Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics

Document de Treball 2021/04 1/48 pág.
Working Paper 2021/04 1/48 pag.

“Rethinking Asset Pricing with Quantile Factor Models”

Jorge M. Uribe, Montserrat Guillen and Xenxo Vidal-Llana

UBIREA Institut de Recerca en Economia Aplicada Regional i Pública

Abstract

Traditional empirical asset pricing focuses on the average cases. We propose a new approach to analyze the cross-section of the returns. We test the predictive power of market-beta, size, book-to-market ratio, profitability, investment, momentum, and liquidity, across the whole conditional distribution of market returns. Our results indicate that the relevant characteristics to explain the winners' tail, the losers' tail and the center of the distribution, in a given period, differ. Indeed, some characteristics can be discarded from our specification if our main interest is to model expected extreme losses (such as traditional momentum), and some others should be kept even if they do not seem particularly significant for the average scenario, because they become significant at the tails (such as size). Book-to-market is mainly a left tail factor, in the sense that it explains to a greater extent the loser's tail than either the center or the tail of the winners. On the contrary, liquidity and investment are right-tail factors, because they explain in greater proportion the winners' tail than the rest of the distribution. Market beta is relevant throughout the whole cross-section, but affect winners and losers in diametrically opposite ways (the effect is positive on the right tail and negative on the left tail). We show that the practice of adding characteristics to our pricing equation should be clearly informed by our particular interests regarding the cross-sectional distribution of the returns, that is, whether we are more interested in a certain fragment of the distribution than in other parts. Our results emphasize the need to consider carefully what factors to include in the pricing equation, which depends on the dynamics that one wants to understand and even on one attitude towards risk. In short, not all factors serve all purposes.

JEL classification: G1, G11, G12, C38.

Keywords: Factor models, Asset characteristics, Tail-risks, Quantile regression.

Jorge M. Uribe: Universitat Oberta de Catalunya. Email: jorge.uribe@ub.edu

Montserrat Guillen: Universitat de Barcelona. Email: mguillen@ub.edu

Xenxo Vidal-Llana: Universitat de Barcelona. Email: juanjose.vidal@ub.edu

1. Introduction

The two fundamental questions in asset pricing are why different assets earn different rates of return and what are the systematic factors that explain movements and commonalities in time of these returns. Answering these questions requires describing and understanding the forces that determine the whole cross-sectional distribution of the returns and not only its central fragment, as is implicitly assumed by virtually all the empirical asset pricing literature. We remedy this limitation of the traditional approach which uses firms' characteristics to explain cross-sectional variations of market returns. The consequences of the limitation that we seek to overcome are considerable. For instance, if the characteristics that explain the right tail, the left tail at the center of the cross-sectional distribution of market returns differ, or if such characteristics exert a different impact on different fragments of the returns' distribution, investment recommendations based on the examination of the risk implied and the return offered by a given asset or portfolio will differ as well. In particular, incorrect specifications of the pricing equation may lead to underestimate or overestimate the risk of such an asset or portfolio, and so-called "smart-beta" strategies that are based on exploiting arbitrage opportunities based on the estimation of the price of risk due to the characteristics appearing in the pricing equation will be incorrectly informed.

Our approach expands the literature in two general ways. First, we expand the cross-sectional asset-pricing model recently revisited by Fama and French (2019) based on characteristics to the quantile-pricing case. We are able to provide quantile factor series which can be thought of as indicators of the time-varying effects of asset characteristics on the whole distribution of cross-sectional returns in the stock market. Second, we show that the particular interest of market participants in different fragments of the stock returns distribution (for instance due to different attitudes towards risk) is a first-order consideration when specifying an asset pricing equation. In short, not all factors serve all purposes. Our main pragmatic message can be summarized as follows: consider carefully what factors to include in your model of the returns, which would depend on the issue or dynamics you want to understand. For instance, if you are interested in modeling the prices of stocks with an expected combination of high-risk, high-reward (i.e. your stock/portfolio tends to be at the cross-sectional tails of the market distribution), the characteristics necessary to understand this kind of assets are different to those necessary to model assets with historical gains and losses much closer to the average scenario. The same message suits the estimation of Value at Risk statistics conditioned on pricing factors used in risk management applications by practitioners and regulators, which should consider their particular interest in the left tail of the market returns distribution.

