Characteristics-driven returns in equilibrium

Guillaume Coqueret*

December 21, 2022

Abstract

We propose an equilibrium construction process of asset prices that generates returns which depend on firm characteristics, possibly in a linear fashion. One key requirement is that agents must have demands that rely separately on firm characteristics and on the log-price of assets. Market clearing via exogenous (non-factor driven) supply, combined with linear demands in characteristics, yields the sought form. The coefficients in the resulting linear expressions are scaled net aggregate demands for characteristics, as well as their variations, and both can be jointly estimated via panel regressions. Empirically, our results reveal that latent demands, which are orthogonal to characteristics, often explain a large proportion of the dispersion in average returns. Characteristics only become relevant when they survive LASSO selections that discard a large majority of their peers.

1. Introduction

A central question in financial economics pertains to why assets experience different returns. There are mainly three streams of explanations to why that may be the case. First, assets may earn contrasting returns because they have idiosyncratic exposures to various sources of risk, or factors, as put forward in Merton (1973), Ross (1976), and Fama and French (1993). A second family of explanations relies on behavioural models, in which investor preferences or beliefs drive demand towards particular stocks, thereby generating heterogeneity in the cross-section of returns (Barberis and Shleifer (2003), Barberis et al. (2015)). Another angle related to these approaches is *mispricing*, whereby agents overestimate or under-appreciate prices, often due to cognitive biases (see, e.g., Lakonishok et al. (1994), Daniel et al. (1998), Hirshleifer (2001) and Stambaugh and Yuan (2017)). Finally, a third portion of the literature, which sometimes overlaps with the first two, argues that returns differ across assets simply because these assets *are different*. It is their characteristics (their size, sector, balance sheet structure, past performance, riskiness of business lines, governance, etc.) which have an impact on their profitability (Daniel and Titman (1997)). As such, these characteristics can be used as independent variables in predictive models, as in Lynch (2001) and Gu et al. (2020), or they can be used to create more efficient risk factors (Daniel et al. (2020)).

*EMLYON Business School, 23 avenue Guy de Collongue, 69130 Ecully, FRANCE. E-mail: coqueret@em-lyon.com

Technically, factors models benefit from the simplicity of parsimony because, in practice, a handful of variables are used to explain the entirety of cross-sectional differences in returns. Assets differ in their exposure to these variables: this is the source of heterogeneity in returns. However, one major issue is the definition or identification of the factors. When they are latent (as in Chamberlain (1983), Connor and Korajczyk (1993), Bai and Ng (2002), or, more recently, in Kelly et al. (2019), and Lettau and Pelger (2020a,b) to cite but a few), they are often hard to understand - though recent advances propose enlightening interpretations (see Clarke (2021)). When they are explicit and based on prior empirical conclusions (e.g., SMB or HML in Fama and French (1993) or WML from Jegadeesh and Titman (1993)), they remain somewhat arbitrary.[1]

On the other hand, firm characteristics are directly observable, though subtleties in measurement can lead to diverging results.[2] Recently, some contributions have sought to compare the efficiency of characteristics-based versus factor-based models in asset pricing.[3] For instance, Hou et al. (2011), Goyal and Jegadeesh (2018), Chordia et al. (2019), Fama and French (2020), Raponi et al. (2020) and Chib et al. (2021)) all report that the former have a better explanatory power over stock returns. Jacobs and Levy (2021) provide some historical perspectives on the differences between the two approaches. In fact, this superiority may come from investors' demand, which is more likely to be driven by characteristics than factors (see Lawrenz et al. (2022)). On the practitioners' side, Brightman et al. (2021) find that characteristics generate more accurate predictions and, thus, are more likely to deliver out-of-sample alpha.

The first purpose of the present article is to provide a theoretical, equilibrium-based, grounding for models that use firm characteristics as predictors of future returns, exactly as in Gu et al. (2020). Indeed, as the latter put it: "*improved predictions do not tell us about economic mechanisms or equilibria*". Here, we propose a definition for a partial equilibrium[4] in which the demand of agents is tailored to yield a special form for asset (log-)returns. The sought linear expression is obtained under particular assumptions. Notably, agents are required to build their allocations (net demands) as linear functions of characteristics plus a multiple of the log-price of assets. We show that these allocations can be decomposed as linear combinations of characteristics if agents believe that expected returns are linear functions of characteristics. Loosely speaking, the model then becomes self-fulfilling because the agents' assumption materializes in equilibrium. Nonetheless, one important suggestion from the model is that returns should also depend on characteristics' change from the previous period, and not only on their levels.

[1] It is nonetheless a very interesting exercise to theoretically justify factors *a posteriori*, akin to transparent HARKing (Hollenbeck and Wright (2017)). This has generated an insightful literature, and we refer for instance to the following common anomalies:

- **size**: Berk (1995);
- **value**: Zhang (2005), Lettau and Wachter (2007);
- **size** and **value**: Gomes et al. (2003), Campbell and Vuolteenaho (2004), Arnott et al. (2015);
- **momentum**: Hong and Stein (1999), Grinblatt and Han (2005), Biais et al. (2010), Vayanos and Woolley (2013), Choi and Kim (2014).

[2] This is well documented for instance for the value factor (Asness and Frazzini (2013) and Hasler (2021)), and in sustainable investing (Dimson et al. (2020) and Avramov et al. (2021)) - not to mention firm-specific sentiment.

[3] It is interesting to note that a recent trend in the literature expects to get the best of both worlds by resorting to characteristics *within* factor models, their alphas and their loadings: see Cederburg and O'Doherty (2015), Cosemans et al. (2016), Dittmar and Lundblad (2017), Kelly et al. (2019), Connor et al. (2021), Ge et al. (2021), Kim et al. (2021a,b) and Windmüller (2022).

[4] Recently, general equilibrium models based on characteristics have emerged. We refer for example to Alti and Titman (2019), Betermier et al. (2021), Betermier et al. (2021). In Buss et al. (2021), demand depends essentially on one latent variable.

The derivation of equilibrium returns provides a novel interpretation for the loadings of predictive panels relying on firm characteristics. All estimated coefficients pertain to *scaled* demands in characteristics, and changes thereof. Therefore, the model is able to extract the relative demand in each characteristic - all without any data on flows or holdings, as in Koijen and Yogo (2019). The outcome gives two types of insights. The first is the sign of the demand, e.g., when investors are on average more "value" than "growth", so to speak. The second is the relative scale of demands across factors, i.e., when agents for instance seek exposure to size more than to momentum, or vice-versa.

Our representation for returns lies at the confluence of two representations in asset pricing. On the one hand, the cross-sectional factors in Fama and French (2020), recalling an interpretation from Fama (1976), link returns and characteristics in a simple linear fashion. However, this link holds for one time period only and is not suited to estimate asset-specific demands that are not driven by characteristics. In contrast, Koijen and Yogo (2019) propose a model which allows latent demands, but which is numerically less tractable and require holdings data. The model we propose can be viewed as a mix between these two approaches. It has a simple formulation as in cross-sectional factors, yet it is meant to include asset-level demands that cannot be captured by characteristics.

The expressions that we obtain for individual returns can be aggregated into sorted portfolio returns. This reveals the importance of characteristics' interactions in the decomposition of anomalies. The prevalence of the latter for asset pricing is also documented in Ross (2021) and Duan et al. (2021). Three terms emerge in the decomposition of long-short portfolio returns: the loading of the characteristic used for the sorting, the average fixed effect of the portfolio, and the covariance terms. Our empirical results indicate that the fixed effect terms often dominate the loadings, and the interaction components compensate this asymmetry. This is in line with the findings of Koijen and Yogo (2019), who also document the prominent impact of latent demands: "*changes in latent demand are the most important, explaining 81 percent of the cross-sectional variance of stock returns*". Nevertheless, the magnitudes of the terms are strongly dependent on the size of the rolling windows that are used for the estimations. As sizes increase, the prevalence of fixed effects decreases.

The final part of the paper raises the following, somewhat controversial, question: what is the value of firm characteristics in explaining future returns? Most of the evidence we propose suggests that standard linear models fail badly in this task, and that they capture mostly noise. This is in line with Green et al. (2017), Freyberger et al. (2020) and Huang et al. (2021), who all conclude that a large majority of characteristics fail to provide valuable predictive information. Our last batch of results suggests that investor preferences (demand estimates) shift in time and that the prevalence of characteristics in explaining future returns is very time-varying, as are factor premia (Gagliardini et al. (2016)). We find that a few dozen of characteristics suffice to explain a large portion of the cross-section of average returns, as long as the set of predictors is allowed to change from one period to another. Relevant characteristics are filtered from LASSO regressions: some are often selected (e.g., momentum), others come in and out of the set. Similarly to *pockets of predictability* (Farmer et al. (2022)), there seem to be *pockets of relevance* for characteristics, which are periods of times during which they are useful in explaining the cross-section of returns.

The remainder of the article is structured as follows. In Section 2, we present the model and some of its theoretical properties. In Section 3, we derive some implications for asset pricing anomalies. Section 4 is dedicated to the baseline empirical study while a focus on the relevance of characteristics is provided in Section 5. Finally, Section 6 concludes. Three proofs and additional material are located in the Appendix.

2. The model

2.1. Discussion on partial economic equilibria

In financial economics, the simplest market clearing equation is the following:

$$\boldsymbol{d}_t(\boldsymbol{p}_t) = \boldsymbol{s}_t(\boldsymbol{p}_t), \tag{1}$$

where the vectors of aggregate demands $\boldsymbol{d}_t$ and supplies $\boldsymbol{s}_t$ depend on the vector of prices $\boldsymbol{p}_t$. To ease the calculations, it is customary to model the demand side only. One rationale for this choice is that researchers prefer to investigate the impact of the demand side on asset prices. Often, investors are separated into heterogeneous groups. One group will craft its allocation decisions based on preferences as well as on some information set, including the price of the asset, while the other group is considered as market maker (liquidity provider, the right side of (1)) and trades independently for the price level. The equation then becomes

$$\boldsymbol{d}_t(\boldsymbol{p}_t) = \boldsymbol{s}_t \quad \Longrightarrow \quad \boldsymbol{p}_t = \boldsymbol{d}_t^{-1}(\boldsymbol{s}_t), \tag{2}$$

where the implication only holds when the inverse (multivariate) mapping $\boldsymbol{d}_t^{-1}$ is well-defined. Going into further detail, the total demand function $\boldsymbol{d}_t$ can be broken down if we consider heterogeneity in the demand of agents, in which case,

$$\sum_{i=1}^{I} A_{t,i}\boldsymbol{w}_{t,i}(\boldsymbol{p}_t) = \boldsymbol{s}_t, \tag{3}$$

where $A_{t,i}$ is the time-t wealth of agent i that is invested on the market and $\boldsymbol{w}_{t,i}$ is the corresponding relative buy or sell quantities (strictly speaking they are not necessarily portfolio compositions and we discuss this nuance later on). One favorable case is when this demand form can be factorized into:

$$\left(\sum_{i=1}^{I} A_{t,i}\boldsymbol{w}_{t,i}\right)(\boldsymbol{p}_t) = \boldsymbol{s}_t \quad \Longrightarrow \quad \boldsymbol{p}_t = \left(\sum_{i=1}^{I} A_{t,i}\boldsymbol{w}_{t,i}\right)^{-1}(\boldsymbol{s}_t), \tag{4}$$

where, again, the implication only holds if the inverse makes sense. The factorization can occur when $\boldsymbol{w}_{t,i}$ is separable, i.e., $\boldsymbol{w}_{t,i} = w_i \times \boldsymbol{w}_t(\boldsymbol{p}_t)$, or when the price-driven part is linear, that is, when $\boldsymbol{w}_{t,i} = a_{t,i,n} + b_{t,i}\boldsymbol{p}_t$. Under reasonable assumptions, the $b_{t,i}$ in the latter form is supposed to be negative, because demand usually decreases with price.[5] Such linear forms can be obtained via mean-variance preferences (see, e.g., Admati (1985) and Kacperczyk et al. (2019)).

[5] There is an ongoing debate on whether demand curves for stocks slope down. Empirical analyses from Shleifer (1986), Kaul et al. (2000), Petajisto (2009), Hau et al. (2010) and Buss et al. (2021) support this conjecture, but those in Cha and Lee (2001) and Jain et al. (2019) do not. Wurgler and Zhuravskaya (2002) conclude that it depends on whether or not the stock has substitutes on the market. In Koijen and Yogo (2019), the uniqueness of prices requires that demands be strictly downward sloping for all investors.

The main issues with most theoretical models is that they yield *prices* and not *returns*. If we want to obtain returns from Equation (4), we must tackle the following expressions (logarithmic versus arithmetic returns):

$$\boldsymbol{r}_{t+1} = \log\left(\text{diag}\left(\left(\sum_{i=1}^{I} A_{t,i}\boldsymbol{w}_{t,i}\right)^{-1}(\boldsymbol{s}_t)\right)^{-1}\left(\sum_{i=1}^{I} A_{t+1,i}\boldsymbol{w}_{t+1,i}\right)^{-1}(\boldsymbol{s}_{t+1})\right), \quad \text{or} \tag{5}$$

$$\boldsymbol{r}_{t+1} = \text{diag}\left(\left(\sum_{i=1}^{I} A_{t,i}\boldsymbol{w}_{t,i}\right)^{-1}(\boldsymbol{s}_t)\right)^{-1}\left(\sum_{i=1}^{I} A_{t+1,i}\boldsymbol{w}_{t+1,i}\right)^{-1}(\boldsymbol{s}_{t+1}) - 1, \tag{6}$$

where $\text{diag}(\boldsymbol{v})$ fills a diagonal matrix with the values of vector $\boldsymbol{v}$. The two expressions above are impractical to work with in all generality. It is therefore imperative to impose a strong structure on the agent demands $\boldsymbol{w}_{t,i}$ to obtain tractable formulae for returns. This is the purpose of the next section. Closed-form expressions are not necessary for empirical applications as long as prices or returns can easily be evaluated numerically (as is done in Koijen and Yogo (2019)). Nonetheless, they often offer insightful interpretations.

2.2. Characteristics-based demands and returns

In this subsection and henceforth, we assume agents allocate according to firms' characteristics. Because characteristics can be of diverse nature, this requirement is not very strong theoretically. Empirically, Bank et al. (2022) acknowledge that "*demand from investors is strongly affected by known stock characteristics*", which provides support for characteristics-driven demands. By construction, as we will show, this generates a characteristics-based structure for assets' log-returns.

Time is discrete and denoted with t. Investors (or agents) on the market are indexed with $i = 1\dots, I$ and they trade between $N > 1$ assets, which are indexed with $n = 1, \dots, N$. We write $p_{t,n}$ for the time-t price of asset n. In addition to their prices, all assets are characterized by exactly K indicators, $c^{(k)}_{t,n}$, for $k = 1, \dots, K$. These indicators are publicly disclosed and available to all agents on the market. Common examples for equities include market capitalization (firm size), accounting and valuation ratios, risk measures (volatility) and past performance (stock momentum).[6] Examples for bonds encompass durations or credit ratings. In the present paper, we will restrict our analysis and examples to the case of stocks, but our theoretical results hold for any asset class, as long as the price is determined by the market clearing mechanism mentioned above.

The $c^{(k)}_{t,n}$ need not be raw values, but can represent synthetic scores which are scaled in the cross-section of stocks, as is now commonplace in the literature on characteristics-based factor models (e.g., Koijen and Yogo (2019), Kelly et al. (2019) and Freyberger et al. (2020)). We write $\boldsymbol{c}_{t,n}$ for the time-t K-dimensional vector of characteristics of asset n.

One central hypothesis of the model is that the weights (or demands) $\boldsymbol{w}_t$ are unconstrained and can be negative. For instance, this can correspond to the case where market clearing operates on *net* demands. Markets and agents would be assumed to be mature so that, at each time step, the latter *adjust* their portfolio by fine-tuning pre-existing positions. Following

[6]It could be debated whether alternative metrics, like stock-specific sentiment or ESG-related data fall into this category. This discussion is outside the scope of this paper.

Koijen and Yogo (2019), we assume that these demands are driven by agent's preferences towards assets' characteristics. This is a rather weak assumption because our definition of characteristics is large and it makes sense that agents form their allocation based on some observable criteria. Reasoning with net demands instead of raw demands shares a clear proximity in spirit with models that rely on *trades* instead of *holdings*, as in van der Beck (2022).

