we face singularity problems.

[31] All the models used in the paper are rejected by the GRS test, with p-values very close to zero.

By using the cross-section of 252 portfolios (Panel B of Table 5), we obtain a marginally higher explanatory ratio of 58% for APT6, which is also strongly significant (1% level). Consequently, the gap in explanatory ratio against the single-factor model is even larger than in the benchmark asset pricing estimation, with significance at the 1% level. On one hand, the average alpha for APT6 is 0.08%, which coincides with the estimate obtained for the full cross-section, and represents about half the magnitude of the average pricing error for APT1 (0.16%). The 6-factor model is not rejected for 26 of the 42 anomalies, compared to 12 in the case of APT1. On the other hand, the number of significant alphas for APT6 is 29, which is less than a third the corresponding estimate for the single-factor model (95).

Another interesting result is that the R_C^2 estimates for APT6 are basically the same as the OLS counterpart estimates discussed above and this holds for both cross-sections. This stems from the fact that the OLS risk price estimates reported above are quite similar to the corresponding factor means reported in Table 2: Only in the case of $\widehat{F}_8$ there is a bigger difference between the factor mean (-0.90) and the OLS risk price estimates (around -0.78). In the next section, we will see that such pattern does not hold for the empirical factor models.

Overall, the results of this section indicate that our APT model, which contains six out of the first nine PCA factors, does a good job in describing the cross-section of 420 equity portfolios considered in this study. The fit of the model is even higher when it comes to explaining the extreme portfolios associated with each anomaly. This performance of the statistical model is not totally surprising: Under the weak restrictions of the APT framework (large fit in the time-series regressions, or alternatively, strong factor structure), the statistical factors were designed in such a way to deliver a large fit for this specific cross-section of equity risk premia. Therefore, the 6-factor model represents a benchmark against which the performance of popular multifactor models is measured, at least when it comes to price our fixed cross-section of testing assets. This represents the focus of analysis in the following sections.

5 Linkage to Factor Models

In this section, we compare our APT model to some of the most popular multifactor models existent in the literature, which represents the main goal of the paper. Following Section 2, we restrict the analysis to models where all the factors represent excess stock returns (zero-cost portfolios).[32] Since the theoretical background of these models is not totally clear in the literature and the factors are traded, we designate these factor models by "empirical" models. This is merely to facilitate the distinction against the benchmark statistical model estimated in the previous section.[33]

In this and the following sections, we consider that, for an empirical factor model to be consistent with the APT, the model should have a similar (in statistical terms) pricing performance (for cross-sectional equity risk premia) relative to the statistical model extracted from the principal component analysis applied to the excess returns of the testing assets. In this sense, the statistical model represents a benchmark, or an upper-bound, to evaluate the pricing performance of empirical factor models.

5.1 Multifactor models

We employ seven multifactor models widely used in the cross-sectional asset pricing literature. The first model is the Carhart's (Carhart 1997) 4-factor model (C4 henceforth),

$$\mathrm{E}(R_{i,t+1} - R_{f,t+1}) = \lambda_M \beta_{i,M} + \lambda_{SMB} \beta_{i,SMB} + \lambda_{HML} \beta_{i,HML} + \lambda_{UMD} \beta_{i,UMD}, \tag{19}$$

which adds a momentum factor (UMD) to the Fama and French (1993) 3-factor model.

The second model is the 4-factor model proposed by Hou, Xue, and Zhang (2015) (HXZ4). This model adds an investment factor (IA, low-minus-high asset growth) and a profitability factor (ROE, high-minus-low return on equity) to the usual market and size (ME) factors:[34]

$$\mathrm{E}(R_{i,t+1} - R_{f,t+1}) = \lambda_M \beta_{i,M} + \lambda_{ME} \beta_{i,ME} + \lambda_{IA} \beta_{i,IA} + \lambda_{ROE} \beta_{i,ROE}. \tag{20}$$

Next, we estimate the 5-factor model proposed by Hou, Mo, Xue, and Zhang (2019, 2020)

[32] Surprisingly, one of the first so-called APT's empirical applications is the multifactor model proposed by Chen, Roll, and Ross (1986), which relies on macro factors.

[33] Specifically, the models proposed by Fama and French (2015, 2016) and Hou, Xue, and Zhang (2015) both contain profitability and investment risk factors. However, while Fama and French (2015) motivate their 5-factor model based on the present-value valuation model of Miller and Modigliani (1961), it turns out that Hou, Xue, and Zhang (2015) rely on the q-theory of investment. On the other hand, Maio and Santa-Clara (2012) and Cooper and Maio (2019a) provide evidence that several of the factors included in the models analyzed here are consistent with Merton's ICAPM (Merton 1973).

[34] The size factor employed in Hou, Xue, and Zhang (2015) is constructed in a slightly different way to the Fama-French size factor (SMB).

(denoted by HMXZ5), which augments HXZ4 by an expected growth factor (EG):

$$\mathrm{E}(R_{i,t+1} - R_{f,t+1}) = \lambda_M \beta_{i,M} + \lambda_{ME} \beta_{i,ME} + \lambda_{IA} \beta_{i,IA} + \lambda_{ROE} \beta_{i,ROE} + \lambda_{EG} \beta_{i,EG}. \tag{21}$$

The fourth model is the 5-factor model of Fama and French (2015, 2016) (FF5), which adds an investment (CMA, low-minus-high asset growth) and a profitability (RMW, high-minus-low operating profitability) factor to the Fama and French (1993) 3-factor model:

$$\mathrm{E}(R_{i,t+1} - R_{f,t+1}) = \lambda_M \beta_{i,M} + \lambda_{SMB^*} \beta_{i,SMB^*} + \lambda_{HML} \beta_{i,HML} + \lambda_{CMA} \beta_{i,CMA} + \lambda_{RMW} \beta_{i,RMW}. \tag{22}$$

Both CMA and RMW are constructed in a different way to the corresponding investment and profitability factors from Hou, Xue, and Zhang (2015).[35] In addition, SMB^* is constructed from different portfolio sorts than the original SMB.[36]

Next, we estimate the 6-factor model proposed by Fama and French (2018), which augments FF5 by the momentum factor (FF6):

$$\mathrm{E}(R_{i,t+1} - R_{f,t+1}) = \lambda_M \beta_{i,M} + \lambda_{SMB^*} \beta_{i,SMB^*} + \lambda_{HML} \beta_{i,HML} + \lambda_{CMA} \beta_{i,CMA} + \lambda_{RMW} \beta_{i,RMW} + \lambda_{UMD} \beta_{i,UMD}. \tag{23}$$

The sixth model is the 6-factor model proposed by Barillas and Shanken (2018) (BS6),

$$\begin{aligned} \mathrm{E}(R_{i,t+1} - R_{f,t+1}) &= \lambda_M \beta_{i,M} + \lambda_{SMB^*} \beta_{i,SMB^*} + \lambda_{HML^*} \beta_{i,HML^*} + \lambda_{UMD} \beta_{i,UMD} \\ &+ \lambda_{IA} \beta_{i,IA} + \lambda_{ROE} \beta_{i,ROE}, \end{aligned} \tag{24}$$

which combines some of the factors contained in the HXZ4, FF5, and C4. Barillas and Shanken (2018) employ the more timely version of HML (denoted by HML^*) proposed by Asness and Frazzini (2013).

Finally, we estimate the 4-factor model of Stambaugh and Yuan (2017) (SY4):

$$\mathrm{E}(R_{i,t+1} - R_{f,t+1}) = \lambda_M \beta_{i,M} + \lambda_{SMB^{**}} \beta_{i,SMB^{**}} + \lambda_{MGMT} \beta_{i,MGMT} + \lambda_{PERF} \beta_{i,PERF}. \tag{25}$$

[35]See Fama and French (2015) and Hou, Xue, and Zhang (2015) for details.
[36]See Fama and French (2015) for details.

This model contains a size factor (SMB^{**}), plus two "mispricing" factors ($MGMT$ and $PERF$). $MGMT$ is constructed from six anomalies that are associated with firms' management actions, while $PERF$ is constructed from five anomalies associated with firms' performance.