From a different but related perspective, the heterogeneities in the relationship between characteristics and cross-sectional returns that we document call from more comprehensive and complementary ways to evaluate asset-pricing models than those currently in practice, based on 'alphas' and residuals of pricing equations focused on the average returns. Pricing errors based on conditional mean-targets, even as those considered by Barillas and Shanken (2018), Giglio and Xu (2019), Giglio et al. (2020), Chib et al. (2020) or Feng et al. (2020), are not well suited to fully evaluate the performance of an asset-pricing model in the tails of the cross-sectional distribution. Complementary statistics such as the local quantile-R^2, or the global average R^2 that weights differently the various fragments of the distribution, should be reported alongside the traditional ones, as a regular practice. Much work is needed seeking to connect the linear pricing errors literature with the factors-in-quantiles literature. Our conservative goal in this respect is to draw the attention of the profession to this important issue and to give a first step in the desired direction.

Modeling the returns using firms' characteristics has the advantage of producing interpretable factors that underlay market movements. This in turn helps us to understand market dynamics, and not only to predict it. The lack of economic insight highlighted for instance by Campbell (2018, Ch. 3), that usually accompanies factor analysis in finance, pioneered by the works of Chamberlain and Rothschild (1983) and Connor and Korajczyk (1988), is an important limitation of the statistical approach that we aim to overcome for the quantile factors case. Hence, we propose how to estimate quantile factor models based on characteristics for modeling asset prices and for risk management applications. Our proposal is an extension of the traditional cross-sectional pricing framework based on characteristics in Fama and French (2019), but instead of limiting our analysis to the average returns, our empirical model considers the effects of assets' characteristics alongside the full conditional distribution of the cross-section of these returns. In particular, our strategy allows us to test the hypothesis that traditional asset characteristics such as firms' market beta, size, book-to-market ratio, liquidity, momentum, profitability or investment, which are arguably the most popular firms' characteristics identified by the literature, explain the tails of the cross-sectional distribution of stock returns to a different extent of what they explain its center. Moreover, the characteristics underlying the dynamics of the winners (right tail) or the losers (left tail) are not necessarily the same, and their economic impact is asymmetric alongside the returns' cross-section. This fact, indeed, should not be very surprising from a theoretical point of view, since many prominent works have pointed out to a nonlinear relationship between state variables and market returns (e.g. Amihud and Mendelson 1986; Campbell and Cochrane, 1999; Bansal and Yaron, 2004; He and Krishnamurthy, 2013), yet the literature has focused almost exclusively on explaining the central part of the returns' cross-sectional distribution.

Our empirical strategy relies on conducting monthly cross-sectional conditional quantile regressions (Koenker and Bassett, 1978) using data from January 1928 to December 2019 for the universe of stocks in the US market, although we focus our main analyses on the last 20 years of the sample. Quantile regression is a powerful tool to analyze stock returns, but nevertheless its usage in empirical asset pricing has been rather limited. Quantiles are order statistics and therefore robust to outliers, which generally plague financial series. Quantile regressions do not rely on strong distributional assumptions on the error term that are inappropriate when modeling returns (e.g. multivariate Gaussian or i.i.d. cross-sectional errors). More importantly, they allow the researcher to estimate the effects of explanatory characteristics on the full conditional distribution of the explained variable. Quantile regressions have been recently used in the financial literature notably to model systemic risk in the banking sector by Adrian and Brunnermeier (2016), or in the macro-financial literature aiming to assess the predictability of economic activity by financial conditions (Adrian, et al. 2019).

Our main empirical results can be summarized as follows: i) not all traditional characteristics are relevant to explain the winners' tail, the losers' tail and average asset returns, to the same degree. Indeed, some characteristics that are relevant to explain returns close to the center of the distribution lose much of their relevance at the tails. That is the case of momentum and, to some extent, operating profitability. The opposite case is also found in our results. Namely, some characteristics, such as size, are far more relevant to explain both tails of the distribution than its center. In the same line, some characteristics are better at explaining only the winners' tail, while some others the losers' tail. We include investment and liquidity in the former group and the book-to-market ratio in the latter. ii) Some characteristics, such as the market beta or size, show diametrically opposite effects on both tails of the market returns distribution, to the point that the sign of the effects changes from negative to positive for the market beta, and from positive to negative for size. iii) The joint ability of the popular characteristics considered in our model varies significantly in time and also across quantiles. It seems easier to explain the tails of the distribution (average R^2=0.10 for both tails), than the center of the distribution (average R^2=0.03). Regarding the time-variation of the joint explanatory power, which is notorious (max R^2=0.26 and min R^2=0.01 for the losers' tail and max R^2=0.35 and min R^2=0.01 for the tail of the winners), we conjecture that the set of factors considered in our specification gain economic relevance under certain macro-economic states. iv) We show that successfully selecting characteristics for our asset pricing equation clearly depends on what part of the distribution of the market returns we are particularly interested in. For instance, if we care more about the right side of the distribution, we would add to our baseline model the *enterprise multiple* proposed by Loughran and Wellman (2011), while if our interest is in the left side of the distribution we would add the *asset growth* variable of Cooper et al. (2008).