Koijen and Yogo (2019) show that characteristics-based demands can be viewed as optimal if characteristics are informative for the evaluation of the first two moments of expected returns. Several studies document the preference of certain investors for particular characteristics (Froot and Teo (2008), Kumar (2009), Cronqvist et al. (2015), Betermier et al. (2017), Koijen et al. (2020) and Balasubramaniam et al. (2021)), so that optimality is not necessarily an imperative requirement. Characteristics, or, more generally, signals, can be thought of providing information about prices, e.g., via dividend innovations, as in Farboodi et al. (2022).

In addition to characteristics-driven investors, there exist *external* agents who trade purely orthogonally to these attributes and act as market makers in our model. Consequently, they provide a net supply for each stock, which we write $s_{t,n}$. We will not further mention these *external* agents, except via this exogenous supply which they provide.

Because we reason in terms of *net* demand, we cannot resort to an exponential function, as in Koijen and Yogo (2019), because net demands can be negative, e.g., when an agent wants to reduce a position in an asset, or sell it short. Instead, we work with the general form

$$w_{t,i,n} = a_{t,i,n} + b_{t,i}^{(0)} f(p_{t,n}) + g_{t,i}(\boldsymbol{c}_{t-1,n}), \tag{7}$$

for the time-t demand of agent i in asset n. We will study particular shapes for the functions f and $g_{t,i}$ subsequently. The above demand is expressed as a percentage of investor i's wealth, i.e., it can be considered as a portfolio composition, even though we do not impose that it sums to one across all N firms.

The rightmost part of (7) implies that the factor-driven agents construct their portfolios based on indicators which they observe or receive at time $t-1$. In practice, it is not uncommon that investors wait for quarterly updates in accounting disclosures before they rebalance their portfolios. This convention does not affect most of the results in the paper. Simply put, if characteristics' time index is lagged ($\boldsymbol{c}_{t-1}$), they are predictors. If it is synchronous ($\boldsymbol{c}_t$), then the model will explain returns but not predict them. Portfolio policies that are linear in firm characteristics are widespread in the literature (see, e.g., Brandt et al. (2009), Hjalmarsson and Manchev (2012), and Ammann et al. (2016)) and they will constitute an important special case subsequently.

The separation of the price $p_{t,n}$ from the other characteristics $\boldsymbol{c}_{t-1,n}$ in (7) is a crucial technical requirement. In any equilibrium-based asset pricing model, the demand function of at least some agents need to be expressed in terms of the price of the asset, which becomes the unknown in a market clearing equation. Solving for this unknown yields the equilibrium price. In Koijen and Yogo (2019), this is implicitly done via market equity, which is factorized into its price versus number of shares components.

The constant $a_{t,i,n}$ in Equation (7) tunes the demand of agent i towards asset n, regardless of the asset's attributes. It could for instance be driven by macro-economics factors, or private information. Technically, it could be made stock-independent, so that $a_{t,i}$ would evaluate the global equity exposure of the agent that is independent from the characteristics. From an estimation standpoint, the $a_{t,i,n}$ will allow for an interesting interpretation which we mention later on. The central component $b_{t,i}^{(0)} f(p_{t,n})$ is an important technical artefact that is used to build the equilibrium price. Let us simply assume for now that the function f is monotonic.

If we denote with $A_{t,i}$ each agent's wealth that is subject to trading at time t, then market clearing imposes that for each asset, total net demand matches total net supply, i.e.,

$$\sum_{i=1}^{I} A_{t,i} w_{t,i,n} = s_{t,n}. \tag{8}$$

The most important assumption of the model is the separation, in the demand, between the *log*-price and the other characteristics, which are unrelated to the former.[7] This implies

$$\sum_{i=1}^{I} A_{t,i} \left(a_{t,i,n} + b_{t,i}^{(0)} \log(p_{t,n}) + g_{t,i}(\boldsymbol{c}_{t-1,n}) \right) = s_{t,n}, \tag{9}$$

i.e.,

$$\log(p_{t,n}) = \frac{\overbrace{\sum_{i=1}^{I} A_{t,i} \left(a_{t,i,n} + g_{t,i}(\boldsymbol{c}_{t-1,n}) \right)}^{\text{total non-price related demand}} - \overbrace{s_{t,n}}^{\text{supply}}}{- \underbrace{\sum_{i=1}^{I} A_{t,i} b_{t,i}^{(0)}}_{\text{agg. demand for log-price}}}. \tag{10}$$

Intuitively, it seems reasonable to assume that the denominator $-\sum_{i=1}^{I} A_{t,i} b_{t,i}^{(0)}$ is positive, because we expect prices to decrease with supply. This amounts to posit that the aggregate demand for log-prices is negative, which is a reasonable postulate. This is in fact very close to the main technical assumption of Koijen and Yogo (2019). Therein, demand slopes for the logarithm of market equity are negative for all investors, while we only require the *aggregate* slope to be negative. This gives a simple formula for logarithmic returns:

$$\begin{aligned}
r_{t+1,n} &= \log\left(\frac{p_{t+1,n}}{p_{t,n}} \right) \\
&= \frac{\sum_{i=1}^{I} A_{t+1,i} \left(a_{t+1,i,n} + g_{t+1,i}(\boldsymbol{c}_{t,n}) \right) - s_{t+1,n}}{\kappa_{t+1}} - \frac{\sum_{i=1}^{I} A_{t,i} \left(a_{t,i,n} + g_{t,i}(\boldsymbol{c}_{t-1,n}) \right) - s_{t,n}}{\kappa_t} \\
&= \sum_{i=1}^{I} B_{t+1,i} \left(a_{t+1,i,n} + g_{t+1,i}(\boldsymbol{c}_{t,n}) \right) - \sum_{i=1}^{I} B_{t,i} \left(a_{t,i,n} + g_{t,i}(\boldsymbol{c}_{t-1,n}) \right) + \frac{s_{t,n}}{\kappa_t} - \frac{s_{t+1,n}}{\kappa_{t+1}} \qquad (11) \\
&= \underbrace{\underbrace{\sum_{i=1}^{I} (B_{t+1,i} a_{t+1,i,n} - B_{t,i} a_{t,i,n})}_{\substack{\text{change in scaled} \\ \text{non-characteristic demand}}} + \underbrace{\sum_{i=1}^{I} (B_{t+1,i} g_{t+1,i}(\boldsymbol{c}_{t,n}) - B_{t,i} g_{t,i}(\boldsymbol{c}_{t-1,n}))}_{\substack{\text{change in scaled} \\ \text{pure characteristic demand}}}}_{g^*(\boldsymbol{c}_{t,n}, \boldsymbol{c}_{t-1,n})} + \underbrace{\underbrace{\frac{s_{t,n}}{\kappa_t} - \frac{s_{t+1,n}}{\kappa_{t+1}}}_{\text{supply shock}}}_{e_{t+1,n}}
\end{aligned} \tag{12}$$

[7] Obviously, market capitalization or valuation ratios incorporate the price of the asset. But we assume for analytical tractability that the scores $\boldsymbol{c}_{t,n}$ are unrelated to asset prices. If characteristics are synthetic indicators which are scaled in the cross-section of stocks, this hypothesis is not too far-fetched.

where $\kappa_t = -\sum_{i=1}^{I} A_{t,i} b_{t,i}^{(0)} > 0$ is minus the aggregate demand for the log price and $B_{t,i} = A_{t,i}/\kappa_t$ are the scaled wealths. On purpose, we simplified the expression in the last line (at the root of the horizontal brackets) by using the same notations as in Equations (1) and (2) from Gu et al. (2020). This underlines that the result can be used as theoretical justification to rely on model-agnostic machine learning techniques (based on characteristics) when modelling the cross-section of asset returns.

Before we further specify the demand function, it is useful to discuss two weaknesses in the above formulation. The first one is that the decomposition only works for logarithmic returns. While they are close to their arithmetic counterparts, some difference may arise in the cross-section of mean returns, especially if assets' volatilities differ. This comes from a simple application of the Taylor series of the mapping $x \mapsto \log(1+x)$ (see Hudson and Gregoriou (2015)). The second drawback of the above expression is that it does not incorporate dividends: returns are hence *price* returns and not *total* returns. The latter are therefore out of the scope of our analysis.

2.3. The case of linear demands

An appealing special case of the demand component $g_{t,i}$ is the linear combination of characteristics. This is for instance exploited in Brandt et al. (2009), or in Koijen and Yogo (2019) when the linear form is exponentiated. The expression is simply:

$$g_{t,i}(\boldsymbol{c}_{t-1,n}) = a_{t,i,n} + \sum_{k=0}^{K} b_{t,i}^{(k)} c_{t-1,n}^{(k)}, \tag{13}$$

where the constants $b_{t,i}^{(k)}$ determine the sign and appetite intensity of agent i for characteristic k. For the sake of consistency with Equation (7), characteristic zero in the above specification is the log-price.

In Koijen and Yogo (2019), such linear forms are obtained when agents believe in a single factor model in which the first two moments of returns are estimated through firm characteristics. In Lemma 1 below, we show that linear demands can be obtained via another theoretical route. We recall that, for agent i, the expression for a mean-variance optimal portfolio has the form

$$\boldsymbol{w}_{t,i,n}^* = \gamma_{t,i}^{-1} \mathbb{V}_{t,i}[\boldsymbol{r}_{t+1}]^{-1}(\bar{\boldsymbol{r}}_t + \delta_{t,i}\mathbf{1}),$$

where $\bar{\boldsymbol{r}}_t$ is the time-t vector of expected returns, $\gamma_{t,i}$ is the time-t risk aversion of agent i , and $\delta_{t,i}$ is a scalar chosen to satisfy the budget constraint. Under particular beliefs, this expression can be factorized in a particular form, as stated below.

Lemma 1. *If agent i believes that returns are driven by*

$$\boldsymbol{r}_{t+1} = \boldsymbol{C}_t \boldsymbol{\beta}_{t+1,i} + \boldsymbol{e}_{t+1}, \tag{14}$$

then the optimal budget-constrained mean-variance portfolio weight for asset n can be written as

$$\boldsymbol{w}_{t,i,n}^* = f_{i,n,1} + \sum_{k=0}^{K} c_{t,n}^{(k)} \times f_{i,n,2}, \tag{15}$$

where $f_{i,n,1} := f_{i,n,1}(\boldsymbol{C}_t, \hat{\boldsymbol{\beta}}_{t,i}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta},i}, \hat{\boldsymbol{\sigma}}_{e,i}^2)$ and $f_{i,n,2} := f_{i,n,2}(\boldsymbol{C}_t, \hat{\boldsymbol{\beta}}_{t,i}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta},i}, \hat{\boldsymbol{\sigma}}_{e,i}^2)$ are scalars that depend on the data $\boldsymbol{C}_t$, as well on agent i's estimations for the terms in Equation (14).

The proof of the lemma is located in Appendix B. Of course, strictly speaking, the weights are not purely linear in the characteristics, because the latter are present in the definition of the functions $f_{n,1}$ and $f_{n,2}$. The only difference between (15) and (13) is the time shift in the characteristics from $c_{t,n}^{(k)}$ to $c_{t-1,n}^{(k)}$. To take into account time lags in the diffusion of the information, we henceforth stick with the latter.

Formally, it is always possible to express agent demands as a linear function of characteristics, as long as an error term is allowed, which is how Koijen and Yogo (2019) proceed. The equation $\boldsymbol{w} = \boldsymbol{C}\boldsymbol{b}$ has exactly one solution if $\boldsymbol{C}$ is square ($N = K$) and non-singular. It has an infinite number of solutions if $K > N$, and it has no solution if $K < N$, in which case $\boldsymbol{w}$ equals $\boldsymbol{C}\boldsymbol{b}$ plus an additional error term. In the specification (13), the constant $a_{t,i,n}$ can be assimilated to this error term, as it captures the demand that is not driven by the characteristics.

The demand form (13) allows to change the notation and include the $a_{t,i,n}$ terms in the sum, so that the linearized form of Equation (12) now reads

$$r_{t+1,n} = \sum_{i=1}^{I}\left(B_{t+1,i}a_{t+1,i,n} - B_{t,i}a_{t,i,n} + \sum_{k=1}^{K}\left(B_{t+1,i}b_{t+1,i}^{(k)}c_{t,n}^{(k)} - B_{t,i}b_{t,i}^{(k)}c_{t-1,n}^{(k)}\right)\right) + \varepsilon_{t+1,n}, \quad (16)$$

where

$$\varepsilon_{t+1,n} = \frac{s_{t,n}}{\eta_t} - \frac{s_{t+1,n}}{\eta_{t+1}} \quad (17)$$

is the innovation from the supply-side. We can then swap the two sums (in i and k) in the central term and, for a given k, we can decompose the central shift in two ways, depending one the factors we put forward:

$$\sum_{i=1}^{I}\left(c_{t,n}^{(k)}B_{t+1,i}b_{t+1,i}^{(k)} - c_{t-1,n}^{(k)}B_{t,i}b_{t,i}^{(k)}\right) \quad (18)$$

$$= c_{t,n}^{(k)} \underbrace{\sum_{i=1}^{I}\left(B_{t+1,i}b_{t+1,i}^{(k)} - B_{t,i}b_{t,i}^{(k)}\right)}_{\beta_{t+1}^{(k)} = \text{change in scaled agg. demand}} + \underbrace{\left(c_{t,n}^{(k)} - c_{t-1,n}^{(k)}\right)}_{\text{past change in char.}} \underbrace{\sum_{i=1}^{I}B_{t,i}b_{t,i}^{(k)}}_{\eta_t^{(k)} = \text{past demand}} \qquad \text{(first identity)} \quad (19)$$

$$= \eta_{t+1}^{(k)}(c_{t,n}^{(k)} - c_{t-1,n}^{(k)}) + c_{t-1,n}^{(k)}\beta_t^{(k)}. \qquad \text{(second identity)} \quad (20)$$

In the above expressions, we use the term "*demand*" slightly improperly. $\eta_t^{(k)}$ and $\beta_t^{(k)}$ defined in the brackets are in fact the purely characteristics driven *components* of the scaled demands. The factor $\beta_t^{(k)}$ can be viewed as the aggregate willingness to be exposed to the characteristics, while $\eta_t^{(k)}$ signals willingness to be exposed to *variations* in characteristics. However, we will henceforth resort to this abuse of language and notation and refer to these terms as demands.

Following Section 9-D in Fama (1976), Fama and French (2020) argue that if returns are linear functions of past values of firm characteristics, then the corresponding loadings (or slopes) are *returns* of long-short portfolios. We underline that this statement only holds if loadings are both time-dependent and are estimated via least square minimization.[8]

[8] Indeed, the expression $\boldsymbol{r} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e}$ has OLS coefficients $\hat{\boldsymbol{b}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{r}$, which is a linear combination of returns, and hence, a portfolio (see also Kirby (2020)). Interestingly, Stevens (1998) proves a reverse identity in which optimal portfolios can be obtained via regressions, though in this case the dependent variables are other assets' returns - see Goto and Xu (2015) and Deguest et al. (2018) for extensions of this approach. The links between regressions and portfolios also include the intuitive interpretations of Theorem 1 in Britten-Jones (1999) and in Leung and Tam (2021).

In the two identities, the slope associated to a particular characteristic is the change in scaled demand for this characteristic. Heuristically, scaled demands (*resp.*, changes in demands) can be viewed as portfolio weights (*resp.*, changes in portfolio weights). But importantly, equilibrium returns are also driven by *variations* in the characteristics as well. This seems to indicate that momentum in firm attributes should be more investigated in the literature. Several contributions tackle this topic (e.g., Ohlson and Shroff (1992), Chen (2003), Novy-Marx (2015), Blank and McLemore (2020), and, to a certain extent, Gabaix and Koijen (2021) when the characteristic is the expected dividend). Given the decompositions outlined above, we express the logarithmic returns in a simple form.