5.2 Data

The data on the equity factors, RM, SMB/SMB^{*}, HML, UMD, CMA, and RMW are obtained from Kenneth French's data library. The data associated with ME, IA, and ROE are obtained from Lu Zhang. The data on HML^{*} are retrieved from the AQR data library, while the data on SMB^{**}, $MGMT$, and $PERF$ are obtained from Yu Yuan's webpage. The descriptive statistics, which are presented in the Internet Appendix, show that the factors with the largest mean are UMD, $MGMT$, and $PERF$, all with average returns above 0.60% per month. The factors with the lowest average are RMW, ME, SMB, and SMB^{*} (around or below 0.30% per month), which confirms previous evidence that the size premium has declined over time. The most volatile factors are the equity premium and the momentum factor, with standard deviations above 4% per month. On the other hand, the investment-based factors (IA and CMA) and EG are the least volatile, with standard deviations below 2% per month. All the factors have relatively low serial correlation, as shown by the small first-order autoregressive coefficients (magnitudes below 20% in all cases).

Regarding the pairwise correlations among the empirical factors (also displayed in the Internet Appendix), it turns out that the different versions of the size factors (SMB, SMB^{*}, ME, and SMB^{**}) are strongly correlated, as indicated by the correlations quite close to 1. We also observe a similar pattern for the asset growth factors (IA and CMA), as shown by the correlation of 0.91. In comparison, the two profitability factors (ROE and RMW) show a smaller correlation (0.67), and the same roughly holds for the two value factors (0.77). Both investment factors are positively correlated with HML (correlations close to 0.70). This indicates that there is some degree of comovement among the value- and investment-based strategies (see also Fama and French 2015; Hou, Xue, and Zhang 2015; Light, Maslov, and Rytchkov 2017; Maio and Santa-Clara 2017). On the other hand, the other value factor (HML^{*}) is less correlated with both investment factors (below 0.50). ROE is positively correlated with UMD (0.49), but the same does not occur with RMW. Hence this suggests that the two profitability factors to a large extent measure different types of risk premia. Additionally, the momentum factor is negatively correlated with HML^{*} (-0.65), but the same does not occur with HML. Further, $MGMT$ is positively correlated with both HML and the investment factors (estimates above 0.70). There is a positive correlation between $PERF$ and both ROE and UMD (above 0.60), whereas $PERF$ is negatively correlated with HML^*. Interestingly, UMD and HML^* are negatively correlated (-0.65), but the same does not occur with HML. Overall, the evidence is that there are relevant correlations among several of the empirical factors.

We run spanning regressions for each of the six statistical factors (in our benchmark model) against the seven multifactor models presented above. The objective is to check if the factors in each of those seven models span our statistical factors (see Barillas and Shanken 2017; Fama and French 2018; Hou, Mo, Xue, and Zhang 2019, among others). The results presented in the online appendix show that there is a substantial amount of information in the statistical factors that is not spanned or subsumed by the empirical factor models. This suggests that the pricing performance of the empirical models (for the cross-section of 42 anomalies) will deviate considerably from that of our benchmark model.

5.3 Cross-sectional asset pricing tests

Next, we test the multifactor models presented above for the broad cross-section of stock returns. We use the same empirical approaches as for the statistical model estimated in the previous section.

The OLS risk price estimates and R^2_{OLS} associated with the empirical models are presented in Table 6 (Panel A). Table 7 (Panel A) presents the evaluation results associated with the time-series method. At a first glance, the seven models show a positive performance in terms of pricing the broad cross-section of stock returns, as indicated by the R^2_{OLS} estimates in the 0.23-0.48 range, with all estimates being statistically significant. However, this level of fit is smaller than the corresponding OLS explanatory ratio obtained for the APT6 model (51%). The underperformance of the empirical models is especially notable in the cases of C4, HXZ4, FF5, FF6, and BS6, with spreads in R^2_{OLS} around or above 15 percentage points. In comparison, the estimates of ρ^2 associated with the empirical models are on the small 0.96-0.97 interval, with strong significance (1% level) in all seven cases.

Perhaps, more striking is the fact that the R^2_C estimates reported in Table 7 are lower than the OLS counterparts by at least 15 percentage points in most cases, the sole exception being C4. This decline in fit is specially relevant in the cases of HMXZ5 (from 0.48 to 0.22), FF5 (from 0.25 to 0.04), SY4 (from 0.42 to 0.18), and even more so in the case of BS6 (from 0.35 to −0.41). This reveals that several of the OLS factor risk price estimates associated with these models differ substantially from the correct estimates (the corresponding factor means). The restriction on the risk price estimates seems less severe in the case of C4, as the difference between R^2_{OLS} and R^2_C is about nine percentage points.

It turns out that both FF5 and BS6 register a rather poor fit for broad cross-sectional equity risk premia, as indicated by the R^2_C estimates of 4% and −41% respectively. For both models, we do not reject the hypothesis of a zero constrained explanatory ratio (at the 10% level), which indicates that such models are useless when it comes to explaining this cross-section of portfolio returns.[37] Critically, the 6-factor statistical model outperforms the seven empirical models when we impose the restriction on the factor risk price estimates: The gaps in R^2_C (denoted by SR^2_2) vary between 28 percentage points (comparison against HMXZ5) and 91 percentage points (relative to BS6), which represent evidence of large economic significance. By using the p-values obtained from the bootstrap simulation, those spreads in R^2_C are statistically significant at the 5% level in most cases. The sole exception occurs for HMXZ5, in which case there is significance only at the 10% level. One way to assess the economic significance of the additional performance associated with the benchmark statistical model is to compute the ratio in R^2_C: It turns out that the best performing empirical models (HMXZ5 and FF) produce an explanatory ratio that is less than half (around 42-43%) the fit corresponding to APT6.

The results for the estimation with the 252 extreme portfolios are displayed in Table 6 (Panel B) and Panel B of Table 7. We can see that both the R^2_{OLS} and R^2_C estimates are somewhat higher than the corresponding estimates obtained in the benchmark test with the 420 portfolios. However, it is still the case that APT6 produces a higher OLS explanatory ratio than these models. The unique model with a comparable performance is HMXZ5 with a R^2_{OLS} of 0.54 (versus 0.58 for APT6). Regarding the estimates of the alternative OLS explanatory ratio (ρ^2), results reported in the appendix show that APT6 delivers a higher fit in all cases, although the differences are relatively small (below three percentage points). Despite the small magnitudes, those differences in ρ^2 are statistically significant at the 1% (comparison with C4, HXZ4, FF5, FF6, and BS6) or 5% level

[37] In untabulated results, all seven models dominate the baseline CAPM in statistical terms, as indicated by small empirical p-values associated with the difference in R^2_C relative to the CAPM. However, this represents a very low hurdle to evaluate the performance of multifactor models for this specific cross-section of risk premia.

(SY4). The unique exception to this pattern is in the comparison with HMXZ5, in which case the estimate of $S\rho^2$ (of 0.4%) is not significant at the 10% level (p-value of 0.51).[38]

Perhaps more importantly, when relying on the more correct R_C^2 metric, the statistical model tends to outperform the empirical models in an economically significant way, with gaps in explanatory ratio that vary between 25 percentage points (comparison with HMXZ5) and 86 percentage points (comparison with BS6). Such differences in R_C^2 are statistically significant at the 5% level in most cases. Again, the sole exception is HMXZ5, in which case the difference in explanatory ratio is not even significant at the 10% level (yet marginally so, with a p-value of 0.103).[39] To have another perspective on the significance of the relative performance, only HMXZ5 originates an explanatory ratio that is more than half (55%) the corresponding estimate for the statistical model.

Therefore, these results largely suggest that most of the empirical factor models clearly lag behind (both in economic and statistical terms) the reference statistical model. The only exception is HMXZ5, which is not dominated in statistical terms (at the 5% level) by the benchmark model. It is interesting to see that apart from FF5 and BS6, the empirical models produce explanatory ratios that are above zero in statistical terms, as indicated by the associated p-values below 5%. However, this represents a rather low hurdle when it comes to judging the performance of those models. A considerably more demanding, but also more relevant, metric is to quantify the difference in explanatory power relative to our benchmark model (and the corresponding statistical significance).[40]

Several other results from Table 7 deserve some discussion. First, both HMXZ5 and FF6 have a similar performance in terms of explaining the complete cross-section of 420 portfolios, as indicated by the constrained explanatory ratios of 0.22 and 0.21, respectively. However, while the first model is not dominated in statistical terms by APT6 (at the 5% level), the second model clearly underperforms the benchmark model (p-values of 0-0.01). In other words, we cannot discriminate between HMXZ5 and APT6 from a statistical point of view, while we are able to do so when comparing FF6 and APT6. This suggests that the factors contained in HMXZ5 are closer (in statistical terms) to the PCA factors than the factors employed in FF6. Second, the expected growth

[38]In the Internet Appendix, we assess the statistical significance of the difference in OLS R^2 between APT6 and the other models by using an unrestricted zero-beta rate, as in Kan, Robotti, and Shanken (2013).