The rest of this document is organized as follows. Section two is a brief revision of the related literature that puts our contribution in perspective. Section three consists of the methodology. In the fourth section we describe our data and sources, which (except for the returns) is publicly available thanks to Chen and Zimmermann (2020), via the *Open source for asset pricing* project. In the fifth section we present our main results and provide intuition for our findings. This section also includes robustness exercises, mainly consisting of changing the working sample, thus conditioning our results on different macroeconomic states, and taking a more historical perspective for our estimations. The sixth section concludes and highlights some future lines of research for the topic of pricing the tails of winners and losers in the stock market.

2. Related Literature and Contribution

Our work is mainly related to the model studied by Fama and French (2019). These authors have emphasized the importance of considering cross-sectional variation of characteristics (as opposed to time-series variation), when explaining market returns. This line of research goes back at least to Fama and McBeth (1973) and emphasizes the time-varying factor loads of economic factors, which are better encapsulated in an asset pricing model by conducting cross-sectional regressions each period, and stacking the results, than by building 'mimic' portfolios according to a characteristic-sorted criterion and then conducting time-series regressions, as it has been done for instance by the same authors in previous opportunities (Fama and French, 1993, 2015). The former strategy works especially well in terms of reducing the pricing errors for individual stock returns (instead of portfolios), which is our unit of interest.

From a statistical point of view our work is related to previous methodological studies that have proposed panel-quantile or other non-linear panel structures to model asset returns or other sets of economic series with time-series and cross-sectional dimensions. For instance, Amengual and Sentana (2018) reject the hypothesis of a Gaussian copula describing the cross-sectional dependence among monthly returns on individual US stocks. Indeed, they find strong nonlinear tail dependences, co-skewness and co-kurtosis among the analyzed returns. Ando and Bai (2020) report that common statistical factor structures explaining asset returns distribution in global financial markets depict divergent patterns on the lower an upper tails of the distribution, after the Global Financial Crisis. Chen et al. (2021) report that the number of statistical quantile-factors is lower at the tails of the returns distribution than at the center. The literature of statistical quantile-factor models, from our perspective, provides the most convincing corpus of evidence of a changing number of statistical factors or a changing influence of such factors across the conditional distribution of the returns, although it offers no clear evidence on what the factors are. In this line, Belloni et al. (2019) and Ma et al. (2020) go one step forward, as they propose panel quantile models that use latent statistical factors as building blocks and other covariates as a complement. They test their models using stock market returns, as a case of study, and document important differences across quantiles of pricing covariates and factors, with a monthly and daily frequency respectively.

Although it is a step in the right direction, the quantile-factors reported by these latter authors also lack economic interpretability by themselves, because they serve the related purpose of inducing time-varying factors loads through which characteristics *jointly* impact the returns, hence such factors are still purely statistical in nature. In other words, the factors remain as a combination at the quantile level of all the information within the sample, which, similar to traditional PCAs, do not offer an economic insight. Another important difference of our model with respect to the aforementioned literature is that this literature needs to resort to a balanced panel structure to estimate their models. In turn, this forces the researcher to retain only the firms present in the market during the whole sample period, which can easily render survivorship bias to the estimates. For instance, only around 23% of firms that were listed in the market in the year 2000, were present at the end of 2019. Such an approach is unsuitable to analyze the stock market from a more historical perspective, as we do in section 5.4 of the results, simply because too few firms were listed in the market during the whole sample period as to construct a representative panel from 1928 to 2019, indeed only 38 in our dataset. Our approach is flexible enough to overcome this problem. Moreover, even if the panels in the previous literature were unbalanced, the strategy followed by these authors keeps the researcher from analyzing the time-varying nature of the documented relationships at the quantile level, which as we will see, is key to understand the heterogeneous flows of information from asset characteristics to excess returns, and to evaluate the robustness of the documented quantile-effects in time. Finally, none of the studies above examine and quantify the importance of considering quantile heterogeneities when we are setting our asset pricing equation (see subsection 5.6), which is one of our two main contributions.