Lemma 2. *We assume that market clearing is defined in* (8) *and demands satisfy* (7) *and* (13)*. In partial equilibrium, it holds that,*

$$r_{t+1,n} = \alpha_{t+1,n} + \sum_{k=1}^{K} \left(\beta_{t+1}^{(k)} c_{t,n}^{(k)} + \eta_t^{(k)} \Delta c_{t,n}^{(k)} \right) + \varepsilon_{t+1,n}, \quad \textit{or} \tag{21}$$

$$r_{t+1,n} = \alpha_{t+1,n} + \sum_{k=1}^{K} \left(\beta_t^{(k)} c_{t-1,n}^{(k)} + \eta_{t+1}^{(k)} \Delta c_{t,n}^{(k)} \right) + \varepsilon_{t+1,n}, \tag{22}$$

where $\Delta c_{t,n}^{(k)} = c_{t,n}^{(k)} - c_{t-1,n}^{(k)}$ *is the local change in the characteristic,* $\beta_t^{(k)}$ *is the change in scaled aggregate demand for characteristic* k*, and* $\eta_t^{(k)}$ *is the scaled demand for characteristic* k *defined in Equation* (19)*. Innovations terms* $\varepsilon_{t+1,n}$ *come from the supply side and are given in* (17)*. Finally, the stock-specific constant is the change in aggregate scaled demand that is not driven by characteristics:* $\alpha_{t+1,n} = \sum_{i=1}^{I}(B_{t+1,i}a_{t+1,i,n} - B_{t,i}a_{t,i,n})$*.*

We list a few comments on this result below.

- In the main equation of the lemma, the time index for the characteristics could be shifted from t to $t+1$ and from $t-1$ to t. This corresponds to the case when agents form their portfolios based on synchronous data in Equation (7).
- In practice, the panel form of the equations will assume that the coefficients α, β and η are not time-dependent. Indeed, because of the $\alpha_{t+1,n}$, the system would be undetermined otherwise. Hence, we will use samples that are narrow chronologically so as to obtain *local* estimates. Then, the $\hat{\alpha}_n$ could be interpreted as a **fixed**, **random**, or **between effects**, depending on the modelling assumptions. Using the terminology of Koijen and Yogo (2019), it is the average **latent demand** for asset n. Strictly speaking, it is the scaled net demand that is formulated by agents for reasons orthogonal to characteristics, e.g., hedging motives.
- One implication of the Lemma is that the time-t conditional expectation of the returns is given by

$$\mathbb{E}_t[r_{t+1,n}] = \sum_{k=1}^{K} \left(\eta_t^{(k)} \Delta c_{t,n}^{(k)} + c_{t,n}^{(k)} \mathbb{E}_t \left[\beta_{t+1}^{(k)} \right] \right) + \mathbb{E}_t[\alpha_{t+1,n} + \varepsilon_{t+1,n}] \tag{23}$$

$$= \sum_{k=1}^{K} \left(\beta_t^{(k)} c_{t-1,n}^{(k)} + \Delta c_{t,n}^{(k)} \mathbb{E}_t \left[\eta_{t+1}^{(k)} \right] \right) + \mathbb{E}_t[\alpha_{t+1,n} + \varepsilon_{t+1,n}]. \tag{24}$$

  In short, and quite naturally, the conditional average returns depend on expected changes in aggregate demand and aggregate supply, plus some term that reflects aggregate preferences that are orthogonal to characteristics. These expressions hold in all generality under the assumptions stated above. One important caveat is that the model is contingent on the choice of the characteristics. It is of course impossible to list all indicators monitored and considered by all agents around the world. Implicitly, indicators that are omitted are integrated in the right part of the equations, that is, in the asset-specific term or in the supply side of the model.

- Because of the simple linear form, the *pricing power* of characteristics and their variations is directly linked to the magnitude of the corresponding demand. A characteristic c with large absolute demand $|\eta|$ will have a sizeable impact via Δc, while a change Δc with an important absolute demand $|\beta|$ will strongly drive returns through c. Reversely, characteristics with negligible demands will only play a marginal role.
- The relationship in the lemma is *predictive* because of the time lag between characteristics $\boldsymbol{c}_t$ and returns $\boldsymbol{r}_{t+1}$. If agents trade rapidly based on $\boldsymbol{c}_t$ in Equation (7), the lag vanishes and the characteristics can only *explain* returns but not forecast them.
- In a sense, Lemma 2 is self-fulfilling. If we posit that all non-supply side agents believe that returns are linear mappings of characteristics as in Lemma 1, then, loosely speaking, their assumption materializes in Lemma 2, if we omit the Δ terms.

2.4. A word on log-price demands

Logarithmic forms such as the one in the left side of Equation (9) are rare in the literature because economists usually prefer to avoid negative demands.[9] In a financial framework in which agents hold shares and adjust their portfolios, negative demands reflect a decision to reduce positions in particular assets. Likewise, if agents are allowed to sell stocks short (via negative portfolio weights), we consider that they are expressing a negative demand.

If the simple links between returns and characteristics in Lemma 2 have some empirical validity, this would imply that agents' demands can be modelled with a logarithmic form.[10]

As a function of the price, the demand is simply written

$$w = a - b \times \log(p), \tag{25}$$

where $a > 0$ is the appeal of the the object (the stock) and $b > 0$ is the slope. The corresponding elasticity is equal to $b(a - b\log(p))^{-1}$. Notably, it increases to infinity when the price shrinks to zero. While this could be an issue, we recall that most studies in asset pricing exclude penny stocks from empirical analyses.

In our model, the appeal is solely driven by the characteristics (plus a constant term). The slope is purely investor-specific. The relationship is illustrated in Figure 1. The demand can be increased by higher appeal, or lower slope, or both.

This form is very opportune for aggregation, which is one property we have largely exploited in the previous subsection. If there are I agents, each with demand $w_i = a_i - b_i \log(p)$, the latter can be summed to obtain (25), where w, a, and b are the sums of the respective individual values. In addition, the form (25), when equated to some exogenous supply s yields $p = \exp\left(\frac{a-s}{b}\right)$ which has intuitive interpretations. The price increases with a, but decreases with the slope and the supply.

[9] Though, in all generality, linear forms such as those in Admati (1985) and Kacperczyk et al. (2019) can also be subject to negative demands.

[10] Conversely, portfolio holdings from retail or institutional investors could corroborate or invalidate this form, thereby confirming or contradicting the result of the lemma. This is out of the scope of the paper.

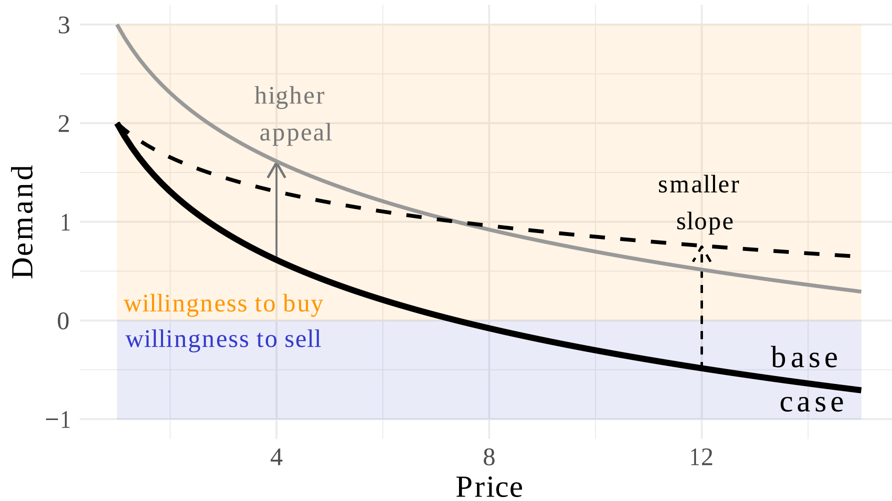

Fig. 1. **Demand curve**. This diagram shows the demand as a function of the price. The functions are equal to demand = appeal − slope × log(price). For the base case, the appeal is equal to two and the slope to one. The grey line has an appeal of three, while the dashed curve has a slope of 1/2.

3. Asset pricing anomalies

3.1. Sorted portfolios

A cornerstone result in financial economics is the capital asset pricing model (the CAPM, for which we refer to Perold (2004) for a historical perspective). The CAPM states that individual stock returns must solely be driven by their exposure to the aggregate market return. Any empirical evidence in opposition to this property can be viewed as an anomaly with respect to the baseline model. One of the workhorses to reveal anomalies is *sorting*, whereby assets are ranked according to some particular attribute and portfolios are built based on these ranks (low values versus high values of the attribute).[11] If, over a sufficiently long period, the average returns of the sorted portfolios are statistically different, it is assumed that an anomaly is uncovered (see, e.g., Cattaneo et al. (2020) for a theoretical account on statistical tests). Often, tests are performed via the null hypothesis that a long-short portfolio has a zero return, with for instance stocks with high values clustered in the long leg minus stocks with low values aggregated in the short leg. In our framework, this equally-weighted portfolio has a return equal to

$$r^{(k)}_{t+1,LS} = r_{t+1,+} - r_{t+1,-} = N_+^{-1} \sum_{n_+=1}^{N_+} r_{t+1,n_+} - N_-^{-1} \sum_{n_-=1}^{N_-} r_{t+1,n_-}, \tag{26}$$

where n_+ and n_- are the indices of the assets with high versus low value of a particular characteristic.[12] Without loss of generality, let us consider that the sought anomaly relates to the

[11]Anomalies can also be revealed by regressions and we refer to Baker et al. (2017) for a detailed account on this matter.

[12]Strictly speaking, the cross-section aggregation of *logarithmic* returns is not rigorous. Nevertheless, we proceed with this approximation. One case where it would be too coarse is if the sorting characteristic is the volatility of returns.

k^{th} characteristic, which explains the superscript in the l.h.s. of Equation (26). For simplicity, we also assume that $N_+ = N_- = N/2$ so that each leg of the portfolio consists of half of the investment universe. This is a simple and common choice (see, e.g., Barberis and Shleifer (2003)), though other fractions, such as thirds, quintiles and deciles are popular alternatives in the asset pricing literature. We then get, using the first identity of the lemma,

$$
\begin{aligned}
r^{(k)}_{t+1,LS} &= \frac{2}{N}\sum_{n_\pm}\left[\sum_{k=1}^{K}\left(\beta^{(k)}_{t+1}(c^{(k)}_{t,n_+} - c^{(k)}_{t,n_-}) + \eta^{(k)}_t(\Delta c^{(k)}_{t,n_+} - \Delta c^{(k)}_{t,n_-})\right)\right] + \Lambda^{(k)}_{t+1} + \Xi^{(k)}_{t+1}, \\
&= \sum_{k=1}^{K}\left(\beta^{(k)}_{t+1}\Psi^{(k)}_t + \eta^{(k)}_t\Phi^{(k)}_t\right) + \Lambda^{(k)}_{t+1} + \Xi^{(k)}_{t+1},
\end{aligned}
$$

with $\Psi^{(k)}_t = \frac{2}{N}\left(\sum_{n_+} c^{(k)}_{t,n_+} - \sum_{n_-} c^{(k)}_{t,n_-}\right)$ being the net portfolio average of characteristics k and $\Phi^{(k)}_t = \frac{2}{N}\left(\sum_{n_+} \Delta c^{(k)}_{t,n_+} - \sum_{n_-} \Delta c^{(k)}_{t,n_-}\right)$ the corresponding change thereof. The last terms in the expressions $\Lambda^{(k)}_{t+1} = \frac{2}{N}\left(\sum_{n_+} \alpha_{t+1,n_+} - \sum_{n_-} \alpha_{t+1,n_-}\right)$ and $\Xi^{(k)}_{t+1} = \frac{2}{N}\left(\sum_{n_+} \varepsilon_{t+1,n_+} - \sum_{n_-} \varepsilon_{t+1,n_-}\right)$ average the latent demands and the pure supply-side components of the portfolio respectively, and, for simplicity, we assume that

$$
\mathbb{E}\left[\Xi^{(k)}_{t+1}\right] = 0, \tag{27}
$$

so that the unconditional average of time-$(t+1)$ returns is

$$
\bar{r}^{(k)}_{t+1,LS} = \mathbb{E}\left[r_{t+1,LS}\right] = \mathbb{E}\left[\Lambda^{(k)}_{t+1} + \underbrace{\beta^{(k)}_{t+1}\Psi^{(k)}_t + \eta^{(k)}_t\Phi^{(k)}_t}_{\text{component driven by char. } k} + \underbrace{\sum_{j\neq k}\beta^{(j)}_{t+1}\Psi^{(j)}_t + \eta^{(j)}_t\Phi^{(j)}_t}_{\text{components driven by other chars.}}\right]. \tag{28}
$$

The hypothesis that the shifts in the supply side are null on average in Equation (27) is again a byproduct of our focusing solely on the demand side. The sources of anomalies that are not driven by characteristics come from $\mathbb{E}\left[\Lambda^{(k)}_{t+1}\right]$ only, which depends on k only through the sorting procedure. In addition, from an estimation standpoint, if we allow for fixed effects in a panel model, then the average error per asset will be zero, which implies that $\Xi^{(k)}_{t+1} = 0$ pointwise, that is, for each characteristic k and estimation sample.

We then write $\mu^{(k)}_X = \mathbb{E}[X^{(j)}_t]$, for the means of random variables, where $X = \{\Psi, \Phi, \beta, \eta\}$, and $\sigma^{(k)}_{X,Y} = \mathbb{E}[(X^{(k)}_t - \mu^{(k)}_t)(Y^{(j)}_s - \mu^{(j)}_s)]$ for the covariance terms, where chronological indices remain flexible to adapt for possible time shifts (e.g., for β and Ψ). The identity

$$
\mathbb{E}[XY] = \sigma_{X,Y} + \mu_X\mu_Y \tag{29}
$$

implies

$$
\bar{r}^{(k)}_{t+1,LS} = \mathbb{E}\left[\Lambda^{(k)}_{t+1}\right] + \sigma^{(k)}_{\beta,\Psi} + \mu^{(k)}_\beta\mu^{(k)}_\Psi + \sigma^{(k)}_{\eta,\Phi} + \mu^{(k)}_\eta\mu^{(k)}_\Phi + \sum_{j\neq k}\left(\sigma^{(j)}_{\beta,\Psi} + \mu^{(j)}_\beta\mu^{(j)}_\Psi + \sigma^{(j)}_{\eta,\Phi} + \mu^{(j)}_\eta\mu^{(j)}_\Phi\right). \tag{30}
$$

3.2. Discussion on distributions

In order to gain further intuition, we must discuss some distributional properties of the elements in Equation (28). The first major assumption that we make is

$$\mu_{\Phi}^{(j)} = \mathbb{E}\left[\Phi_t^{(j)}\right] = 0, \tag{31}$$

because there is no reason why, a priori, a sorted portfolio should see a non-zero shift in the variation of its characteristic scores. This is also confirmed in the data and results are available upon request.

The second important hypothesis we make is on the distribution of characteristics. We suppose that, at for any given time, the data has been processed such that, across the cross-section of stocks, characteristic k has a standard Gaussian distribution (with zero mean and unit variance). In recent articles in asset pricing (Kelly et al. (2019), Freyberger et al. (2020)), characteristics have uniform distributions. From that, it is easy to recover Gaussian laws by applying the inverse cumulative distribution function to the properly "*uniformized*" values of the characteristics. More generally, we posit that the cross-section of characteristics follows a multivariate Gaussian law:

$$\boldsymbol{c}_t \overset{d}{=} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t), \quad \text{with} \quad [\boldsymbol{\Sigma}_t]_{i,i} = 1 \text{ and } [\boldsymbol{\Sigma}_t]_{j,k} = \rho_t^{(j,k)}. \tag{32}$$

If the number of assets, N, is large enough, we can then approximate the aggregate portfolio score for characteristic $j = \{1, \dots, K\}$ with

$$\Psi_t^{(j)} = \frac{2}{N}\left(\sum_{n_+} c_{t,n_+}^{(j)} - \sum_{n_-} c_{t,n_-}^{(j)}\right) \underset{N\to\infty}{\longrightarrow} \mathbb{E}_{\rho_t^{(j,k)}}\left[c_t^{(j)}|c_t^{(k)} > 0\right] - \mathbb{E}_{\rho_t^{(j,k)}}\left[c_t^{(j)}|c_t^{(k)} < 0\right] = \frac{4}{\sqrt{2\pi}}\rho_t^{(j,k)}, \tag{33}$$

where the zero threshold in the conditional expectations is the median value of the characteristic (we recall that the sorting procedure operates on characteristic k). Under Gaussian assumptions, the strong law of large numbers warrants that the left part converges almost surely to $\rho_t^{(j,k)}$, which is the time-t correlation between characteristic j and characteristic k (see Appendix A for a proof). One key point is that the left-hand side of the above equation is random, thus we are obliged to envisage that the correlation term is stochastic as well. Empirically, this makes perfectly sense because correlations between characteristics are likely to vary with time.[13] Technically, this is why, in Equation (33), the expectations are conditional on this correlation value.