[39]We get identical results by using the 9-factor APT model instead of APT6.

[40]Untabulated results show that both HMXZ5 and SY4 perform similarly to APT6 in what concerns the number of anomalies that pass the GRS-test (about 31-32), while the other empirical models do worse on such metric. However, we note that such measure provides only an approximate assessment of the joint pricing power for the 420 portfolios.

factor is determinant for the performance of HMXZ5: HXZ4 clearly lags behind the statistical model, as indicated by the differences in R_C^2 above 33 percentage points, which are significant at the 1% or 5% level.

Third, although most models are dominated statistically by the benchmark model, BS6 stands up in terms of negative performance when it comes to explaining the 42 anomalies.[41] Indeed, this 6-factor model is the only empirical model producing negative explanatory ratios, that is, its performance is worse than a trivial model predicting constant risk premia in the cross-section. The differences in R_C^2 (relative to the statistical model) are around 90 percentage points, and these estimates are significant at the 1% level (p-values around zero). To get another perspective of the underperformance of such model, unreported results show that the negative spreads in R_C^2 relative to the other six empirical models are statistically significant at the 1% level in all cases. Therefore, our findings suggest that different criteria (cross-sectional asset pricing test versus Bayesian asset pricing test) can lead to quite opposite results in terms of ranking alternative factor models.[42]

The results of this section also imply that there is a dramatic room for improving most of the current workhorses in the literature in terms of pricing a broad cross-section of equity risk premia. This can include adding new empirical factors or replacing (some of) the existing factors by better factors.[43] The only potential exception to this pattern appears to be HMXZ5, which is not dominated in statistical terms by the benchmark statistical model, especially when we focus attention on the more interesting extreme deciles in each anomaly.[44]

We note that our statistical model represents only an approximation to the underlying true multifactor model governing our cross-section of risk premia. This is well illustrated by the fact that the cross-sectional R^2 produced by APT6 is substantially below one. To obtain a statistical model that would generate explanatory ratios closer to one, we would need to include additional statistical

[41] This is consistent with the evidence provided in Hou, Mo, Xue, and Zhang (2020).

[42] Barillas and Shanken (2018) use a Bayesian procedure to obtain the best combination of the factors associated with C4, FF5, HXZ4, as well as HML^*. Yet, their statistical procedure does not take into account the testing assets.

[43] Another possibility is to derive and estimate conditional versions of the current factors models, which is the route adopted in Cooper and Maio (2019b).

[44] On a related note, we expect a priori that SY4 would be the most obvious candidate when it comes to reaching equivalent (in statistical terms) pricing performance to the benchmark APT model. The reason is that the two key factors in SY4, $PERF$ and $MGMT$, are "composite" factors, that is, they are related by construction with several market anomalies (and thus, with several segments of the cross-section of testing assets). However, our results show that SY4 is statistically dominated by the benchmark model at the 5% level. In other words, constructing factors that are mechanically related with several segments of the cross-section does not imply a large pricing power for the broader set of anomalies.

factors (i.e., higher-order principal components) in our model. However, such higher-dimensional statistical model would be of limited utility to our main objective in the paper—comparison with the empirical models existent in the literature—as the number of factors would be substantially higher than in those models. If anything, the hurdle on the empirical models would be higher, that is, the dominance of such augmented statistical model over these models would be even stronger than what we already document in this section.

We perform several robustness checks to the analysis conducted above. To save space, the results are presented and discussed in detail in the Internet Appendix. Specifically, the main findings documented above are robust to employing an alternative bootstrap simulation to assess the statistical significance of SR^2, using a shorter time-series in the asset pricing tests, excluding some anomalies from the testing assets, or employing an alternative statistical inference. We also estimate restricted versions of APT6 to gauge the pricing contribution of the different statistical factors.

5.4 Alternative testing portfolios

We estimate both the statistical and empirical factor models by using alternative equity portfolios as testing assets. This represents an "out-of-sample" asset pricing test for our statistical model over the cross-sectional dimension, as we force the model to price different portfolios than those employed in the construction of the statistical factors. The objective is to control for a possible in-sample "over-fitting" of the statistical model, which may cause its good pricing performance, and corresponding statistical dominance over the empirical factor models, to disappear when attempted to explain alternative equity risk premia.

Specifically, we force the model to price other anomalies that are also employed in Green, Hand, and Zhang (2017), but were excluded from the cross-section used in the rest of the paper. Unreported results indicate that the corresponding return spreads associated with the extreme two or three portfolios (on each leg) generate significant (at the 10% level) CAPM alphas.[45] Hence, there is significant cross-sectional dispersion in risk premia to be priced when we consider the full spectrum of portfolios within each of these additional portfolio groups (rather than relying exclusively on

[45] Specifically, the return spread is given by $\frac{1}{3}(r_8 + r_9 + r_{10}) - \frac{1}{3}(r_1 + r_2 + r_3)$ or $\frac{1}{2}(r_9 + r_{10}) - \frac{1}{2}(r_1 + r_2)$, where r_j denotes the return on the jth decile.

the very top and bottom deciles). In total, we consider eight new anomalies corresponding to 80 portfolios, which are described in the internet appendix.

The results, which are tabulated in the Internet Appendix, show that our statistical model produces a R^2_C estimate of 0.34, which is statistically significant at the 1% level (p-value around zero). In comparison, the R^2_C estimates associated with the C4, HXZ4, FF5, and BS6 models are either negative or very close to zero. This means that these four models do worse (or the same) than a trivial model (containing only the intercept) when it comes to pricing the cross-sectional dispersion in risk premia for those 80 portfolios. Critically, the spreads in R^2_C between APT6 and each of those four models are statistically significant (at the 5% or 1% level). The best performing empirical model is SY4, with an explanatory ratio of 0.21, which is significant at the 5% level. The performance of both HMXZ5 and FF6 is somewhat more modest, with a R^2_C estimate of 0.14 in both cases and only for FF6 is there significance at the 5% level. However, despite lagging the statistical model by more than 20 percentage points, both models (as well as SY4) are not statistically dominated by APT6, as indicated by the p-values (associated with SR^2) above 10% in all three cases (marginally so in the case of FF6, with a p-value of 11.5%).

The estimation results for the cross-section including only the extreme deciles associated with the eight portfolio groups point to a qualitatively similar picture to the baseline cross-sectional estimation. The statistical model produces an explanatory ratio of 0.27, which is significant at the 1% level. In comparison, most of the empirical models produce R^2_C estimates that are not above zero in statistical terms. The sole exception is SY4 with an estimate of 0.19, which is significant at the 5% level. Similar to the estimation with the 80 portfolios, the estimates of SR^2 associated with HMXZ5 and FF6, despite showing sizable magnitudes (about 20 percentage points), are not significant at the 10% level. Hence, the large statistical uncertainty implies that we cannot discriminate statistically between APT6 on one side and either HMXZ5 or FF6 on the other side.[46]

Overall, the results of this subsection suggest that the statistical dominance of APT6 against the empirical models does not deteriorate in a substantial way when it comes to pricing "out-of-sample" CAPM anomalies, that is, employing other portfolios (as testing assets) than those used in the

[46]One limitation of the asset pricing test described above is that the dimension of the cross-section is substantially smaller than the "in-sample" tests employed in the rest of the paper (8 versus 42 anomaly groups). To overcome this limitation, we conduct a second "out-of-sample" asset pricing test on the cross-sectional dimension, which includes 11 new anomalies (in addition to the eight groups employed in the first out-of-sample test). The results, which are discussed in the appendix, are qualitatively similar to the baseline out-of-sample asset pricing estimation.

construction of the statistical factors. In fact, the main qualitative results obtained in the previous section remain robust. This also suggests that there is no excessive in-sample "over-fitting" associated with the statistical model. One way for interpreting these results is that the 42 anomaly sorts used in the construction of the PCA factors are fairly representative of the broader cross-section of CAPM anomalies.