Another line of studies related to ours, develops different frameworks seeking to incorporate asset characteristics to inform latent factor estimation and to offer the possibility of a nonlinear relationship between explanatory factors and asset returns. For instance, Connor et al. (2012), Feng et al. (2019), Lettau et al. (2020), Kelly et al. (2019), Kelly et al. (2020) all seek to estimate latent factors and factor-loadings which are theoretically informed by covariates. Gu et al. (2020 a,b) and Chen et al. (2020), pursue the same objective but they use Machine Learning tools such as (deep) neural networks and decision trees to synthetize the rich information contained in the characteristics dimension. Deep neural networks, in particular Autoencoders, can be understood as a generalization of PCA to a nonlinear space (see for instance Feng et al. (2018) and Andreini et al. (2020)). This line of studies seeks to encapsulate nonlinearities in the relationship between returns and characteristics that otherwise could not be incorporated into the model, because they are not reducible to a linear approximation via some transformation of the original variables (or some double or triple sorting conditional on characteristics). This latter set of studies is totally silent about our main concern. Namely, we care about the whole conditional distribution of the cross-section of the returns, and not only the average. In other words, our focus is on the whole cross-sectional distribution of the returns (i.e., left tail, right tail, center) and not on the shape of the relationship between characteristics and the average cross-sectional returns.

3. Methodology

We follow the notation in Fama and French (2019). In the following model, returns in month *t*, R_{it}, $i = 1, \dots, n_t$, are explained by past values of size (MC_{it-1}), book-to-market ratio (BM_{it-1}), operating profitability (OP_{it-1}), and the rate of growth of assets (INV_{it-1}):

$$R_{it} = R_{z_t} + R_{MC_t} MC_{it-1} + R_{BM_t} BM_{it-1} + R_{OP_t} OP_{it-1} + R_{INV_t} INV_{it-1} + e_{it}, \qquad (1)$$

where R_{z_t} is the intercept of the cross-section regression in month *t*. and n_t the number of different companies with information in month *t*. As emphasized by Fama and French (2019) the slope estimates in Equation 1 (i.e. R_{MC_t}, R_{BM_t}, R_{OP_t} and R_{INV_t}) are portfolio returns that can be interpreted as factors once they have been stacked. The intercept is the month *t* return common to all assets, which is not captured by the explanatory variables in the regression (see as well Fama and McBeth (1973) and Fama (1976)). e_{it} is a residual term.

A small rearrangement of terms in Equation 1 shows that, indeed, stacking the slopes of the cross-sectional regressions for all months in the sample, Equation 1 can be interpreted as a factor-asset-pricing model with time varying factor loadings given by firms' characteristics, in the following way:

$$R_{it} - R_{z_t} = MC_{it-1}R_{MC_t} + BM_{it-1}R_{BM_t} + OP_{it-1}R_{OP_t} + INV_{it-1}R_{INV_t} + e_{it}. \quad (2)$$

We have interchanged the order of the characteristics and the time-varying slopes, in order to emphasize the common factor structure. In both versions of the model given by equations 1 or 2 the relationship is linear. That is, it consists of a mean-to-mean mapping that measures average movements of the Right Hand Side (RHS) variables on average responses of the Left Hand Side (LHS) returns.

We expand the model in equations 1 and 2 in order to explain the whole conditional distribution of returns R_{it} in a given month, which we propose to summarize via an interpretable quantile-factor structure. To this end, we generalize Equation 1, in the following way:

$$Q^{\theta}_{R_{it}} = R^{\theta}_{z_t} + R^{\theta}_{MC_t}MC_{it-1} + R^{\theta}_{BM_t}BM_{it-1} + R^{\theta}_{OP_t}OP_{it-1} + R^{\theta}_{INV_t}INV_{it-1}, \quad (3)$$

where $Q^{\theta}_{R_{it}}$ is the θth conditional quantile of R_{it} given asset characteristics on the RHS of the equation. We can alternatively present the model using matrices notation as $Q^{\theta}_{R_{it}} = R^{\theta}_{Xt}\, X'_{it-1}$, where $X_{it} = \{MC_{it}, BM_{it}, OP_{it}, INV_{it}\}$, such that $Q^{\theta}_{R_{it}} = \inf\{r_t{:}\, F_i(r|X'_{it-1}) \geq \theta\}$ where $F_i(\cdot)$ is the cumulative distribution function. Note that Equation 3 does not contain a random term because $Q^{\theta}_{R_{it}}$ characterizes R_{it} but it is deterministic in nature. Note that the slopes have superscripts because they change for each value of θ. So, going back to Equation 3 we have that R^{θ}_{Xt}:

$$R^{\theta}_{Xt} = \text{argmin}_{\beta} E\left[\rho_{\theta}\left(R_{it} - R^{\theta}_{z_t} - \beta\, X'_{it-1}\right)\right], \quad (4)$$

where $\rho_{\theta}(\cdot)$ is a loss function, given by $\rho_{\theta}(e) = (1-\theta)I_{\{e<0\}}|\text{e}| + \theta I_{\{e>0\}}|\text{e}|$, where $I_{\{e<0\}}$ is an indicator function that is equal to 1 when the subscript is true and 0 otherwise (see Uribe and Guillen (2020. Ch. 3)). As it is well known, the mathematical formulation in Equation 4 leads to the solution of a linear programming optimization problem. Its basic structure and the counterpart algorithm solution can be found in Koenker (2005), and are omitted here.