This is however not true for $\Psi_t^{(k)}$, which remains constant because when $j = k$ there is no interaction with other characteristics and the aggregate portfolio score is solely driven by the way the characteristics are normalized. Thus, $\Psi_t^{(k)} = 4(2\pi)^{-1/2}$ is fixed and equal to its mean.

A similar reasoning can be applied to $\Phi_t^{(j)}$, which captures the aggregate change in characteristic score of the long-short portfolio. Again, we assume that these changes are normally distributed, with zero mean, but we do not impose any covariance structure. The overarching distribution between the vectors $\Phi_t^{(j)}$ and $\Psi_t^{(j)}$ is a $2K$ Gaussian vector for which some rows

[13] In machine learning, the change in the covariance structure of predictors is referred to as *covariate shift* - see, e.g., Moreno-Torres et al. (2012).

and columns of the covariance matrix are filled with zeros. Indeed, by definition, almost surely, $\Phi_t^{(j)} = 0$ for $j = k$.

In turn, when N is large enough,

$$\Phi_t^{(j)} = \frac{2}{N}\left(\sum_{n_+} \Delta c_{t,n_+}^{(j)} - \sum_{n_-} \Delta c_{t,n_-}^{(j)}\right) \underset{N\to\infty}{\longrightarrow} \frac{4}{\sqrt{2\pi}} \rho_{t,\Delta}^{(j,k)} \mathbb{V}\left[\Delta c_t^{(j)}\right]^{1/2},$$

where $\rho_{j,k}^{\Delta}$ is the correlation between the *change* $\Delta c_{t,n}^{(j)}$ and the *level* of the characteristic $c_{t,n}^{(k)}$. The scaling constant comes from the identity in Appendix A. The bilinearity of the covariance operator implies the following result, which shows that the average return of a characteristic-sorted portfolio depends on the relationships with all other characteristics.

Lemma 3. *Assume that (27) and (31) hold and that characteristics and their variations follow centered Gaussian distributions. As the number of assets increases to infinity, we have that*

$$\bar{r}_{t+1,LS}^{(k)} \underset{N\to\infty}{\longrightarrow} \mathbb{E}\left[\Lambda_{t+1}^{(k)}\right] + \frac{4}{\sqrt{2\pi}}\left(\mu_\beta^{(k)} + \mathbb{C}\text{ov}\left(\sum_{j\neq k}\beta_{t+1}^{(j)}, \sum_{j\neq k}\rho_t^{(j,k)}\right) + \mathbb{C}\text{ov}\left(\sum_{j\neq k}\eta_t^{(j)}, \sum_{j\neq k}\rho_{t,\Delta}^{(j,k)}\mathbb{V}\left[\Delta c_t^{(j)}\right]^{1/2}\right)\right). \tag{34}$$

The lemma suggests that the average return is captured by four components:

1. A first term (pure latent demands) that depends on the characteristic k independently from the demand in this characteristic (and in any characteristic).
2. The average demand for the characteristic.
3. The covariance between *i)* summed demand shifts for all other characteristics and *ii)* the summed correlation between characteristic k and all other characteristics;
4. The covariance between *i)* total past (scaled) demands for all other characteristics and *ii)* the summed (over j) scaled correlations between shifts in characteristic j and the level of characteristic k.

If characteristics are demeaned, $\mathbb{E}[XY] + \mathbb{E}[XZ] = \mathbb{E}[X(Y + Z)]$ implies that the third component is positive when the aggregate change in demand for all other characteristics is positively correlated with the correlation between the characteristic and the sum of all other characteristics. The fourth term is positive when the demand in all other characteristics is positively related to the scaled correlation between the characteristic and the summed changes in all other characteristics.

4. Estimation

The fundamental goal of the paper is to propose simple decompositions of asset returns. Given the linear forms obtained above, we opt for simple panel models which have intuitive outputs and interpretations. A notebook with the code used to generate our baseline results and a link to the dataset is available here.

4.1. Data

The starting point for our material is the dataset used in Gu et al. (2020), updated until December 2021. Because there are large fluctuations in coverage in the early part of the sample which are likely to perturb estimations, we trim off observations prior to January 1984. This leaves 3.13 million firm-month pairs. The number of firms fluctuates between 5,471 and 9,140.

Each month, characteristics are processed to follow a standard Gaussian law, truncated beyond ± 3. This is achieved by uniformizing their distribution on [0.001, 0.999], and then applying the quantile function of the normal law. This is a technical tweak that ensures that the assumptions in Section 3.2 are satisfied. Finally, we evaluate differences in characteristics to obtain the $\Delta c_{t,n}^{(k)}$. To ease readability when plotting estimation output, we will focus on three firm characteristics: firm size, book-to-market ratio, and past performance (prior 12 month to 1 month return) as a proxy for momentum. Estimations are however performed for *all* variables.

4.2. Baseline output

The linear relationships outlined in Lemma 2 call for a panel-based estimation of scaled demands and changes in demands, which we assumed fixed over chronologically compact samples. Returns, characteristics (scores) and changes in characteristics are observed so that the $\alpha^{(k)}$, $\beta^{(k)}$ and $\eta^{(k)}$ are the unknowns. Our baseline configuration encompasses all 93 characteristics plus the 93 changes thereof as independent variables. We use rolling samples for dynamic estimation and our default window comprises two year of data, which makes at most 24 monthly points for each firm. The rationale for this choice is the following: first, we want *local* estimates. Indeed, if loadings are time-invariant they should be so on short periods. But at the same time, short samples (e.g., 6 to 12 months) yield noisy matrices,[14] because the ratio between the number of observations and the (large) number of predictors is too small. This issue will be confirmed in Section 4.3 below. Two years is thus a reasonable trade-off, which anyway will soon be relaxed. Estimations are updated every twelve months, which implies a one year overlap between consecutive samples.

In Figure 2, we plot the time-series of t-statistics pertaining to the $\hat{\beta}_t^{(k)}$ and $\hat{\eta}_t^{(k)}$ rolling estimates of Equation (21). The values when considering Equation (22) as model are qualitatively very similar. We do not produce them as they do not provide incremental added value. We show the t-statistics instead of coefficient values because they are a better indicator of the significance of the demands for characteristics. Only the three main asset pricing characteristics are depicted because it plotting 93 curves would bear little insight.

The values we report correspond to the least squares dummy variables (LSDV) estimator based on rolling samples of 24 months. We do not consider random effect models because the stock-specific constant (i.e., latent demand) in the expression of the returns is a quantity that should be estimated, just like the loadings on the other variables. Samples of 12 or 36 months yield qualitatively similar results and are available upon request.

The most salient pattern we observe is the overwhelmingly negative values associated with market capitalization, which corroborates the size effect (Banz (1981), Van Dijk (2011)). *Changes* in equity are linked to coefficients that are less significant, however. The t-statistics for book-to-market are mostly positive, but also less significant, and the figures for changes in value

[14] Because some fields are only updated quarterly or annually, estimations with fewer than 4 months of data are not even defined because of colinearity.

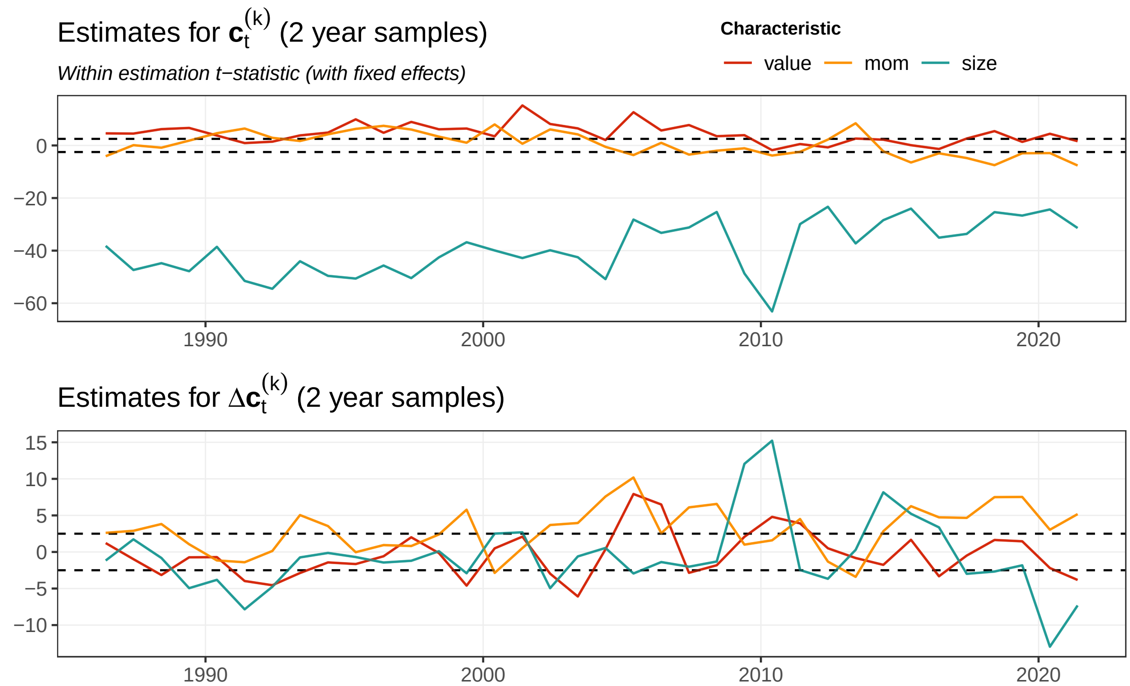

Fig. 2. **Time-series of t-statistics**. We plot the time-series of the t-statistics pertaining to $\hat{\beta}_t^{(k)}$ (upper panel) and $\hat{\eta}_t^{(k)}$ (lower panel) estimated from Equation (21) - with locally constant coefficients. Estimations are run on the prior 24 months with ***all*** 93 characteristics and changes thereof. Dashed black lines mark the ± 2.5 thresholds, which correspond to a 1% significance level.

fluctuate around zero. For the momentum anomaly, the t-statistics also oscillate, often within the non-significance boundaries. Lastly, we underline the strong difference in absolute values of t-statistics between the two panels: there is much more significance for the characteristics, and less for their variations.

4.3. Fixed effects and fit: the sample size effect

To further investigate the properties of our models and estimations, we turn to in-sample fit, as measured by the traditional R^2. In the left panel of Figure 3, we produce the time-series of the R^2 and we extend the 24 samples to shorter and longer sizes. The most salient pattern is that the R^2 decreases with sample size. This was expected, because, as the sample size increases, the ratio between the number of predictors and the number of observations decreases, which reduces the potential to fit.

A legitimate question pertains to *out-of-sample* fit. While this is somewhat out of the scope of the paper, we briefly mention some unreported results. The values we obtained are also quite time-varying but mostly negative and hence rather disappointing. This lower predictive accuracy may be due to over-fitting for instance.[15]

[15]The time-varying accuracy of dynamic forecasting models is not surprising and was recently documented in Farmer et al. (2022) for the prediction of the equity premium. In addition, negative out-of-sample R^2 are also not uncommon (see Kelly et al. (2021), Coqueret (2022), Cakici and Zaremba (2022), and the linear models of

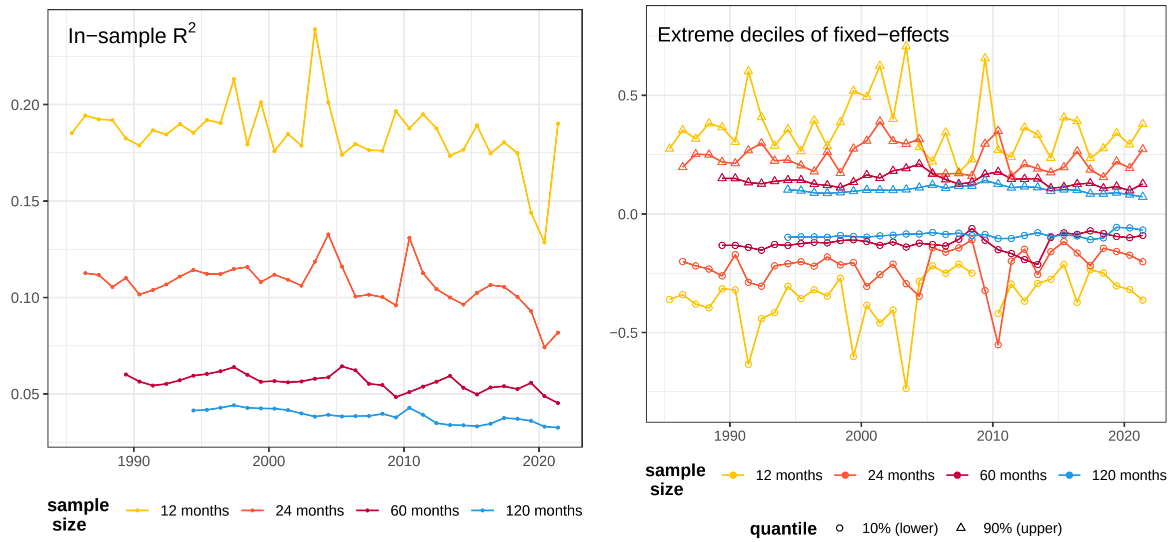

Fig. 3. **R-squared and fixed effects**. We plot the time-series of the in-sample R^2 (left) and of extreme cross-sectional quantiles of $\hat{\alpha}_{t,n}$ estimated from Equation (21) (right). Estimations are performed on rolling samples of the prior 12, 24, 60 or 120 months with all 93 characteristics and their variations.

Next, in the right panel of Figure 3, we plot the the time-series of the extreme deciles of the fixed effects ($\hat{\alpha}_{t,n}$, estimated from Equation (21)). Here again, the sample size is a clear driver of our results. As it increases, the distribution of fixed effects narrows. Fundamentally, the dispersion of fixed effects is an important indicator because it reveals if the independent variables (characteristics) are able to capture and explain the diversity in the cross-section of returns. Given the scales in the right panel of Figure 3 (± 0.5 for the extreme deciles), it appears that fixed effects, or "*latent demands*", as Koijen and Yogo (2019) call them, remain important, in absolute terms. We will further investigate this question in Section 5.

The curves in Figure 3 show that our estimates can be highly dependent on the sample size. Below, we explain why that is the case.

If we stack the coefficients into $\boldsymbol{\zeta} = (\boldsymbol{\beta}, \boldsymbol{\eta})$, then Equation (10.50) in Wooldridge (2010) implies that

$$\hat{\boldsymbol{\zeta}} = \left(\sum_{n=1}^{N} \sum_{t=1}^{T} \ddot{\boldsymbol{z}}'_{t,n} \ddot{\boldsymbol{z}}_{t,n} \right)^{-1} \left(\sum_{n=1}^{N} \sum_{t=1}^{T} \ddot{\boldsymbol{z}}'_{t,n} \ddot{r}_{t+1,n} \right), \tag{35}$$

where $\ddot{\boldsymbol{z}}_{t,n}$ is the vector of predictors that is demeaned (column-wise) at the stock level and

Gu et al. (2020)), as they often stem from exceedingly large predictions in the numerator of

$$R^2_{\text{oos}} = 1 - \frac{\sum_{n=1}^{N_t} (r_{t+1,n} - \tilde{r}_{t+1,n})^2}{\sum_{n=1}^{N_t} r_{t+1,n} - \bar{r}_{t+1,n})^2},$$

where $\bar{r}_{t+1,n}$ in the denominator is the cross-sectional average of realized returns and $\tilde{r}_{t+1,n}$ in the numerator is the prediction from the model based on estimates computed from the sample strictly prior to $t+1$. In other contexts, the disappointing performance of conditional pricing models has previously been reported in Ghysels (1998), Lewellen and Nagel (2006), and Simin (2008).