6 Out-of-Sample Asset Pricing Tests

In this section, we focus on the stability over time of the pricing performance of both the statistical model and the empirical factor models.[47] This represents an "out-of-sample" asset pricing test over the time-series dimension. The empirical factors are constructed from observable measures while the statistical factors are estimated. It is therefore interesting to examine the pricing performance of empirical factors relative to that of PCA factors out-of-sample, complementing our main analysis in the previous sections.

Specifically, we use factor betas and factor risk prices estimated in-sample to forecast one-month ahead realized portfolio excess returns. Hence, we have a decoupling between the sample used in the estimation of both the factor betas and risk prices and the sample used in the computation of the pseudo pricing errors.[48] We then compute an "out-of-sample" cross-sectional R^2 for both the statistical model and the empirical factor models to evaluate their relative performance in explaining cross-sectional risk premia.

Given the out-of-sample nature of this test, for the statistical model, we re-estimate PCA factors each month t using portfolio returns up to month $t-1$. As in Section 3, we choose the number of PCA factors based on the IC_2 information criterion suggested by Bai and Ng (2002). We refer to this set of PCA factors as the recursive statistical model, which represents our main focus. For comparison purposes, we also consider the fixed statistical model with six statistical factors (APT6) estimated once over the full sample period. If the estimated PCs are relatively stable over time, the pricing performance of the two statistical models will be relatively similar. The sample period for testing starts in 1978:01 so that the first estimation involves 60 months.

[47] We thank the referee for suggesting this analysis.

[48] The pricing errors represent the difference between one-step ahead excess portfolio returns and fitted risk premia, hence the designation of "pseudo" pricing errors. Simin (2008) uses the terminology of "forecast errors".

The estimation procedure of factor betas is as follows. Taking the C4 model as an example, for each portfolio i and for each month t, we estimate in-sample betas by running the following regression using returns up to month $t-1$,

$$R^e_{i,\tau} = \alpha_i + \beta_{i,M} RM_\tau + \beta_{i,SMB} SMB_\tau + \beta_{i,HML} HML_\tau + \beta_{i,UMD} UMD_\tau + \varepsilon_{i,\tau},\ \tau \in [1, t-1], \quad (26)$$

where the notation is self-explanatory. As in Simin (2008), we estimate in-sample factor risk prices as the time-series average of the factor returns up to $t-1$, since all factors are traded. We then calculate the forecasted portfolio risk premium as the sum of the risk premiums (products of estimated beta and risk price) across all factors (using information up to $t-1$),

$$\widehat{R}^e_{i,t} \equiv \widehat{\beta}_{i,M} \overline{RM} + \widehat{\beta}_{i,SMB} \overline{SMB} + \widehat{\beta}_{i,HML} \overline{HML} + \widehat{\beta}_{i,UMD} \overline{UMD},$$

where $\overline{f}$ denotes the trailing mean of factor f.[49]

Defining the out-of-sample one-step ahead forecasting error as $\zeta_{i,t} = R^e_{i,t} - \widehat{R}^e_{i,t}$, we have $R^e_{i,t} = \widehat{R}^e_{i,t} + \zeta_{i,t}$. Averaging over the time-series, we have $\overline{R^e_i} = \overline{\widehat{R}^e_i} + \overline{\zeta}_i$. We then define the out-of-sample cross-sectional R^2 to closely resemble R^2_C in our main analysis, which is given by

$$R^2_{OOS} = 1 - \frac{\text{Var}_N(\overline{\zeta}_i)}{\text{Var}_N(\overline{R^e_i})}. \quad (27)$$

In line with our main analysis conducted in the previous sections, we conduct two sets of estimations using different testing assets. In the first set of analysis, we use all the 420 equity portfolios as testing assets. We find that R^2_{OOS} is 0.34 for the recursive statistical model and 0.24 for the fixed statistical model. In comparison, we obtain explanatory ratios of −0.38, −0.02, 0.14, 0.05, 0.08, −0.37, and −0.15 for the C4, HXZ4, HMXZ5, FF5, FF6, BS6, and SY4 models, respectively. Thus, both statistical models produce a larger fit than that of any of the empirical factor models

[49] This method shares some similarities with the Fama and MacBeth (1973) (FM) method used in cross-sectional tests of asset pricing models. In both methods, both the factor betas and risk price estimates are implicitly allowed to vary over time by using recursive (or rolling) samples. However, there are two key differences relative to the FM procedure. First, in each period, the factor risk price estimates correspond to the recursive (rolling) means of the factors rather than being estimated from a cross-sectional regression of realized excess returns onto factor loadings. Second, the risk price estimates (and corresponding fitted total portfolio risk premia) employed in computing the pricing errors in t are obtained from the recursive (rolling) sample (containing information up to $t-1$) rather than using information from the current period (t).

considered and such difference in performance is substantial in most cases. Indeed, several empirical models (C4, HXZ4, BS6, and SY4) have a very poor performance, as indicated by the negative explanatory ratios. Hence, SY4 performs substantially worse along this metric than in our baseline "in-sample" analysis in the previous section. On the other hand, HMXZ5 stands up as the best performing empirical model on this metric, in line with the evidence in the rest of the paper.

In the second set of estimations, we use the extreme three deciles (on each leg) within each group of decile portfolios as testing assets. We find that R^2_{OOS} is 0.38 for the recursive statistical model and 0.30 for the fixed statistical model. Hence, the performance of both models is marginally higher than in the estimation with all portfolios. In comparison, such metric assumes the values of −0.31, 0.07, 0.20, 0.10, 0.14, −0.25, and −0.09 for the C4, HXZ4, HMXZ5, FF5, FF6, BS6, and SY4 models, respectively. Again, both statistical models show a substantially higher fit than that of any of the empirical factor models considered. Similarly to the estimation with the full-cross-section, it turns out that HMXZ5 stands up as the best empirical model.[50]

Overall, the results from this section suggest that the relative outperformance of a model containing the statistical factors against the empirical factor models remains quite robust out-of-sample.[51]

7 Conclusion

In this paper, we aim to examine systematically the performance of the current multifactor models in the empirical asset pricing literature by using a novel approach. Specifically, we compare these models against a statistical benchmark model that is consistent with the general framework of the Arbitrage Pricing Theory (APT) of Ross (1976). We follow the empirical APT literature in terms

[50] Given that all factor models analyzed in the paper are unconditional, our OOS analysis is based on recursive or expanding windows. However, as a robustness check, we also consider 60-months rolling samples in the estimation of the factor loadings and risk prices, a common practice in the literature. Untabulated results show that the statistical model continues to dominate the empirical models by a large degree. Specifically, in the estimation with all 420 equity portfolios as testing assets, we find that the estimated R^2_{OOS} for the fixed statistical model (APT6) is 0.41. In comparison, we obtain estimates of −0.16, 0.10, 0.15, 0.26, 0.22, −0.16, and −0.02 for the C4, HXZ4, HMXZ5, FF5, FF6, BS6, and SY4 models, respectively. In the estimation with the extreme deciles, the statistical model produces an R^2_{OOS} marginally higher than in the full-cross section test, at 0.46. In comparison, such metric assumes the values of −0.11, 0.14, 0.19, 0.30, 0.27, −0.11, and 0.00 for the C4, HXZ4, HMXZ5, FF5, FF6, BS6, and SY4 models, respectively.

[51] We also consider an alternative measure of out-of-sample R^2 used in the literature (see e.g. Gu, Kelley, and Xiu (2020)). Unreported results indicate that both versions of the statistical model outperform the empirical models based on such new measure of OOS forecasting performance.

of estimating common statistical factors by applying asymptotical principal components analysis (PCA) to a large cross-section of stock returns associated with 42 anomalies or portfolio sorts. market anomalies.

The asset pricing results show that a 6-factor model (denoted by APT6), containing the first, second, fourth, sixth, eighth, and ninth principal components as risk factors, explains 51% of the cross-sectional variation in the risk premia. When we impose the constraint that the factor risk price estimates are equal to the factor means, we obtain a similar fit. We conduct an alternative asset pricing test with 252 portfolios associated with the extreme three deciles (on each leg) for each market anomaly. The results show an even slightly larger fit for the statistical model.