$\ddot{r}_{t+1,n}$ is the similarly demeaned future return.

According to Equation (10.58) in Wooldridge (2010), the expression for fixed-effects obtained from the LSDV estimator is

$$\hat{\alpha}_n = \bar{r}_n - \sum_{k=1}^{K} \left(\hat{\beta}^{(k)} \bar{c}_n^{(k)} + \hat{\eta}^{(k)} \bar{\Delta c}_n^{(k)} \right), \tag{36}$$

where $\hat{\beta}^{(k)}$ and $\hat{\eta}^{(k)}$ are the LSDV estimates from (35) and $\bar{r}_n$, $\bar{c}_n^{(k)}$ and $\bar{\Delta c}_n^{(k)}$ the sample average of future returns, characteristics, and their changes.

In Figure 3, the magnitude of extreme fixed effects for small sample sizes (e.g., 12 months in the right panel) is several orders larger than that pertaining to longer sample sizes (e.g., 120 months). This sizeable dispersion in fixed effects has at least two origins. First, it comes from the estimated coefficients, which correspond to the right term (sum) in Equation (36). When samples are small, the inverse in (35) is not well conditioned, which implies larger magnitudes for $\hat{\boldsymbol{\zeta}}$ (i.e., $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\eta}}$), all other things equal. Indeed, small sample sizes imply that the smallest eigenvalues of the matrix to be inverted are close to zero. Upon inversion, the related values become arbitrarily large, which generates sizeable elements in the matrix and thus in the estimates.[16] For a 12 month sample of 6,000 firms (72,000 observations), there are 6,186 columns in the original data matrix, encompassing all dummy variables plus the 186 predictors. Moreover, some characteristics are highly correlated, which generates multi-colinearity. This contributes to the bad conditioning of the first matrix in Equation (35) and hence to dispersion in coefficients.

The second source of dispersion lies in the sample averages in (36). Under an i.i.d. assumption on data generation, it is easy to show that large sample sizes decrease the variance of cross-sectional average. In Appendix C, we formally prove this statement and show in Figure 9 that multiplying the sample size five-fold (e.g., from 12 to 60 or from 24 to 120) divides the dispersion (standard deviation of mean returns) by two on average. This creates a strong dependence on the sample size for the estimation of $\Lambda_{t+1}^{(k)}$ in Equation (34). In practice however, returns are not i.i.d., and, locally, firms can experience extreme positive or negative price fluctuations, which generates a lot of heterogeneity in the cross-section. In contrast, in the long term, performance is smoothed and discrepancies between firms are less pronounced. These two sources of dispersion could of course cancel out, but, as is shown below, this will not be the case empirically.

4.4. Anomaly decomposition

Lemma 3 decomposes the average return of long-short portfolios based on characteristic sorting. There are four terms: the idiosyncratic demand term (based on fixed effects posterior to estimation), one average demand term and two covariance terms. Each one of them requires the computation of an expectation.

In Figure 4, we plot estimates of $\Lambda_{t+1}^{(k)}$ and $\beta_{t+1}^{(k)}$ for three characteristics (in sub-panels), along with average returns of rolling samples. The grey lines show the difference between the

[16] This is a very qualitative statement. Let us assume that the predictors are i.i.d. and multivariate Gaussian. While asymptotic results are well covered by the literature when the samples are large, the law of the spectrum of covariance matrices is less documented for finite samples (see Rudelson and Vershynin (2010)). The distribution of the smallest eigenvalue (which matters substantially in the conditioning of the matrix) is treated in Edelman (1991). Further results on condition numbers are obtained in Edelman (1988) and Edelman and Sutton (2005).

average returns and the sum of the two terms and thus serve as proxy for the last two terms in (34), which consist of covariances. For the size factor (middle panel), the loadings and fixed effects cancel out, meaning that the covariance terms (thick grey line) are relatively marginal. For the other two anomalies, this is not the case and fixed effects outweigh the loadings in amplitude, leaving room for the covariance interactions. In sum, the compensation of large fixed effects requires important interaction terms for the value and momentum anomalies. As we reveal in the next section, there are other compensation effects involving latent demands in the decomposition of average returns.

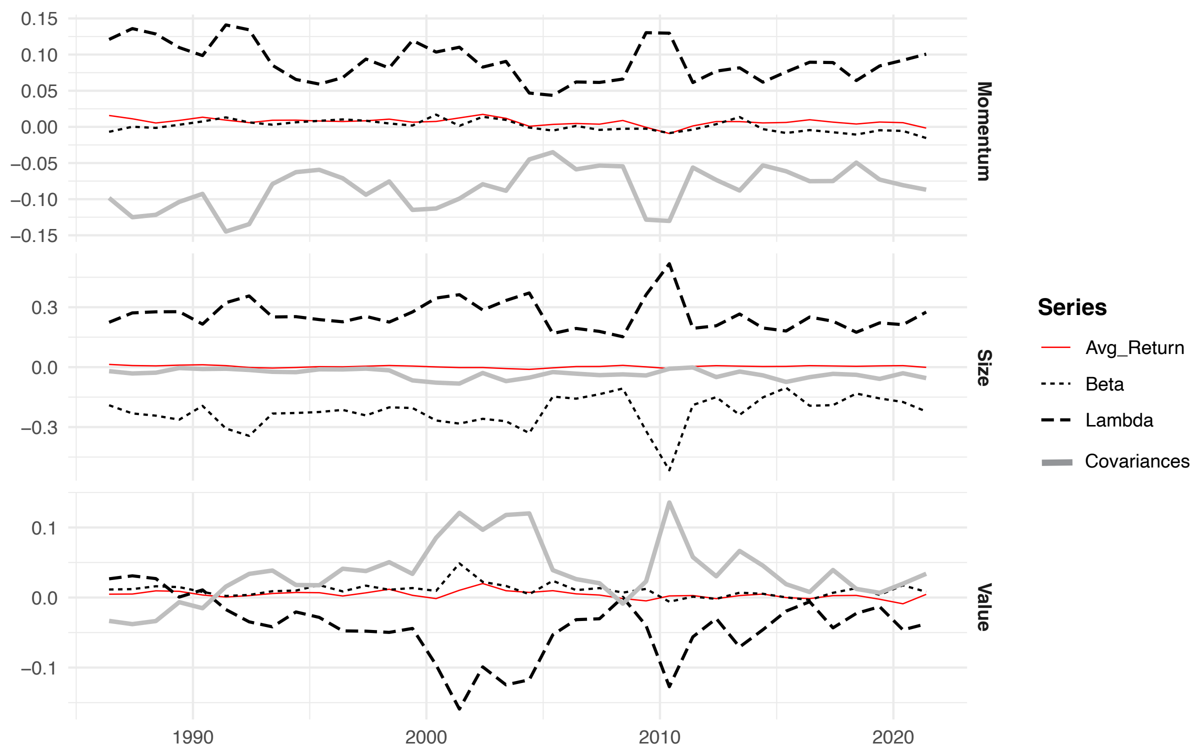

Fig. 4. **Anomaly decomposition**. We plot the time-series of the sample average of $\bar{r}^{(k)}_{t+1,LS}$ (thin red line), of sample estimates of $\hat{\beta}^{(k)}_{t+1}$ (times $4/\sqrt{\pi}$) and $\hat{\Lambda}^{(k)}_{t+1}$ (dotted lines). The thick grey curves show the average return minus the two terms (loadings and fixed effects), and approximate the sum of covariance terms in (34). Estimates rely on rolling samples of 24 months.

5. Characteristics versus latent demands

5.1. *Reframing the question*

In this section, we revert to our initial problem which asks why firms experience different returns. The evidence from the previous subsection (and Figures 3 and 4) suggests that a substantial portion of average returns can be explained with latent demands (fixed effects). This challenges the usefulness of characteristics in capturing the diversity in the cross-section of average returns.

Simply put, if we reorder Equation (36), we have that average forward returns are equal to

$$\underbrace{\bar{r}_n}_{\text{average future return}} = \underbrace{\hat{\alpha}_n}_{\text{latent demand}} + \underbrace{\bar{f}(\text{characteristics}_n)}_{\text{average char.-based value}}. \tag{37}$$

Hence, the models, which are based on characteristics, are valuable in explaining the cross-section of future returns only if the $\hat{\alpha}_n$ are small in magnitude, as the latter are the components of returns that are *not* explained by the model. In a standard panel model, if we are very conservative, we can even impose that the $\hat{\alpha}_n$ should all be statistically indistinguishable from zero, and this is exactly the specification of the null hypothesis in the so-called GRS test (Gibbons et al. (1989)) in the context of factor relevance.

This test is known to be prone to high likelihood of rejection, even in standard contexts where factors (on the r.h.s. of the equation) are synchronous with returns (on the l.h.s.). For instance, Barillas and Shanken (2018) report significant GRS statistics and note that "*models are routinely rejected at conventional levels by the GRS test*". More sophisticated approaches, e.g. the one of Ang and Kristensen (2012), also often reject the null that averaged conditional alphas are zero. In a specification where the model is predictive (i.e., there is a time lag between the l.h.s. variables and those in the r.h.s.), as in Equation (21), the odds of rejecting the null are even larger because the model cannot be expected to capture the entirety of variations in future returns.

If the $\hat{\alpha}_n$ are all jointly equal to zero, it is clear that the model (and hence the characteristics) does a great job at explaining or predicting average returns. However, the model may also be useful under much less stringent conditions. Fundamentally, we are interested in the diversity of average returns in the cross-section. This can be captured simply by taking the standard deviation of the $\bar{r}_n$, which we write σ_r. If $\sigma_r = 0$, then all average returns are the same and there is no variation in cross-sectional performance. Of course, this is never the case, and the core goal of asset pricing is to understand why asset deliver different returns. Based on Equation (37), we therefore propose a decomposition for the cross-sectional variance of average returns:

$$\begin{aligned}\sigma_r^2 = \underbrace{\mathbb{C}\text{ov}[\bar{r}_n, \bar{r}_n]}_{\substack{\text{variation in} \\ \text{mean returns}}} &= \mathbb{C}\text{ov}[\bar{r}_n, \hat{\alpha}_n + \bar{f}(\boldsymbol{x}_n)] \\ &= \underbrace{\mathbb{C}\text{ov}[\bar{r}_n, \hat{\alpha}_n]}_{\text{covariance with effects}} + \underbrace{\mathbb{C}\text{ov}[\bar{r}_n, \bar{f}(\boldsymbol{x}_n)]}_{\text{covariance with char. model}}.\end{aligned} \tag{38}$$

Based on this decomposition, we specify two types of models in Definition 4.

Definition 4. *Given the decomposition in Equation* (38)*, we say that the estimated model is*

- **characteristics-relevant (CR)** *if* $\mathbb{C}ov(\bar{r}_n, \hat{f}(\boldsymbol{x}_n)) > 0$*, and*
- **non-CR** *otherwise.*

Simply put, we consider that a model makes efficient use of characteristics if its characteristics-driven output is positively correlated with average returns. Indeed, if not, as we will see below, this means that the portion of the model that explains returns (with the correct sign) comes from fixed effects (latent demands) and is therefore unrelated to characteristics. In Figure 5, we plot the corresponding sample quantities, computed after each estimation. The element that is decomposed is the yellow line in the middle of the panels. The blue line, which depicts the covariance with the characteristics-based elements, lies below zero, meaning that the estimation is not CR. As the sample size increases (lower panels), the dispersion in the cross-section shrinks and the phenomenon is slightly less marked. However, our results unambiguously point to a limited relevance of characteristics in *linear* asset pricing. This is very consistent with the findings of Koijen and Yogo (2019), albeit with a substantially different estimation method.

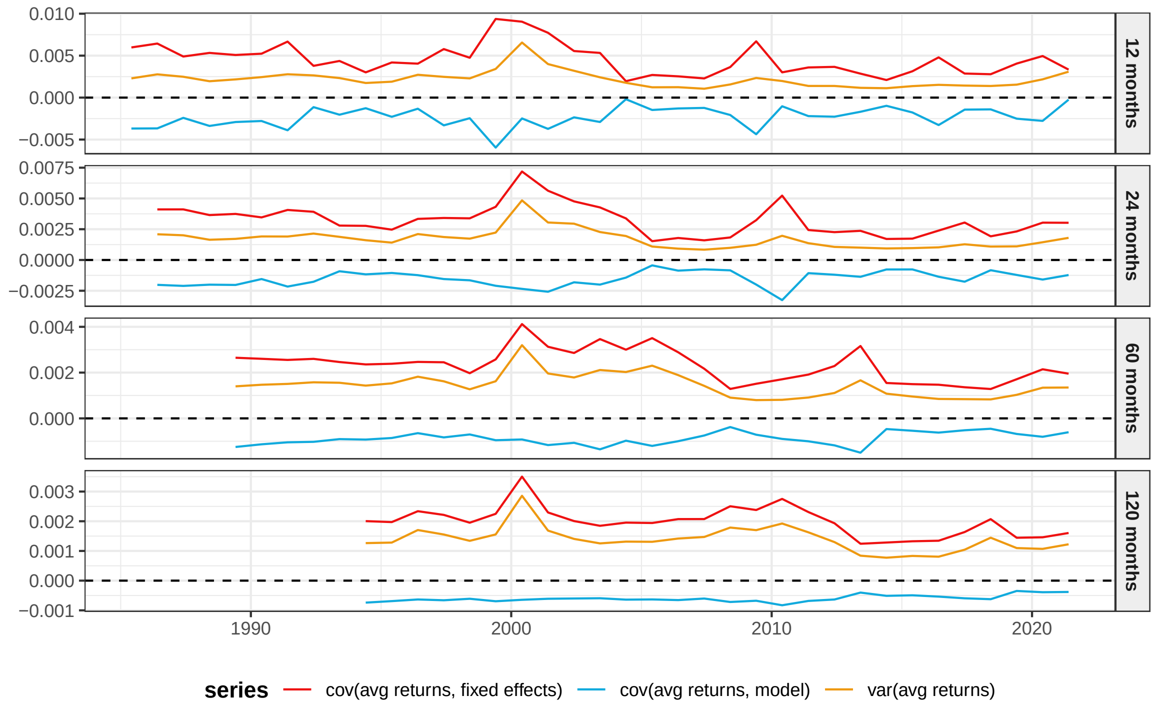

Fig. 5. **Decomposition of cross-sectional dispersion in returns**. We plot the time-series of the sample quantities in Equation (38). First, models are estimated, yielding average returns, fixed effects and predictions. Second, we compute the variance and covariance terms across all firms in the sample. Four sizes are depicted, from 12 months (upper panel) to 120 months (lower panel). For the lower panels, the curves start later because of the required estimation buffer period.

5.2. Alternative configurations

There is an ongoing debate among asset pricing researchers about the optimal number of characteristics that should be used in models. Some advocate parsimony (Fama and French (2015), Hou et al. (2015)), while others contend that large cross-sections of attributes improve the explanatory power (DeMiguel et al. (2020), Bryzgalova et al. (2021), He et al. (2021) and Han et al. (2022)). The middle ground, with 10 to 15 useful characteristics, is recommended by Green et al. (2017) and Freyberger et al. (2020).

Because of this lack of agreement, we complement our initial results by estimating coefficients for both a low-dimensional model and a large one. The low dimensional model consists in retaining only the three baseline characteristics (size (*mvel1*), value (*bm*) and momentum

($mom12m$)) for the estimation, along with their changes.

For the large model, we augment our dataset with macro-economic variables, as in Gu et al. (2020). Indeed, it is likely that macro-economic conditions influence agents' investing decisions. In the most general form of our model, in Equation (7), this would mean that the functions $g_{t,i}$ not only depend on characteristics, but also on proxies that capture the state of the economic environment. As an extension of our baseline results, we follow Gu et al. (2020) and consider that macro-economic variables have a multiplicative effect on demands. Thus, it suffices to run the original models with the predictors augmented $(M+1)$-fold, where M is the number of chosen macro proxies. We consider $M = 3$ variables, taken from the study of Welch and Goyal (2008).[17] This makes $94 \times (3+1) \times 2 = 752$ predictors in total if we count characteristics and their changes.

In Figure 6, we depict the decomposition from Equation (38). The results are very close to the second panel of Figure 5 (24 month samples), and hence, both models are, too, not CR. In particular, the scale of the peaks (≈ 0.007) are almost identical. Consequently, and surprisingly, the size of model, as proxied by the number of its independent variables, does not change the outcome, so that, again, the characteristics-based component yield mostly values that have to be compensated by the latent demands.

Fig. 6. **Variance decomposition with alternative predictor sets**. We plot the time-series of the sample quantities in Equation (38). Estimations are performed on rolling samples of the prior 24 months with 3+3 (upper panel) or 376+376 (lower panel) characteristics and their variations. In the first case, the characteristics are market capitalization (size), book-to-market (value), and momentum.

[17]They include: aggregate book-to-market ratio (bm), Treasury-bill rate (tbl), and default spread (dfy).

5.3. *Nonlinear demands*

If we go back to the general formulation of returns from Equation (12), we can simplify the expression to

$$r_{t+1,n} = a^{\Delta}_{t+1,n} + \underbrace{\sum_{i=1}^{I} B_{t+1,i}g_{t+1,i}(\boldsymbol{c}_{t,n})}_{N_1(\boldsymbol{c}_{t,n})} - \underbrace{\sum_{i=1}^{I} B_{t,i}g_{t,i}(\boldsymbol{c}_{t-1,n})}_{N_2(\boldsymbol{c}_{t-1,n})} + \epsilon_{t+1,n}, \tag{39}$$

where $a^{\Delta}_{t+1,n} = \sum_{i=1}^{I}(B_{t+1,i}a_{t+1,i,n} - B_{t,i}a_{t,i,n})$ is the stock-specific, characteristics-independent, constant. Econometrically, the above equation can be specified and estimated as the sum of two neural networks, N_1 and N_2, each with I units in their last layer (before aggregation) and each unit taking as input a possibly complex sub-network to generate nonlinearities (in g_{t+1} and g_t).[18] The constants $a^{\Delta}_{t+1,n}$ are the biases in the last layer of the overarching network. Practically, they will be estimated via a large number of dummy variables that will be added to characteristics as inputs to the global network.

Below, we test two configurations for the two neural networks N_1 and N_2:

- in the **baseline** configuration, both networks have **two** intermediate layers with 32 and 16 units, respectively. The first layer has a rectified linear unit as activation function. The second layer has a hyperbolic tangent activation in order to scale the output (see Aldridge and Avellaneda (2019)). Importantly, units in the last layer do not have internal biases, as we impose them exogenously with the dummy variables. This allows us to capture the latent demands.
- in the **sophisticated** configuration, both networks have **three** intermediate layers, with 32, 16 and 8 units. The first layer has a rectified linear unit as activation function. The second and third layers have a hyperbolic tangent activation. Again, units do not have internal biases. In addition, as in Gu et al. (2020), we add a **penalization** term for all units in the first two layers. This term is the L^1 norm of weights with a scaling constant of 0.01, the default value, and it acts as constraint on the magnitude of the layers' weights.

During training, we resort to batches of 2,000 randomly selected observations and run 100 epochs in total for each time period (i.e., training sample). All of our model parameters, which determine the architecture of the networks and how they are trained, are arguably standard and close or equal to those used by Gu et al. (2020). As in our baseline model, samples consist of 24 months of data, thus if there are 6,000 firms, this makes 144,000 observations for a given training set.

The decomposition of the cross-sectional variance in average returns is shown in Figure 7 for models based on neural networks. An interesting feature is that the scale of covariances is slightly larger in magnitude, compared to the second panel of Figure 5, implying that complexity does not increase the relevance of the characteristics-based component of the model. The pairs of curves are close, meaning that conclusions are robust to changes in the architecture of the networks.

[18] Combinations of neural networks have recently emerged in the asset pricing literature, see for example Gu et al. (2021) in a different context. More generally, nonlinearities in asset pricing are documented in Kirby (2020).

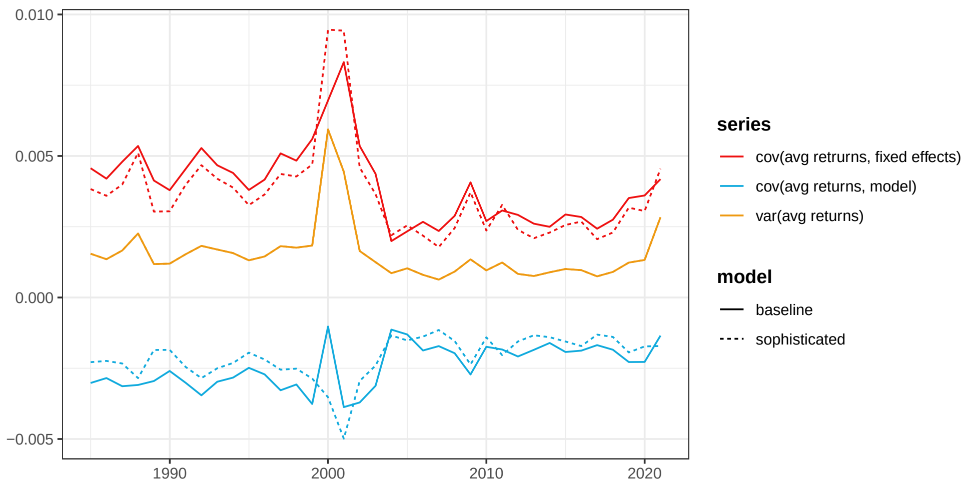

Fig. 7. **Variance decomposition with neural networks**. We plot the time-series of the sample quantities in Equation (38). Estimations are performed on rolling samples of the prior 24 months. Values for the **baseline** configuration are shown with hard lines, while those for the **sophisticated** model are represented with dotted lines.

5.4. Sparse demands

The evidence from the previous sections suggest that the characteristics-based components of estimated models are negatively correlated with average returns. One possible explanation is that characteristics carry too much noise. One representation may be that a few useful characteristics are drowned in a majority of irrelevant ones.

In light of these conclusions, we propose a last batch of models that are designed to overcome these drawbacks. The idea is basic: we seek to reduce the magnitude and number of nonzero model coefficients via LASSO regressions (Tibshirani (1996)) so as to select only the most pertinent characteristics. We thus run our baseline experiment, but with penalized regressions and in this case, we cannot use the simple form from Equation (35). The dummy variables must be generated to extract fixed effects, and we apply the LASSO estimator on the resulting data. The only, but crucial, degree of freedom is the penalization intensity (λ) that is applied to the L^1 norm of coefficients. To illustrate its impact, we will report two series of results, one for $\lambda = 0.0003$ (low penalization) and one for $\lambda = 0.001$ (high penalization).

In Figure 8, we depict the variance decomposition in the left panels, for the low penalization regime (upper plot) and high penalization regime (lower plot). In contrast to all previous models, both configurations are characteristics-relevant, as the covariance of mean returns and model terms is always above zero.

There are two main differences between the two penalization levels. First, the left panel of Figure 8 show that the dominating terms in the decomposition are not the same. With low penalization, the covariance with fixed effects (in red) lies most of the time above that of model values (in blue). However, when penalization is large, the situation reverses.

The second notable difference between the low and high penalization regimes lies in the right panels, which show the proportion of coefficients that survive the LASSO selection. Mechanically, a higher penalization implies a smaller proportion of surviving predictors. With limited

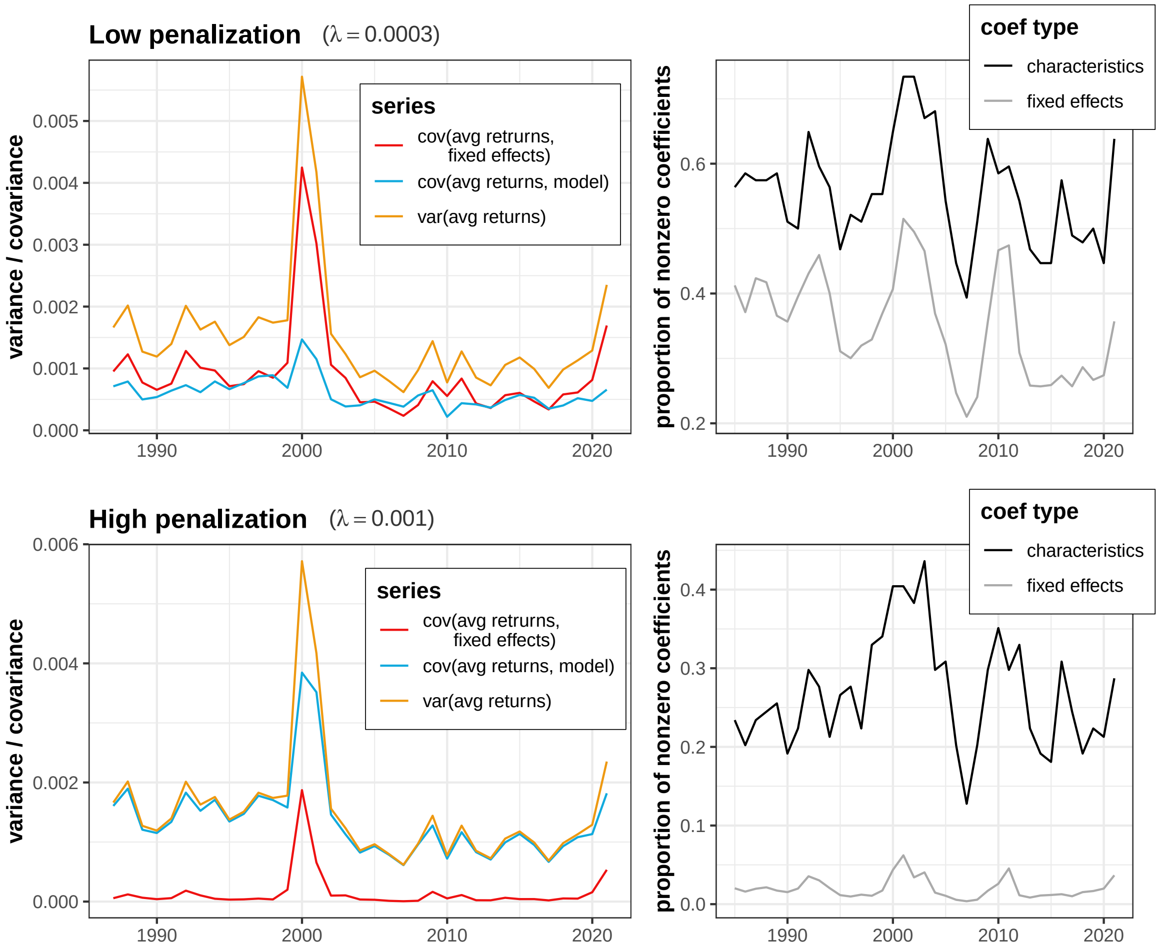

Fig. 8. **Sparse demands**. In the left panels, we plot the time-series of the sample quantities in Equation (38). The upper panel pertains to a small penalization constant ($\lambda = 0.0003$) and the upper one to a large constant ($\lambda = 0.001$). The right panels show the proportion of coefficients that have not been shrunk to zero by the penalization. Estimations are performed on rolling samples of the prior 24 months.

penalization, slightly more than half of characteristics remain, on average, while the number is only around one third for the dummy variables (fixed effects, i.e., latent demands). With regard to the high penalization regime, the numbers are logically lower: close to 25% on average for characteristics, but well below 5% for the fixed effects. In both situations, the two types in independent variables are not impacted equally: characteristics are less impacted - but they are also less numerous to begin with.

In short, the relative importance of the characteristics-driven term increases with the LASSO penalty. This means that reducing the number of predictors improves the relevance of characteristics. It is as if the cross-section of characteristics generates substantial noise, but when wisely hand-picking a reasonable number of them, this noise, as well as the influence of latent demands, vanishes. The optimal number of characteristics is neither large or small, but rather around a few dozens. Crucially, the relevant predictors must be allowed to change in time.

Variable	Average coefficient ($\times 10^3$)		Proportion of nonzero coefficient	
	high penalization	low penalization	high penalization	high penalization
mom1m	-3.474	-7.556	0.919	1.000
std_turn	1.783	4.578	0.622	0.838
turn	-1.117	-5.071	0.622	0.973
mvel1	-1.127	-4.892	0.324	0.757
rd_mve	1.556	2.771	0.459	0.784
mom36m	-0.954	-2.982	0.622	0.865
mom12m	0.213	-4.105	0.865	0.730
idiovol	-0.775	-2.454	0.568	0.649
retvol	-1.175	-1.901	0.568	0.838
mom6m	0.240	-3.270	0.270	0.838
zerotrade	-0.630	-2.058	0.216	0.514
lev	0.580	1.741	0.432	0.757

Table 1: **Most resilient characteristics**. This table lists the 12 characteristics that survive the LASSO selection the most often. The first two numerical columns gather the average coefficient (across all sampling dates), while the last two show the proportion of times that the variables have a nonzero coefficient in the model. Low (*resp.* high) penalization corresponds to $\lambda = 0.0003$ (*resp.* to $\lambda = 0.001$)

To complete our analysis, we report in Table 1 the characteristics that are the most often chosen by the LASSO. Interestingly, many of them (momentum, turnover and size (via *mvel1*) also appear at the top of the list of important variables in Gu et al. (2020). In particular, momentum-driven variables are ubiquitous.

6. Conclusion

In this article, we derive expressions for asset returns that depend on assets' characteristics and changes thereof, possibly in a linear form. The main modelling assumption is that agents have preferences that can be expressed in two separable components: the logarithm of the price of the asset, and a linear combination of its characteristics. From an estimation standpoint, returns and characteristics are given, and loadings are estimated via panel regressions on short samples and with fixed effects. The main novelty compared to standard models is that we include *changes* in characteristics. According to the equilibrium formula, the coefficients pertaining to these changes can be viewed as the scaled demands for characteristics which measure the net appetite for these characteristics.

Our contribution is theoretical in nature and proposes an interpretation for estimates of panel models in asset pricing. Our empirical results rely on rolling estimates that crucially depend on sample sizes. The latter have a sizeable impact on the magnitude of the fixed effects, i.e., latent demands. Nevertheless, these demands that cannot be explained through characteristics are prominent in most of the model specifications we propose. The models that do the most justice to traditional asset pricing characteristics are those with a selection layer whereby the LASSO dynamically removes a large portion of unnecessary characteristics (70%-80%) and latent demands (90%-95%). Just like there are *pockets of predictability*, there are pockets of relevance for characteristics. This finding contributes to the literature on the optimal number of characteristics or factors in empirical asset pricing.

Finally, a natural extension would be to generalize the model to account for an increasing set of predictors, because a framework that handles the *dynamic* inclusion of new characteristics would make more sense. The rise of alternative data based on sustainability, sentiment, now-casted earnings and macroeconomic variables, etc., paves the way to sophisticated models in which the number of features slowly increases with time - as technology progressively provides investors with myriads of characteristics.

Appendix A. Identity on a conditional expectation

Consider a bivariate Gaussian distribution (Y, Z) with zero mean, variances σ_y^2 and σ_z^2, and correlation $\rho \in [-1, 1]$. Y is the sorting variable. We are interested in

$$\frac{\mathbb{E}[Z1_{\{Y>m\}}]}{P[Y>m]} - \frac{\mathbb{E}[Z1_{\{Y<m\}}]}{P[Y<m]}.$$

If m is the median of Y (here, zero), this simplifies to

$$\begin{aligned}
2\mathbb{E}[Z(1_{\{Y>0\}} - 1_{\{Y<0\}})] &= 2\mathbb{E}[Z(1_{\{Y>0\}} - 1_{\{Y<0\}})] = 2\mathbb{E}[Z(1 - 2 \times 1_{\{Y<0\}})] = 4\mathbb{E}[Z1_{\{Y>0\}}] \\
&= 4\int_0^\infty \frac{e^{-y^2/(2\sigma_y^2(1-\rho^2))}}{2\pi\sigma_y^2\sigma_z^2\sqrt{1-\rho^2}} \left(\int_{\mathbb{R}} ze^{-(z^2-2\rho yz\sigma_z/\sigma_y)/(2\sigma_z^2(1-\rho^2))}dz\right) dy \\
&= 4\int_0^\infty \frac{e^{-y^2/(2\sigma_y^2(1-\rho^2))}}{2\pi\sigma_y\sigma_z\sqrt{1-\rho^2}} \left(\rho y\frac{\sigma_z}{\sigma_y}\sqrt{2\sigma_z^2\pi(1-\rho^2)}e^{(\rho y)^2/(2\sigma_y^2\sigma_z^2(1-\rho^2))}\right) dy \\
&= 4\rho\frac{\sigma_z}{\sigma_y^2}\int_0^\infty y\frac{e^{-y^2/(2\sigma_y^2)}}{\sqrt{2\pi}}dy = \frac{4}{\sqrt{2\pi}}\rho\sigma_z
\end{aligned}$$

where the integral result in the third line comes from Gradshteyn and Ryzhik (2007), equation 3.462-6, and the final equality is relatively standard and it is a simple case of equation 3.462-5.