The central analysis in the paper is to compare our APT model to some of the most popular multifactor models existent in the literature (denoted by empirical models) in terms of pricing the 420 portfolios. The asset pricing tests show that the performance of the seven empirical models lags behind the fit of the statistical model, with differences in cross-sectional R^2 between 28 and 91 percentage points. When using the shorter and more interesting cross-section associated with the extreme portfolios, the gaps in explanatory ratio relative to the benchmark model vary between 25 and 86 percentage points. Critically, the APT model dominates most of the empirical models in statistical terms (at the 5% level).

The overall conclusions from this paper are simple, but important. Several of the current empirical workhorses employed in the asset pricing literature fail to be good empirical proxies for the APT. That is, they deviate significantly, both in economic and statistical terms, from a benchmark statistical model that is designed in such a way (APT intuition) to explain well a rich cross-section of equity risk premia. Therefore, assuming that explaining the broad cross-section of stock returns is the main goal of a successful empirical multifactor model, our results suggest that most of the models proposed in the literature fail considerably on such dimension.

Following most of the empirical asset pricing literature, our empirical design relies on a cross-section of equity portfolios sorted on several prominent CAPM anomalies. By design, there is significant cross-sectional dispersion in risk premia to be priced by the candidate models. However, a successful asset pricing model should be able to price the risk premia associated with any risky asset. Examples include individual equities, corporate bonds, Treasury bonds, or currencies. Extending the analysis to a broader cross-section of asset risk premia is left for future research.

Table 1: **List of portfolio sorts**

This table lists the 42 alternative anomalies or portfolio sorts employed in the empirical analysis. "Category" refers to the broad classification employed by Hou, Xue, and Zhang (2015, 2020).

Symbol	Anomaly	Category
acc	Working capital accruals	Investment
agr	Asset growth	Investment
baspread	Bid-ask spread	Trading frictions
beta	Beta	Trading frictions
bm_ia	Industry-adjusted book-to-market ratio	Value-growth
bm	Book-to-market ratio	Value-growth
cashpr	Cash productivity	Profitability
cfp	Cash-flow-to-price ratio	Value-growth
chatoia	Industry-adjusted change in asset turnover	Profitability
chmom	Change in 6-month momentum	Momentum
currat	Current ratio	Intangibles
dolvol	Dollar trading volume	Trading frictions
ear	Earnings announcement return	Momentum
egr	Growth in common shareholder equity	Investment
ep	Earnings-to-price ratio	Value-growth
grcapx	Growth in capital expenditures	Investment
grltnoa	Growth in long-term net operating assets	Investment
idiovol	Idiosyncratic return volatility	Trading frictions
indmom	Industry momentum	Momentum
invest	Capital expenditures and inventory	Investment
lgr	Growth in long-term debt	Investment
maxret	Maximum daily return	Trading frictions
mom6m	6-month momentum	Momentum
mom12m	12-month momentum	Momentum
operprof	Operating profitability	Profitability
pchcapx	% change in capital expenditures	Investment
pchcurrat	% change in current ratio	Intangibles
pchdepr	% change in depreciation	Investment
pchsale_pchinvt	% change in sales − % change in inventory	Intangibles
pchsaleinv	% change in sales-to-inventory ratio	Intangibles
quick	Quick ratio	Intangibles
roaq	Return on assets	Profitability
roeq	Return on equity	Profitability
roic	Return on invested capital	Profitability
rsup	Revenue surprise	Momentum
salecash	Sales-to-cash ratio	Intangibles
saleinv	Sales-to-inventory ratio	Intangibles
salerec	Sales-to-receivables ratio	Intangibles
sgr	Sales growth	Value-growth
sp	Sales-to-price ratio	Value-growth
std_dolvol	Volatility of liquidity	Trading frictions
turn	Share turnover	Trading frictions

Table 2: **Descriptive statistics for statistical factors**

This table reports some descriptive statistics for the common factors ($\widehat{F}_j, j = 1, ..., 9$) estimated from 420 equity portfolios. ϕ designates the first-order autocorrelation coefficient. v_j represents the cumulative proportion of the total variance in the raw portfolio returns explained by the factors $\widehat{F}_1$ to $\widehat{F}_j$. The sample is 1973:01–2016:12.

	$\widehat{F}_1$	$\widehat{F}_2$	$\widehat{F}_3$	$\widehat{F}_4$	$\widehat{F}_5$	$\widehat{F}_6$	$\widehat{F}_7$	$\widehat{F}_8$	$\widehat{F}_9$
Mean(%)	0.53	−0.59	0.01	−0.67	0.09	0.47	−0.03	−0.90	−0.55
ϕ	0.08	0.00	0.11	−0.03	0.11	0.05	0.15	0.04	0.02
v_j	0.85	0.88	0.89	0.90	0.90	0.91	0.91	0.91	0.92

Table 3: **Anomalies and statistical factors**

This table reports R^2 estimates from single regressions of return spreads onto the estimated common factors ($\widehat{F}_j$). The "high-minus-low" return spreads are associated with 42 market anomalies. See Table 1 for a description of the different portfolio sorts. The sample is 1973:01–2016:12.

	$\widehat{F}_1$	$\widehat{F}_2$	$\widehat{F}_3$	$\widehat{F}_4$	$\widehat{F}_5$	$\widehat{F}_6$	$\widehat{F}_7$	$\widehat{F}_8$	$\widehat{F}_9$
acc	0.00	0.06	0.01	0.02	0.03	0.02	0.05	0.01	0.01
agr	0.03	0.04	0.18	0.14	0.08	0.00	0.00	0.00	0.09
baspread	0.42	0.45	0.02	0.00	0.00	0.00	0.01	0.00	0.00
beta	0.50	0.38	0.00	0.00	0.03	0.00	0.00	0.00	0.00
bm_ia	0.07	0.18	0.01	0.01	0.03	0.00	0.00	0.15	0.00
bm	0.01	0.00	0.59	0.16	0.00	0.00	0.01	0.00	0.03
cashpr	0.04	0.11	0.42	0.05	0.01	0.04	0.02	0.00	0.02
cfp	0.05	0.28	0.02	0.00	0.08	0.00	0.06	0.05	0.06
chatoia	0.01	0.02	0.00	0.03	0.11	0.00	0.01	0.01	0.00
chmom	0.09	0.01	0.01	0.14	0.01	0.03	0.00	0.33	0.00
currat	0.37	0.28	0.01	0.00	0.03	0.03	0.03	0.01	0.01
dolvol	0.00	0.08	0.45	0.09	0.00	0.06	0.09	0.02	0.00
ear	0.01	0.00	0.11	0.05	0.00	0.01	0.02	0.00	0.00
egr	0.06	0.05	0.10	0.13	0.13	0.00	0.01	0.01	0.08
ep	0.10	0.32	0.05	0.01	0.07	0.03	0.00	0.02	0.06
grcapx	0.04	0.12	0.10	0.12	0.01	0.03	0.00	0.02	0.03
grltnoa	0.00	0.09	0.08	0.04	0.00	0.04	0.03	0.01	0.00
idiovol	0.41	0.52	0.00	0.01	0.01	0.00	0.00	0.00	0.00
indmom	0.01	0.00	0.17	0.38	0.01	0.00	0.00	0.04	0.00
invest	0.04	0.01	0.08	0.04	0.11	0.08	0.02	0.06	0.02
lgr	0.03	0.03	0.10	0.11	0.04	0.00	0.00	0.01	0.11
maxret	0.37	0.47	0.01	0.00	0.00	0.00	0.01	0.01	0.00
mom6m	0.07	0.02	0.24	0.47	0.00	0.03	0.00	0.03	0.00
mom12m	0.05	0.02	0.34	0.42	0.01	0.01	0.00	0.02	0.00
operprof	0.00	0.03	0.26	0.05	0.04	0.07	0.10	0.01	0.00
pchcapx	0.04	0.17	0.08	0.05	0.00	0.05	0.00	0.00	0.03
pchcurrat	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00
pchdepr	0.00	0.08	0.01	0.02	0.05	0.00	0.00	0.02	0.01
pchsale_pchinvt	0.00	0.01	0.00	0.06	0.01	0.00	0.02	0.00	0.01
pchsaleinv	0.00	0.01	0.01	0.05	0.01	0.00	0.01	0.00	0.00
quick	0.24	0.35	0.02	0.00	0.00	0.09	0.00	0.00	0.02
roaq	0.11	0.29	0.25	0.00	0.11	0.00	0.00	0.00	0.01
roeq	0.09	0.33	0.21	0.01	0.15	0.00	0.00	0.00	0.01
roic	0.14	0.25	0.17	0.07	0.09	0.01	0.01	0.00	0.00
rsup	0.00	0.01	0.09	0.06	0.04	0.03	0.02	0.00	0.06
salecash	0.11	0.14	0.01	0.03	0.00	0.39	0.01	0.04	0.00
saleinv	0.26	0.06	0.03	0.00	0.08	0.05	0.00	0.02	0.00
salerec	0.06	0.01	0.13	0.01	0.01	0.36	0.15	0.01	0.00
sgr	0.04	0.05	0.24	0.05	0.12	0.00	0.00	0.00	0.10
sp	0.01	0.01	0.48	0.15	0.01	0.07	0.01	0.00	0.02
std_dolvol	0.00	0.13	0.25	0.14	0.01	0.02	0.14	0.00	0.00
turn	0.36	0.42	0.03	0.00	0.01	0.00	0.01	0.00	0.00