Appendix B. Linear demands

Agent i believes the returns are driven by: $\boldsymbol{r}_{t+1} = \boldsymbol{C}_t\boldsymbol{\beta}_{t+1,i} + \boldsymbol{e}_{t+1}$, where the $(N \times 1)$ vector of errors is independent from all other terms and has a zero mean vector and a covariance matrix $\text{diag}(\boldsymbol{\sigma}_{e,i}^2)$, where $\boldsymbol{\sigma}_{e,i}^2$ is the vector of variances of errors. Thus,

$$\bar{\boldsymbol{r}}_t = \mathbb{E}_{t,i}[\boldsymbol{r}_{t+1}] = \boldsymbol{C}_t\mathbb{E}_{t,i}[\boldsymbol{\beta}_{t+1,i}] = \boldsymbol{C}_t\hat{\boldsymbol{\beta}}_{t,i}, \tag{40}$$

$$\mathbb{V}_{t,i}[\boldsymbol{r}_{t+1}] = \mathbb{E}_t[(\boldsymbol{r}_{t+1} - \bar{\boldsymbol{r}}_t)(\boldsymbol{r}_{t+1} - \bar{\boldsymbol{r}}_t)'] = \boldsymbol{C}_t\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta},t,i}\boldsymbol{C}_t' - \bar{\boldsymbol{r}}_t\bar{\boldsymbol{r}}_t' + \text{diag}(\hat{\boldsymbol{\sigma}}_{e,i}^2) \tag{41}$$

with $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta},t,i} = \mathbb{E}_{t,i}[\boldsymbol{\beta}_{t+1}\boldsymbol{\beta}_{t+1}']$ being agent i's time-t estimate covariance structure of the loadings. Likewise, $\hat{\boldsymbol{\beta}}_{t,i}$ is the time-t estimate of the vector of expected loadings and $\hat{\boldsymbol{\sigma}}_{e,i}^2$ the expected (or estimated) variances of errors (we omit the time index for notational simplicity).
Using the Sherman-Morrison identity, we obtain a simplified expression for the estimation of the covariance matrix:

$$\mathbb{V}_{t,i}[\boldsymbol{r}_{t+1}]^{-1} = \boldsymbol{M}^{-1}\left(\boldsymbol{I}_N + \frac{\bar{\boldsymbol{r}}_t\bar{\boldsymbol{r}}_t'\boldsymbol{M}^{-1}}{1 + \bar{\boldsymbol{r}}_t'\boldsymbol{M}^{-1}\bar{\boldsymbol{r}}_t}\right), \tag{42}$$

with $\boldsymbol{M} := \boldsymbol{M}(\boldsymbol{C}_t, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta},t,i}, \hat{\boldsymbol{\sigma}}_{e,i}^2) = \boldsymbol{C}_t\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta},t,i}\boldsymbol{C}_t' + \text{diag}(\hat{\boldsymbol{\sigma}}_{e,i}^2)$. An application of a more general Woodbury identity[19] yields

$$\boldsymbol{M}^{-1} = \left(\boldsymbol{I}_N - \text{diag}(\hat{\boldsymbol{\sigma}}_{e,i}^2)^{-1}\boldsymbol{C}_t(\boldsymbol{I}_K + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta},t,i}\boldsymbol{C}_t'\text{diag}(\hat{\boldsymbol{\sigma}}_{e,i}^2)^{-1}\boldsymbol{C}_t)^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta},t,i}\boldsymbol{C}_t'\right)\text{diag}(\hat{\boldsymbol{\sigma}}_{e,i}^2)^{-1}. \tag{43}$$

[19] Namely, $(\boldsymbol{I} + \boldsymbol{AB})^{-1} = \boldsymbol{I} - \boldsymbol{A}(\boldsymbol{I} + \boldsymbol{BA})^{-1}\boldsymbol{B}$. Our application of this identity first factors out the diagonal matrix of error variance to obtain the identity matrix.

Now, the traditional mean-variance solution is

$$\boldsymbol{w}_{t,i}^{*} = \underset{\boldsymbol{w}}{\operatorname{argmax}} \left\{ \boldsymbol{w}'\bar{\boldsymbol{r}}_t - \frac{\gamma_{t,i}}{2}\boldsymbol{w}'\mathbb{V}_{t,i}[\boldsymbol{r}_{t+1}]\boldsymbol{w},\ \text{s.t } \boldsymbol{w}'\boldsymbol{1} = b \right\} = \gamma_{t,i}^{-1}\mathbb{V}_{t,i}[\boldsymbol{r}_{t+1}]^{-1}(\bar{\boldsymbol{r}}_t + \delta_{t,i}\boldsymbol{1}),$$

where the constant $\delta_{t,i}$ is chosen to satisfy the budget constraint. From this, we infer that the optimal weights satisfy the proportionality relationship

$$\boldsymbol{w}_{t,i}^{*}(\boldsymbol{C}_t, \hat{\boldsymbol{\beta}}_{t,i}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta},t,i}, \hat{\boldsymbol{\sigma}}_{e,i}^2) = \boldsymbol{M}^{-1}\left(\boldsymbol{I}_N + \frac{\boldsymbol{C}_t\hat{\boldsymbol{\beta}}_{t,i}\hat{\boldsymbol{\beta}}_{t,i}'\boldsymbol{C}_t'\boldsymbol{M}^{-1}}{1+\hat{\boldsymbol{\beta}}_{t,i}'\boldsymbol{C}_t'\boldsymbol{M}^{-1}\boldsymbol{C}_t\hat{\boldsymbol{\beta}}_{t,i}}\right)(\boldsymbol{C}_t\hat{\boldsymbol{\beta}}_{t,i} + \delta_{t,i}\boldsymbol{1}).$$

Thus, for one asset, the form of $\boldsymbol{M}^{-1}$ given in (43) allows to write

$$w_{t,i,n} = f_{i,n,1}(\boldsymbol{C}_t, \hat{\boldsymbol{\beta}}_{t,i}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta},i}, \hat{\boldsymbol{\sigma}}_{e,i}^2) + \sum_{k=0}^{K} c_{t,n}^{(k)} \times f_{i,n,2}(\boldsymbol{C}_t, \hat{\boldsymbol{\beta}}_{t,i}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta},i}, \hat{\boldsymbol{\sigma}}_{e,i}^2),$$

with

$$f_{i,n,2} \propto -\hat{\sigma}_{e,i,n}^{-2}\Big[(\boldsymbol{I}_K + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta},t,i}\boldsymbol{C}_t'\text{diag}(\hat{\boldsymbol{\sigma}}_{e,i}^2)^{-1}\boldsymbol{C}_t)^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta},t,i}\boldsymbol{C}_t'\text{diag}(\hat{\boldsymbol{\sigma}}_{e,i}^2)^{-1} \tag{44}$$

$$\times\left(\boldsymbol{I}_N + \frac{\boldsymbol{C}_t\hat{\boldsymbol{\beta}}_{t,i}\hat{\boldsymbol{\beta}}_{t,i}'\boldsymbol{C}_t'\boldsymbol{M}^{-1}}{1+\hat{\boldsymbol{\beta}}_{t,i}'\boldsymbol{C}_t'\boldsymbol{M}^{-1}\boldsymbol{C}_t\hat{\boldsymbol{\beta}}_{t,i}}\right)(\boldsymbol{C}_t\hat{\boldsymbol{\beta}}_{t,i} + \delta_{t,i}\boldsymbol{1})\Big]_{n,\cdot} \tag{45}$$

$$f_{i,n,1} \propto -\hat{\sigma}_{e,i,n}^{-2}\left[\left(\boldsymbol{I}_N + \frac{\boldsymbol{C}_t\hat{\boldsymbol{\beta}}_{t,i}\hat{\boldsymbol{\beta}}_{t,i}'\boldsymbol{C}_t'\boldsymbol{M}^{-1}}{1+\hat{\boldsymbol{\beta}}_{t,i}'\boldsymbol{C}_t'\boldsymbol{M}^{-1}\boldsymbol{C}_t\hat{\boldsymbol{\beta}}_{t,i}}\right)\right]_{n,\cdot}(\boldsymbol{C}_t\hat{\boldsymbol{\beta}}_{t,i} + \delta_{t,i}\boldsymbol{1}) \tag{46}$$

where $[\boldsymbol{M}]_{n,\cdot}$ stands for the n^{th} row vector of matrix $\boldsymbol{M}$ and $\hat{\sigma}_{e,i,n}$ is the n^{th} element of $\hat{\boldsymbol{\sigma}}_{e,i}$.

Appendix C. Dispersion of cross-sectional sample means

We consider a variable $z_{t,n}$ observed at time t across N dimensions indexed by n (e.g., assets). We assume that the N-dimensional vector $\boldsymbol{z}_t$ is i.i.d. through time and follows some Gaussian distribution with mean vector $\boldsymbol{\mu} = (\mu_n)_{\{1\le n\le N\}}$ and covariance matrix $\boldsymbol{\Omega} = (\omega_{ij})_{\{1\le i,j\le N\}}$. The average dispersion (variance) in cross-sectional sample means[20] is equal to

$$
\begin{aligned}
C &= \mathbb{E}\left[N^{-1}\sum_{n=1}^{N}\left(\sum_{t=1}^{T}\frac{z_{t,n}}{T}-\sum_{l=1}^{N}\sum_{t=1}^{T}\frac{z_{t,l}}{NT}\right)^2\right] \\
&= \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\left[\sum_{s,t}\frac{z_{t,n}z_{s,n}}{T^2}+\sum_{l,m}\sum_{s,t}\frac{z_{t,l}z_{s,m}}{(NT)^2}-2\sum_{l=1}^{N}\sum_{s,t}\frac{z_{t,n}z_{s,l}}{NT^2}\right] \\
&= \frac{1}{N}\sum_{n=1}^{N}\left[\frac{\omega_{n,n}+T\mu_n^2}{T}+\sum_{l,m}\frac{\omega_{l,m}+T\mu_l\mu_m}{N^2T}-2\sum_{l=1}^{N}\frac{\omega_{l,n}+T\mu_l\mu_n}{NT}\right] \\
&= \frac{1}{N}\sum_{n=1}^{N}\left[\frac{\omega_{n,n}+T\mu_n^2}{T}-\sum_{l=1}^{N}\frac{\omega_{l,n}+T\mu_l\mu_n}{NT}\right] \\
&= \frac{1}{N}\sum_{n=1}^{N}\left(\mu_n^2-\frac{1}{N}\sum_{l=1}^{N}\mu_l\mu_n\right)+\frac{1}{NT}\sum_{n=1}^{N}\left(\omega_{n,n}-\frac{1}{N}\sum_{l=1}^{N}\omega_{l,n}\right)
\end{aligned}
$$

Both terms are positive for the same reasons. The second is, e.g., larger than $\frac{1}{2NT}\sum_{l,m}(\omega_{l,l}-\omega_{m,m})^2$.[21] Therefore, as the sample size increases, the dispersion in sample means decreases to the *true* variance in means.

In Figure 9, we plot the time-series of the square root of the cross-sectional dispersion defined as $N^{-1}\sum_{n=1}^{N}\left(\sum_{t=1}^{T}\frac{z_{t,n}}{T}-\sum_{l=1}^{N}\sum_{t=1}^{T}\frac{z_{t,l}}{NT}\right)^2$ when $z_{t,n}$ is the return of the stocks in the sample. We observe that multiplying the sample size by five divides the standard deviation by a factor 2-2.5 on average. This translates into a dispersion that is four times smaller for the deeper sample.

[20]When $T=1$, i.e., the dispersion is not over means but over simple returns, we refer to Grant and Satchell (2016) for theoretical results.

[21]This comes from the correlations being smaller than one in magnitude and, loosely speaking, from the identity $\sum_{i,j}(a_i^2-a_ia_j)=\frac{1}{2}\sum_{i,j}(a_i-a_j)^2$.

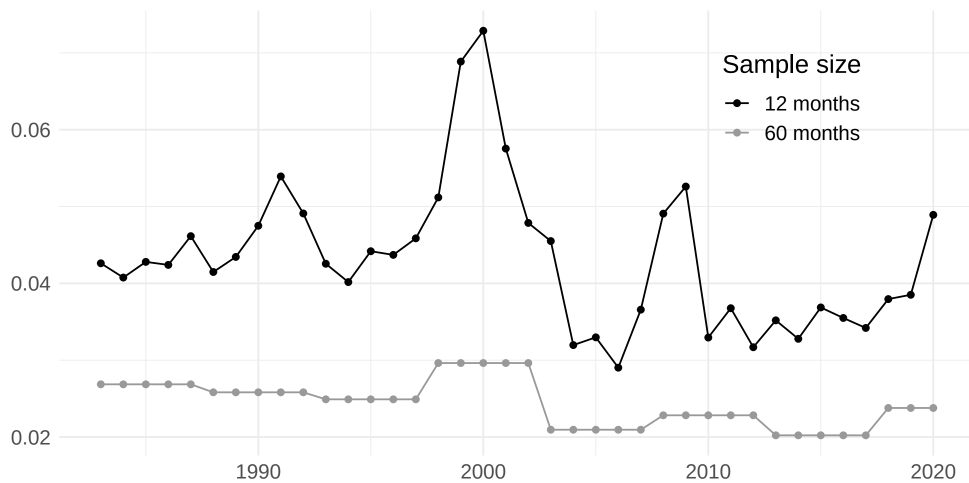

Fig. 9. **Dispersion of sample means**. We plot the time-series of the standard deviation of average returns in the cross-section. Two sample sizes (T) are tested: 12 months and 60 months. The latter remain constant for five years. The values for 2021 are omitted because they are evaluated on an incomplete sample.

References

Admati, A. R. (1985). A noisy rational expectations equilibrium for multi-asset securities markets. *Econometrica 53*(3), 629–657.

Aldridge, I. and M. Avellaneda (2019). Neural networks in finance: Design and performance. *Journal of Financial Data Science 1*(4), 39–62.

Alti, A. and S. Titman (2019). A dynamic model of characteristic-based return predictability. *Journal of Finance 74*(6), 3187–3216.

Ammann, M., G. Coqueret, and J.-P. Schade (2016). Characteristics-based portfolio choice with leverage constraints. *Journal of Banking & Finance 70*, 23–37.

Ang, A. and D. Kristensen (2012). Testing conditional factor models. *Journal of Financial Economics 106*(1), 132–156.

Arnott, R. D., J. C. Hsu, J. Liu, and H. Markowitz (2015). Can noise create the size and value effects? *Management Science 61*(11), 2569–2579.

Asness, C. and A. Frazzini (2013). The devil in HML's details. *Journal of Portfolio Management 39*(4), 49–68.

Avramov, D., S. Cheng, A. Lioui, and A. Tarelli (2021). Investment and asset pricing with ESG disagreement. *Journal of Financialk Economics Forthcoming*.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Baker, M., P. Luo, and R. Taliaferro (2017). Detecting anomalies: The relevance and power of standard asset pricing tests. Technical report, Harvard Business School Working Paper.

Balasubramaniam, V., J. Y. Campbell, T. Ramadorai, and B. Ranish (2021). Who owns what? A factor model for direct stockholding. *SSRN Working Paper 3795521*.

Bank, M., J. Lawrenz, and F. Insam (2022). Taste for characteristics or risk factor aversion? Evidence from institutional demand. *SSRN Working Paper 4070626*.