Table 4: **APT model: Factor risk premia estimates**

This table reports the factor risk price estimates for the APT 9- and 6-factor models. The common factors are estimated from PCA applied to 420 equity portfolios. The empirical method is the two-step regression approach, where the second step consists of an OLS cross-sectional regression of average portfolio excess returns on factor betas. The testing assets represent 420 portfolios associated with 42 portfolio sorts. See Table 1 for a description of the different portfolio sorts. In Panel B, the testing assets are the extreme three deciles (on each leg) for each group. λ_j denotes the risk price estimate (in %) for the jth common factor ($\widehat{F}_j$). Below the risk price estimates are displayed t-statistics based on Shanken's standard errors (in parentheses). The column labeled R^2_{OLS} denotes the cross-sectional OLS R^2. The values in parentheses denote empirical p-values (obtained from a bootstrap simulation) for the null hypothesis $R^2_{OLS} = 0$. ρ^2 represents an alternative cross-sectional OLS R^2, with the values in parentheses denoting asymptotic p-values for the null hypothesis $\rho^2 = 0$. The sample is 1973:01–2016:12. Italic, underlined, and bold t-ratios denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	R^2_{OLS}	ρ^2
						A. All deciles					
1	0.53	−0.58	0.01	−0.64	0.08	0.43	−0.03	−0.78	−0.48	0.51	0.97
	(**2.81**)	(**−3.08**)	(0.04)	(**−3.41**)	(0.41)	(2.26)	(−0.14)	(**−4.18**)	(−2.54)	(0.000)	(0.004)
2	0.53	−0.58		−0.64		0.42		−0.78	−0.48	0.51	0.97
	(**2.81**)	(**−3.08**)		(**−3.41**)		(2.26)		(**−4.18**)	(−2.53)	(0.000)	(0.004)
						B. Extreme deciles					
1	0.53	−0.59	−0.00	−0.62	−0.03	0.48	−0.08	−0.78	−0.59	0.58	0.97
	(**2.81**)	(**−3.15**)	(−0.01)	(**−3.31**)	(−0.14)	(2.53)	(−0.41)	(**−4.16**)	(**−3.09**)	(0.001)	(0.006)
2	0.53	−0.59		−0.62		0.48		−0.79	−0.59	0.58	0.97
	(**2.81**)	(**−3.15**)		(**−3.31**)		(2.53)		(**−4.17**)	(**−2.98**)	(0.000)	(0.006)

Table 5: **Time-series tests for APT model**

This table reports the evaluation results for the 6-factor APT model (APT6). The empirical method is time-series regressions applied to each testing portfolio. The testing assets represent 420 portfolios associated with 42 portfolio sorts (Panel A). See Table 1 for a description of the different portfolio sorts. In Panel B, the testing assets are the extreme three deciles (on each leg) for each group. MAA is the mean absolute alpha (in %). $\#GRS$ denotes the number of portfolio groups in which the model is not rejected by the GRS-test (at the 5% level). $\#t$ represents the number of portfolios with statistically significant alphas (at the 5% level). R^2_C is the cross-sectional constrained R^2 and the numbers in parentheses represent the respective empirical p-values to test the null that the explanatory ratio is zero (obtained from a bootstrap simulation). SR^2 represents the difference in R^2_C between APT6 and the single-factor statistical model (APT1), with the respective empirical p-values presented in parenthesis. The sample is 1973:01–2016:12.

	A. All deciles		B. Extreme deciles	
	APT1	APT6	APT1	APT6
$MAA(\%)$	0.13	0.08	0.16	0.08
$\#GRS$	13	32	12	26
$\#t$	122	42	95	29
R^2_C	−0.58	0.50	−0.60	0.58
	(1.000)	(0.000)	(1.000)	(0.000)
SR^2		1.09		1.18
		(0.000)		(0.000)

Table 6: **Factor risk premia estimates for empirical models**

This table reports the factor risk price estimates for the empirical multifactor models. The empirical method is the two-step regression approach, where the second step consists of an OLS cross-sectional regression of average portfolio excess returns on factor betas. In Panel A, the testing assets represent 420 portfolios associated with 42 portfolio sorts. In Panel B, the testing assets represent 252 portfolios (extreme three deciles on each leg) associated with the 42 portfolio sorts. See Table 1 for a description of the different portfolio groups. λ_M, $\lambda_{SMB}/\lambda_{SMB^*}$, λ_{HML}, and λ_{UMD} denote the risk price estimates (in %) for the market, size, value, and momentum factors, respectively. λ_{ME}, λ_{IA}, λ_{ROE}, and λ_{EG} represent the risk prices associated with the Hou–Mo–Xue–Zhang size, investment, profitability, and expected growth factors, respectively. λ_{RMW} and λ_{CMA} denote the risk price estimates for the Fama–French profitability and investment factors. $\lambda_{SMB^{**}}$, λ_{MGMT}, and λ_{PERF} represent the risk prices for the Stambaugh–Yuan size, management and performance factors, respectively. λ_{HML^*} is the risk price for the Asness–Frazzini value factor. Below the risk price estimates are displayed t-statistics based on Shanken's standard errors (in parentheses). The column labeled R^2_{OLS} denotes the baseline cross-sectional OLS R^2. The values in parentheses denote empirical p-values (obtained from a bootstrap simulation) for the null hypothesis $R^2_{OLS} = 0$. ρ^2 represents an alternative cross-sectional OLS R^2, with the values in parentheses denoting asymptotic p-values for the null hypothesis $\rho^2 = 0$. The sample is 1973:01–2016:12. Italic, underlined, and bold t-ratios denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	λ_M	$\lambda_{SMB}/\lambda_{SMB^*}/\lambda_{SMB^{**}}$	$\lambda_{HML}/\lambda_{HML^*}$	λ_{UMD}	λ_{RMW}	λ_{CMA}	λ_{ME}	λ_{IA}	λ_{ROE}	λ_{EG}	λ_{MGMT}	λ_{PERF}	R^2_{OLS}	ρ^2
A. All deciles														
C4	0.56	−0.04	0.35	0.53									0.23	0.96
	(**2.81**)	(−0.25)	(2.37)	(2.57)									(0.007)	(0.004)
HXZ4	0.53						0.28	0.25	0.39				0.31	0.96
	(**2.66**)						(*1.83*)	(2.29)	(**2.85**)				(0.001)	(0.004)
HMXZ5	0.55						0.34	0.28	0.31	0.45			0.48	0.97
	(**2.75**)						(2.15)	(2.51)	(2.28)	(**3.57**)			(0.000)	(0.004)
FF5	0.52	0.21	0.06		0.26	0.31							0.25	0.96
	(**2.60**)	(1.49)	(0.41)		(2.22)	(**3.08**)							(0.007)	(0.004)
FF6	0.54	0.21	0.17	0.48	0.21	0.25							0.36	0.97
	(**2.69**)	(1.43)	(1.22)	(2.35)	(*1.79*)	(2.53)							(0.004)	(0.004)
BS6	0.54	0.17	−0.00	0.52				0.20	0.32				0.35	0.96
	(**2.70**)	(1.14)	(−0.00)	(2.52)				(1.96)	(2.47)				(0.003)	(0.004)
SY4	0.55	0.31									0.36	0.45	0.42	0.97
	(**2.74**)	(2.15)									(2.35)	(2.29)	(0.000)	(0.004)
B. Extreme deciles														
C4	0.55	−0.05	0.36	0.53									0.27	0.94
	(**2.77**)	(−0.29)	(2.45)	(**2.58**)									(0.011)	(0.006)
HXZ4	0.52						0.32	0.25	0.44				0.39	0.95
	(**2.60**)						(2.02)	(2.32)	(**3.08**)				(0.001)	(0.006)
HMXZ5	0.54						0.36	0.28	0.34	0.47			0.54	0.96
	(**2.72**)						(2.19)	(2.53)	(2.45)	(**3.63**)			(0.000)	(0.006)
FF5	0.51	0.27	0.04		0.34	0.30							0.32	0.95
	(2.54)	(*1.82*)	(0.29)		(**2.85**)	(**2.95**)							(0.008)	(0.006)
FF6	0.52	0.25	0.16	0.27	0.24	0.49							0.42	0.96
	(**2.63**)	(*1.71*)	(1.13)	(2.34)	(2.41)	(2.36)							(0.005)	(0.006)
BS6	0.52	0.21	−0.03	0.51				0.20	0.38				0.41	0.95
	(**2.63**)	(1.45)	(−0.17)	(2.50)				(2.02)	(**2.92**)				(0.004)	(0.006)
SY4	0.54	0.32									0.37	0.48	0.47	0.96
	(**2.72**)	(2.11)									(2.41)	(2.39)	(0.000)	(0.006)

Table 7: **Time-series tests for empirical models**

This table reports the evaluation results for the empirical multifactor models. The empirical method is time-series regressions applied to each testing portfolio. The testing assets represent 420 portfolios associated with 42 portfolio sorts (Panel A). See Table 1 for a description of the different portfolio sorts. In Panel B, the testing assets are the extreme three deciles (on each leg) for each group. R^2_C is the cross-sectional constrained R^2 and the numbers in parentheses represent the respective empirical p-values to test the null that the explanatory ratio is zero (obtained from a bootstrap simulation). SR^2 represents the difference in R^2_C between the 6-factor APT model (APT6) and each multifactor model, with the respective empirical p-values presented in parenthesis. "Ratio" denotes the ratio in R^2_C between a given empirical model and APT6. This measure is not computed for the BS6 model, as such model produces a negative R^2_C estimate. The multifactor models are the Carhart 4-factor model (C4), Hou–Xue–Zhang 4-factor model (HXZ4), Hou–Mo–Xue–Zhang 5-factor model (HMXZ5), Fama–French 5-factor model (FF5), Fama–French 6-factor model (FF6), Barillas–Shanken 6-factor model (BS6), and the Stambaugh–Yuan 4-factor model (SY4). The sample is 1973:01–2016:12.

	APT6	C4	HXZ4	HMXZ5	FF5	FF6	BS6	SY4
				A. All deciles				
R^2_C	0.50	0.14	0.14	0.22	0.04	0.21	−0.41	0.18
	(0.000)	(0.007)	(0.016)	(0.004)	(0.127)	(0.002)	(0.990)	(0.007)
SR^2		0.36	0.36	0.28	0.46	0.30	0.91	0.33
		(0.001)	(0.010)	(0.076)	(0.000)	(0.013)	(0.000)	(0.025)
Ratio		0.28	0.28	0.44	0.08	0.42	NA	0.36
				B. Extreme deciles				
R^2_C	0.58	0.18	0.25	0.32	0.11	0.29	−0.28	0.25
	(0.000)	(0.004)	(0.001)	(0.001)	(0.030)	(0.000)	(0.945)	(0.002)
SR^2		0.40	0.33	0.25	0.47	0.29	0.86	0.33
		(0.001)	(0.014)	(0.103)	(0.001)	(0.019)	(0.000)	(0.027)
Ratio		0.31	0.43	0.55	0.19	0.50	NA	0.43

References

Ahn, S. C., and A. R. Horenstein. 2013. Eigenvalue ratio test for the number of factors. *Econometrica* 81: 1203–1227.

Ahn, S. C., A. R. Horenstein, and N. Wang. 2018. Beta matrix and common factors in stock returns. *Journal of Financial and Quantitative Analysis* 53: 1417–1440.

Ang, A., J. Liu, and K. Schwarz. 2020. Using stocks or portfolios in tests of factor models. *Journal of Financial and Quantitative Analysis* 55:709–750.

Asness, C., and A. Frazzini. 2013. The devil in HML's details. *Journal of Portfolio Management* 39: 49–68.

Back, K. E. 2017. *Asset pricing and portfolio choice theory.* New York, NY: Oxford University Press.

Bai, J., and S. Ng. 2002. Determining the number of factors in approximate factor models. *Econometrica* 70: 191–221.

Balakrishnan, K., E. Bartov, and L. Faurel. 2010. Post loss/profit announcement drift. *Journal of Accounting and Economics* 50: 20–41.

Barillas, F., and J. Shanken. 2017. Which alpha? *Review of Financial Studies* 30: 1316–1338.

Barillas, F., and J. Shanken. 2018. Comparing asset pricing models. *Journal of Finance* 73: 715–754.

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

Basu, S. 1983. The relationship between earnings yield, market value, and return for NYSE common stocks: Further evidence. *Journal of Financial Economics* 12: 129–156.

Belo, F., and X. Lin. 2011. The inventory growth spread. *Review of Financial Studies* 25: 278–313.

Black, F., M. C. Jensen, and M. S. Scholes. 1972. The capital asset pricing model: Some empirical tests. In *Studies in the theory of capital markets*, eds. M. C. Jensen, 79–121. Santa Barbara, CA: Praeger Publishers.

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

Brennan, M. J., A. W. Wang, and Y. Xia. 2004. Estimation and test of a simple model of intertemporal capital asset pricing. *Journal of Finance* 59: 1743–1775.

Campbell, J. Y., J. Hilscher, and J. Szilagyi. 2008. In search of distress risk. *Journal of Finance* 63: 2899–2939.

Campbell, J. Y., and T. Vuolteenaho. 2004. Bad beta, good beta. *American Economic Review* 94: 1249–1275.

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

Chamberlain, G., and M. Rothschild. 1983. Arbitrage, factor structure, and mean variance analysis on large asset markets. *Econometrica* 51: 1281–1304.

Chen, Z., G. Connor, and R. A. Korajczyk. 2018. A performance comparison of large-**n** factor estimators. *Review of Asset Pricing Studies* 8: 153–182.

Chen, N. F., R. Roll, and S. A. Ross. 1986. Economic forces and the stock market. *Journal of Business* 59: 383–403.

Cochrane, J. H. 2005. *Asset pricing*. Princeton, NJ: Princeton University Press.

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

Connor, G., and R. A. Korajczyk. 1986. Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics* 15: 373–394.

Connor, G., and R. A. Korajczyk. 1988. Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics* 21: 255–289.

Connor, G., R. A. Korajczyk, and R. T. Uhlaner. 2015. A synthesis of two factor estimation methods. *Journal of Financial and Quantitative Analysis* 50: 825–842.

Cooper, I., and P. Maio. 2019a. Asset growth, profitability, and investment opportunities. *Management Science* 65, 3988–4010.

Cooper, I., and P. Maio. 2019b. New evidence on conditional factor models. *Journal of Financial and Quantitative Analysis* 54: 1975–2016.

Cooper, M. J., H. Gulen, and M. J. Schill. 2008. Asset growth and the cross-section of stock returns. *Journal of Finance* 63: 1609–1651.

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

Delikouras, S., and A. Kostakis. 2019. A single-factor consumption-based asset pricing model. *Journal of Financial and Quantitative Analysis* 54: 789–827.

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

Fama, E. F., and K. R. French. 1996. Multifactor explanations of asset pricing anomalies. *Journal of Finance* 51: 55–84.

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

Fama, E. F., and K. R. French. 2016. Dissecting anomalies with a five-factor model. *Review of Financial Studies* 29: 69–103.