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

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer (2015). X-CAPM: An extrapolative capital asset pricing model. *Journal of Financial Economics 115*(1), 1–24.

Barberis, N. and A. Shleifer (2003). Style investing. *Journal of Financial Economics 68*(2), 161–199.

Barillas, F. and J. Shanken (2018). Comparing asset pricing models. *Journal of Finance 73*(2), 715–754.

Berk, J. B. (1995). A critique of size-related anomalies. *Review of Financial Studies 8*(2), 275–286.

Betermier, S., L. E. Calvet, and E. Jo (2021). A supply and demand approach to capital markets. *SSRN Working Paper 3440147*.

Betermier, S., L. E. Calvet, S. Knüpfer, and J. Kvaerner (2021). What do the portfolios of individual investors reveal about the cross-section of equity returns? *SSRN Working Paper 3795690*.

Betermier, S., L. E. Calvet, and P. Sodini (2017). Who are the value and growth investors? *Journal of Finance 72*(1), 5–46.

Biais, B., P. Bossaerts, and C. Spatt (2010). Equilibrium asset pricing and portfolio choice under asymmetric information. *Review of Financial Studies 23*(4), 1503–1543.

Blank, B. and C. McLemore (2020). Corporate equity performance and changes in firm characteristics. *Journal of Investment Strategies 10*(1).

Brandt, M. W., P. Santa-Clara, and R. Valkanov (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies 22*(9), 3411–3447.

Brightman, C., F. Henslee, V. Kalesnik, F. Li, and J. Linnainmaa (2021). Surprise! factor betas don't deliver factor alphas. *Research Affiliates Working Paper*.

Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *Journal of Finance 54*(2), 655–671.

Bryzgalova, S., M. Pelger, and J. Zhu (2021). Forest through the trees: Building cross-sections of stock returns. *SSRN Working Paper 3493458*.

Buss, A., R. Uppal, and G. Vilkov (2021). Dynamics of asset demands with confidence heterogeneity. *SSRN Working Paper 3928738*.

Cakici, N. and A. Zaremba (2022). Empirical asset pricing via machine learning: The global edition. *SSRN Working Paper 4028525*.

Campbell, J. Y. and T. Vuolteenaho (2004). Bad beta, good beta. *American Economic Review 94*(5), 1249–1275.

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and E. Schaumburg (2020). Characteristic-sorted portfolios: Estimation and inference. *Review of Economics and Statistics 102*(3), 531–551.

Cederburg, S. and M. S. O'Doherty (2015). Asset-pricing anomalies at the firm level. *Journal of Econometrics 186*(1), 113–128.

Cha, H.-J. and B.-S. Lee (2001). The market demand curve for common stocks: Evidence from equity mutual fund flows. *Journal of Financial and Quantitative Analysis 36*(2), 195–220.

Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models. *Econometrica 51*(5), 1305–1323.

Chen, H.-l. (2003). On characteristics momentum. *Journal of Behavioral Finance 4*(3), 137–156.

Chib, S., X. Lin, Yi Chun, K. Pukthuanthong, and X. Zeng (2021). Slope factors outperform: Evidence from a large comparative study. *SSRN Working Paper 3966807*.

Choi, S. M. and H. Kim (2014). Momentum effect as part of a market equilibrium. *Journal of Financial and Quantitative Analysis 49*(1), 107–130.

Chordia, T., A. Goyal, and J. A. Shanken (2019). Cross-sectional asset pricing with individual stocks: betas versus characteristics. *SSRN Working Paper 2549578*.

Clarke, C. (2021). The level, slope and curve factor model for stocks. *Journal of Financial Economics Forthcoming*.

Connor, G. and R. A. Korajczyk (1993). A test for the number of factors in an approximate factor model. *the Journal of Finance 48*(4), 1263–1291.

Connor, G., S. Li, and O. B. Linton (2021). A dynamic semiparametric characteristics-based model for optimal portfolio selection. *SSRN Working Paper 3803193*.

Coqueret, G. (2022). Persistence in factor-based supervised learning models. *Journal of Finance and Data Science 8*, 12–34.

Cosemans, M., R. Frehen, P. C. Schotman, and R. Bauer (2016). Estimating security betas using prior information based on firm fundamentals. *Review of Financial Studies 29*(4), 1072–1112.

Cronqvist, H., S. Siegel, and F. Yu (2015). Value versus growth investing: Why do different investors have different styles? *Journal of Financial Economics 117*(2), 333–349.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998). Investor psychology and security market under-and overreactions. *Journal of Finance 53*(6), 1839–1885.

Daniel, K., L. Mota, S. Rottke, and T. Santos (2020). The cross-section of risk and returns. *Review of Financial Studies 33*(5), 1927–1979.

Daniel, K. and S. Titman (1997). Evidence on the characteristics of cross sectional variation in stock returns. *Journal of Finance 52*(1), 1–33.

Deguest, R., L. Martellini, and V. Milhau (2018). A reinterpretation of the optimal demand for risky assets in fund separation theorems. *Management Science 64*(9), 4333–4347.

DeMiguel, V., A. Martin-Utrera, F. J. Nogales, and R. Uppal (2020). A transaction-cost perspective on the multitude of firm characteristics. *Review of Financial Studies 33*(5), 2180–2222.

Dimson, E., P. Marsh, and M. Staunton (2020). Divergent ESG ratings. *Journal of Portfolio Management 47*(1), 75–87.

Dittmar, R. F. and C. T. Lundblad (2017). Firm characteristics, consumption risk, and firm-level risk exposures. *Journal of Financial Economics 125*(2), 326–343.

Duan, Z., Z. Gong, and Q. Qi (2021). Factorization asset pricing. *SSRN Working Paper 3940074*.

Edelman, A. (1988). Eigenvalues and condition numbers of random matrices. *SIAM journal on matrix analysis and applications 9*(4), 543–560.

Edelman, A. (1991). The distribution and moments of the smallest eigenvalue of a random matrix of wishart type. *Linear algebra and its applications 159*, 55–80.

Edelman, A. and B. D. Sutton (2005). Tails of condition number distributions. *SIAM journal on matrix analysis and applications 27*(2), 547–560.

Fama, E. (1976). *Foundations of Finance - Portfolio decisions and securities prices*. Basic Books, Inc.

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

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

Farboodi, M., D. Singal, L. Veldkamp, and V. Venkateswaran (2022). Valuing financial data. *SSRN Working Paper 3947931*.

Farmer, L., L. Schmidt, and A. Timmermann (2022). Pockets of predictability. *Journal of Finance Forthcoming*.

Freyberger, J., A. Neuhierl, and M. Weber (2020). Dissecting characteristics nonparametrically. *Review of Financial Studies 33*(5), 2326–2377.

Froot, K. and M. Teo (2008). Style investing and institutional investors. *Journal of Financial and Quantitative Analysis 43*(4), 883–906.

Gabaix, X. and R. S. Koijen (2021). In search of the origins of financial fluctuations: The inelastic markets hypothesis. *SSRN Wroking Paper 3686935*.

Gagliardini, P., E. Ossola, and O. Scaillet (2016). Time-varying risk premium in large cross-sectional equity data sets. *Econometrica 84*(3), 985–1046.

Ge, S., S. Li, and O. Linton (2021). Dynamic peer groups of arbitrage characteristics. *SSRN Working Paper 3638105*.

Ghysels, E. (1998). On stable factor structures in the pricing of risk: Do time-varying betas help or hurt? *Journal of Finance 53*(2), 549–573.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989). A test of the efficiency of a given portfolio. *Econometrica 57*(5), 1121–1152.

Gomes, J., L. Kogan, and L. Zhang (2003). Equilibrium cross section of returns. *Journal of Political Economy 111*(4), 693–732.

Goto, S. and Y. Xu (2015). Improving mean variance optimization through sparse hedging restrictions. *Journal of Financial and Quantitative Analysis 50*(6), 1415–1441.

Goyal, A. and N. Jegadeesh (2018). Cross-sectional and time-series tests of return predictability: What is the difference? *Review of Financial Studies 31*(5), 1784–1824.

Gradshteyn, I. S. and I. M. Ryzhik (2007). *Table of integrals, series, and products.* Academic Press.

Grant, A. and S. Satchell (2016). Theoretical decompositions of the cross-sectional dispersion of stock returns. *Quantitative Finance 16*(2), 169–180.

Green, J., J. R. Hand, and X. F. Zhang (2017). The characteristics that provide independent information about average us monthly stock returns. *Review of Financial Studies 30*(12), 4389–4436.

Grinblatt, M. and B. Han (2005). Prospect theory, mental accounting, and momentum. *Journal of Financial Economics 78*(2), 311–339.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *Review of Financial Studies 33*(5), 2223–2273.

Gu, S., B. Kelly, and D. Xiu (2021). Autoencoder asset pricing models. *Journal of Econometrics 222*(1), 429–450.

Han, Y., A. He, D. Rapach, and G. Zhou (2022). Expected stock returns and firm characteristics: E-enet, assessment, and implications. *SSRN Working Paper 3185335*.

Hasler, M. (2021). Is the value premium smaller than we thought? *SSRN Working Paper 3886984*.

Hau, H., M. Massa, and J. Peress (2010). Do demand curves for currencies slope down? Evidence from the MSCI global index change. *Review of Financial Studies 23*(4), 1681–1717.

He, A., D. Huang, M. Yuan, and G. Zhou (2021). Tests of asset pricing models with a large number of assets. *SSRN Working Paper 3143752*.

Hirshleifer, D. (2001). Investor psychology and asset pricing. *Journal of Finance 56*(4), 1533–1597.

Hjalmarsson, E. and P. Manchev (2012). Characteristic-based mean-variance portfolio choice. *Journal of Banking & Finance 36*(5), 1392–1401.

Hollenbeck, J. R. and P. M. Wright (2017). Harking, sharking, and tharking: Making the case for post hoc analysis of scientific data. *Journal of Management 43*(1), 5–18.

Hong, H. and J. C. Stein (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance 54*(6), 2143–2184.

Hou, K., G. A. Karolyi, and B.-C. Kho (2011). What factors drive global stock returns? *Review of Financial Studies 24*(8), 2527–2574.

Hou, K., C. Xue, and L. Zhang (2015). Digesting anomalies: An investment approach. *Review of Financial Studies 28*(3), 650–705.

Huang, T., M. I. Spiegel, and H. Zhang (2021). An anatomy of characteristics in dynamic trading. *SSRN Working Paper 3896367*.

Hudson, R. S. and A. Gregoriou (2015). Calculating and comparing security returns is harder than you think: A comparison between logarithmic and simple returns. *International Review of Financial Analysis 38*, 151–162.

Jacobs, B. I. and K. N. Levy (2021). Factor modeling: The benefits of disentangling cross-sectionally for explaining stock returns. *Journal of Portfolio Management Forthcoming*.

Jain, A., P. Tantri, and R. S. Thirumalai (2019). Demand curves for stocks do not slope down: Evidence using an exogenous supply shock. *Journal of Banking & Finance 104*, 19–30.

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

Kacperczyk, M., J. Nosal, and L. Stevens (2019). Investor sophistication and capital income inequality. *Journal of Monetary Economics 107*, 18–31.

Kaul, A., V. Mehrotra, and R. Morck (2000). Demand curves for stocks do slope down: New evidence from an index weights adjustment. *Journal of Finance 55*(2), 893–912.

Kelly, B. T., S. Malamud, et al. (2021). The virtue of complexity in machine learning portfolios. *SSRN Working Paper 3984925*.

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

Kim, S., R. A. Korajczyk, and A. Neuhierl (2021a). Arbitrage portfolios. *Review of Financial Studies Forthcoming*.

Kim, S., R. A. Korajczyk, and A. Neuhierl (2021b). Characteristic-based returns: Alpha or smart beta? *Journal of Investment Management Forthcoming*.

Kirby, C. (2020). Firm characteristics, cross-sectional regression estimates, and asset pricing tests. *Review of Asset Pricing Studies 10*(2), 290–334.

Koijen, R. S., R. Richmond, and M. Yogo (2020). Which investors matter for equity valuations and expected returns? *SSRN Working Paper 3378340*.

Koijen, R. S. and M. Yogo (2019). A demand system approach to asset pricing. *Journal of Political Economy 127*(4), 1475–1515.

Kumar, A. (2009). Dynamic style preferences of individual investors and stock returns. *Journal of Financial and Quantitative Analysis 44*(3), 607–640.

Lakonishok, J., A. Shleifer, and R. W. Vishny (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance 49*(5), 1541–1578.

Lawrenz, J., M. Bank, and F. Insam (2022). Taste for characteristics or risk factor aversion? Evidence from institutional demand. *SSRN Working Paper 4070626*.

Lettau, M. and M. Pelger (2020a). Estimating latent asset-pricing factors. *Journal of Econometrics 218*(1), 1–31.

Lettau, M. and M. Pelger (2020b). Factors that fit the time series and cross-section of stock returns. *Review of Financial Studies 33*(5), 2274–2325.

Lettau, M. and J. A. Wachter (2007). Why is long-horizon equity less risky? A duration-based explanation of the value premium. *Journal of Finance 62*(1), 55–92.

Leung, R. C. and Y.-M. Tam (2021). Statistical arbitrage risk premium by machine learning. *arXiv Preprint* (2103.09987).

Lewellen, J. and S. Nagel (2006). The conditional capm does not explain asset-pricing anomalies. *Journal of Financial Economics 82*(2), 289–314.

Lynch, A. W. (2001). Portfolio choice and equity characteristics: Characterizing the hedging demands induced by return predictability. *Journal of Financial Economics 62*(1), 67–130.

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

Moreno-Torres, J. G., T. Raeder, R. Alaiz-Rodríguez, N. V. Chawla, and F. Herrera (2012). A unifying view on dataset shift in classification. *Pattern Recognition 45*(1), 521–530.

Novy-Marx, R. (2015). Fundamentally, momentum is fundamental momentum. *SSRN Working Paper 2572143*.

Ohlson, J. A. and P. K. Shroff (1992). Changes versus levels in earnings as explanatory variables for returns: Some theoretical considerations. *Journal of Accounting Research 30*(2), 210–226.

Perold, A. F. (2004). The capital asset pricing model. *Journal of Economic Perspectives 18*(3), 3–24.

Petajisto, A. (2009). Why do demand curves for stocks slope down? *Journal of Financial and Quantitative Analysis 44*(5), 1013–1044.

Raponi, V., C. Robotti, and P. Zaffaroni (2020). Testing beta-pricing models using large cross-sections. *Review of Financial Studies 33*(6), 2796–2842.

Ross, L. J. (2021). Are characteristic interactions important to the cross-section of expected returns? *SSRN Working Paper 3862847*.

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

Rudelson, M. and R. Vershynin (2010). Non-asymptotic theory of random matrices: extreme singular values. In *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures*, pp. 1576–1602.

Shleifer, A. (1986). Do demand curves for stocks slope down? *Journal of Finance 41*(3), 579–590.

Simin, T. (2008). The poor predictive performance of asset pricing models. *Journal of Financial and Quantitative Analysis 43*(2), 355–380.

Stambaugh, R. F. and Y. Yuan (2017). Mispricing factors. *Review of Financial Studies 30*(4), 1270–1315.

Stevens, G. V. (1998). On the inverse of the covariance matrix in portfolio analysis. *Journal of Finance 53*(5), 1821–1827.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological) 58*(1), 267–288.

van der Beck, P. (2022). On the estimation of demand-based asset pricing models. *SSRN Working Paper 4111329*.

Van Dijk, M. A. (2011). Is size dead? A review of the size effect in equity returns. *Journal of Banking & Finance 35*(12), 3263–3274.

Vayanos, D. and P. Woolley (2013). An institutional theory of momentum and reversal. *Review of Financial Studies 26*(5), 1087–1145.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*(4), 1455–1508.

Windmüller, S. (2022). Firm characteristics and global stock returns: A conditional asset pricing model. *Review of Asset Pricing studies 12*(2).

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Wurgler, J. and E. Zhuravskaya (2002). Does arbitrage flatten demand curves for stocks? *Journal of Business 75*(4), 583–608.

Zhang, L. (2005). The value premium. *Journal of Finance 60*(1), 67–103.