Fama, E. F., and K. R. French. 2018. Choosing factors. *Journal of Financial Economics* 128: 234–252.

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

Feng, G., S. Giglio, and D. Xiu. 2020. Taming the factor zoo: A test of new factors. *Journal of Finance* 75: 1327–1370.

Foster, G., C. Olsen, and T. Shevlin. 1984. Earnings releases, anomalies, and the behavior of security returns. *Accounting Review* 59: 574–603.

Gibbons, M. R., S. A. Ross, and J. Shanken. 1989. A test of the efficiency of a given portfolio. *Econometrica* 57: 1121–1152.

Giglio, S., and D. Xiu. 2019. Asset pricing with omitted factors. Working Paper, Yale School of Management.

Goyal, A., C. Pérignon, and C. Villa. 2008. How common are common return factors across the NYSE and Nasdaq? *Journal of Financial Economics* 90: 252–271.

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

Gu, S., B. T. Kelly, and D. Xiu. 2020. Empirical asset pricing via machine learning. *Review of Financial Studies* 33: 2223–2273.

Guo, H., and P. Maio. 2020. ICAPM and the accruals anomaly. *Quarterly Journal of Finance* 10: 2050014.

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

Hafzalla, N., R. Lundholm, and E. M. Van Winkle. 2011. Percent accruals. *Accounting Review* 86: 209–236.

Hou, K., H. Mo, C. Xue, and L. Zhang. 2019. Which factors? *Review of Finance* 23: 1–35.

Hou, K., H. Mo, C. Xue, and L. Zhang. Forthcoming. An augmented q-factor model with expected growth. *Review of Finance*.

Hou, K., C. Xue, and L. Zhang. 2015. Digesting anomalies: An investment approach. *Review of Financial Studies* 28: 650–705.

Hou, K., C. Xue, and L. Zhang. 2020. Replicating anomalies. *Review of Financial Studies* 33: 2019–2133.

Ingersoll, J.E. 1987. *Theory of financial decision making*. Lanham, MD: Rowman & Littlefield Publishers.

Jagannathan, R., and Z. Wang. 1996. The conditional CAPM and the cross-section of expected returns. *Journal of Finance* 51: 3–53.

Jagannathan, R., and Z. Wang. 1998. An asymptotic theory for estimating beta-pricing models using cross-sectional regressions. *Journal of Finance* 53: 1285–1309.

Jegadeesh, N., and J. Livnat. 2006. Revenue surprises and stock returns. *Journal of Accounting and Economics* 41: 147–171.

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

Kan, R., C. Robotti, and J. Shanken. 2013. Pricing model performance and the two-pass cross-sectional regression methodology. *Journal of Finance* 68: 2617–2649.

Kan, R., and C. Zhang. 1999. Two-pass tests of asset pricing models with useless factors. *Journal of Finance* 54: 203–235.

Kandel, S., and R. F. Stambaugh. 1995. Portfolio inefficiency and the cross section of expected returns. *Journal of Finance* 50: 157–184.

Kelly, B. T., S. Pruitt, and Y. Su. 2019. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics.* 134: 501–524.

Kim, S., R. A. Korajczyk, and A. Neuhierl. Forthcoming. Arbitrage portfolios. *Review of Financial Studies.*

Kim, S., and G. Skoulakis. 2018. Ex-post risk premia estimation and asset pricing tests using large cross sections: The regression-calibration approach. *Journal of Econometrics* 204: 159–188.

Kirby, C. 2020. Firm characteristics, cross-sectional regression estimates, and asset pricing tests. *Review of Asset Pricing Studies* 10: 290–334.

Kozak, S., S. Nagel, and S. Santosh. 2018. Interpreting factor models. *Journal of Finance* 73: 1183–1223.

Lakonishok, J., A. Shleifer, and R. W. Vishny. 1994. Contrarian investment, extrapolation, and risk. *Journal of Finance* 49: 1541–1578.

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

Lettau, M., M. Maggiori, and M. Weber. 2014. Conditional risk premia in currency markets and other asset classes. *Journal of Financial Economics* 114: 197–225.

Lewellen, J., and S. Nagel. 2006. The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics* 82: 289–314.

Lewellen, J., S. Nagel, and J. Shanken. 2010. A skeptical appraisal of asset-pricing tests. *Journal of Financial Economics* 96: 175–194.

Light, N., D. Maslov, and O. Rytchkov. 2017. Aggregation of information about the cross section of stock returns: A latent variable approach. *Review of Financial Studies* 30: 1339–1381.

Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47: 13–37.

Lioui, A., and P. Maio. 2014. Interest rate risk and the cross-section of stock returns. *Journal of Financial and Quantitative Analysis* 49: 483–511.

Ludvigson, S. C. 2013. Advances in consumption-based asset pricing: Empirical tests. In *Handbook of the economics of finance*, eds. G. Constantinides, M. Harris, and R. Stulz, 799–906. Amsterdam: Elsevier.

Ludvigson, S. C., and S. Ng. 2007. The empirical risk-return relation: A factor analysis approach. *Journal of Financial Economics* 83: 171–222.

Ludvigson, S. C., and S. Ng. 2009. Macro factors in bond risk premia. *Review of Financial Studies* 22: 5027–5067.

Ludvigson, S. C., and S. Ng. 2010. A factor analysis of bond risk premia. In *Handbook of empirical economics and finance*, eds. A. Ulah and D. E. A. Giles, 313–372. Boca Raton, FL: Chapman and Hall.

Maio, P. 2013a. Intertemporal CAPM with conditioning variables. *Management Science* 59: 122–141.

Maio, P. 2013b. Return decomposition and the Intertemporal CAPM. *Journal of Banking & Finance* 37: 4958–4972.

Maio, P., and D. Philip. 2015. Macro variables and the components of stock returns. *Journal of Empirical Finance* 33: 287–308.

Maio, P., and D. Philip. 2018. Economic activity and momentum profits: Further evidence. *Journal of Banking & Finance* 88: 466–482.

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

Maio, P., and P. Santa-Clara. 2017. Short-term interest rates and stock market anomalies. *Journal of Financial and Quantitative Analysis* 52: 927–961.

Merton, R. C. 1973. An intertemporal capital asset pricing model. *Econometrica* 41: 867–887.

Miller, M. H., and F. Modigliani. 1961. Dividend policy, growth, and the valuation of shares. *Journal of Business* 34: 411–433.

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

Pennacchi, G. 2008. *Theory of asset pricing*. Boston, MA: Pearson Addison Wesley.

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

Pukthuanthong, K., R. Roll, and A. Subrahmanyam. 2019. A protocol for factor identification. *Review of Financial Studies* 32: 1573–1607.

Reisman, H. 1988. A general approach to the arbitrage pricing theory (APT). *Econometrica* 56: 473–476.

Richardson, S. A., R. G. Sloan, M. T. Soliman, and I. Tuna. 2005. Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics* 39: 437–485.

Roll, R., and S. A. Ross. 1980. An empirical investigation of the arbitrage pricing theory. *Journal of Finance* 35: 1073–1103.

Rosenberg, B., K. Reid, and R. Lanstein. 1985. Persuasive evidence of market inefficiency. *Journal of Portfolio Management* 11: 9–17.

Ross, S. A. 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13: 341–360.

Shanken, J. 1992. On the estimation of beta pricing models. *Review of Financial Studies* 5: 1–34.

Shanken, J., and G. Zhou. 2007. Estimating and testing beta pricing models: Alternative methods and their performance in simulations. *Journal of Financial Economics* 84: 40–86.

Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442.

Simin, T. 2008. The poor predictive performance of asset pricing models. *Journal of Financial and Quantitative Analysis* 43: 355–380.

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

Stambaugh, R. F., and Y. Yuan. 2017. Mispricing factors. *Review of Financial Studies* 30: 1270–1315.

Titman, S., K. C. J. Wei, and F. Xie. 2004. Capital investments and stock returns. *Journal of Financial and Quantitative Analysis* 39: 677–700.

White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48: 817–838.

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

Yan, X., and L. Zheng. 2017. Fundamental analysis and the cross-section of stock returns: A data-mining approach. *Review of Financial Studies* 30: 1382–1423.

Yogo, M. 2006. A consumption-based explanation of expected stock returns. *Journal of Finance* 61: 539–580.