To estimate the model in equations 3-4 we follow the most popular specification for the estimation of conditional quantiles due to Koenker and Basset (1978), which has been also followed by the recent literature in statistical quantile-factor models (see Ando and Bai (2019); Belloni et al. (2019); Chen et al. (2021)). In short, the interpretable quantiles pricing that we propose can be obtained by stacking the quantile slopes of cross-sectional quantile regressions of the returns on firms' characteristics for all months in the working sample. Our approach, on top of providing interpretability of the quantile-factors, by construction, also relaxes the assumption of constant slopes associated with each characteristic made in past quantile-factor literature. Our model is also a generalization of the factor model proposed by Fama and French (2019), which has time varying factors loads, but does not consider changing slopes across the quantiles of R_{it}. Moreover it relaxes, in words of Fama and French (2019, p. 1893): "the unrealistic assumption that the disturbances in (1) are cross-sectionally iid". Equation 5 gives an alternative representation of the model in Equation 3, closer to traditional asset pricing specifications:

$$R_{it} = R_{z_t}^{\theta} + R_{Xt}^{\theta} X_{it-1}' + e_{it}^{\theta}, \tag{5}$$

where $R_{Xt}^{\theta} = \left\{R_{MC_t}^{\theta}, R_{BM_t}^{\theta}, R_{OP_t}^{\theta}, R_{INV_t}^{\theta}\right\}$ is a vector or parameters to be estimated, the quantile-dependent idiosyncratic error e_{it}^{θ} is assumed to satisfy the following quantile restriction $P\left[e_{it}^{\theta} \leq 0 \middle| R_{z_t}^{\theta}, R_{Xt}^{\theta}\right] = \theta$. We emphasize that Equation 5 is not a regression. It is a quantile-pricing model in which the factors may be different according to the quantiles that one aims to analyze, and the time-varying factor loads are idiosyncratic characteristics of the firms.

In our results section we include three additional characteristics in our baseline specification. Namely, "momentum", MOM, "liquidity", LIQ, and the "market beta" $(beta)$, which renders the following model:

$$\begin{aligned} R_{it} = R_{z_t}^{\theta} + MC_{it-1} R_{MC_t}^{\theta} + BM_{it-1} R_{BM_t}^{\theta} + OP_{it-1} R_{OP_t}^{\theta} + INV_{it-1} R_{INV_t}^{\theta} \\ + MOM_{it-1} R_{MOM_t}^{\theta} + beta_{it-1} R_{beta_t}^{\theta} + + LIQ_{it-1} R_{LIQ_t}^{\theta} + e_{it}^{\theta}, \end{aligned} \tag{6}$$

The inclusion of these three characteristics obeys a large body of empirical studies that has consistently documented the importance of these characteristics (see for instance Campbell (2017, Ch.3), Bali et al. (2016) for a review of the subject).

3.1. Time and quantile goodness of fit

We assess the relative effectiveness of the cross-sectional characteristics to explain returns alongside the quantiles and across time. To evaluate individual asset characteristics we estimate the t-statistics associated to the quantile coefficients at each cross-sectional estimation step. Complementarily, to examine the joint local adjustment of the model we estimate the pseudo-R squared statistic, $R^2(\theta)$, proposed by Koenker and Machado (1999) given by:

$$R^2(\theta) = 1 - \hat{V}(\theta)/\tilde{V}(\theta), \tag{7}$$

where $\tilde{V}(\theta)$ stands for the square residuals of a restricted optimization problem that only includes an intercept within the set of regressors, specific to the θ quantile, and $\hat{V}(\theta)$ is the residual variation left unexplained by a given conditional quantile asset pricing model. Note that by construction $R^2(\theta)$ lies between zero and one. Unlike a traditional R^2 statistic which measures the relative performance of two linear models while explaining the conditional mean function, $R^2(\theta)$ is a measure of the relative success of the corresponding quantile regression. Therefore, $R^2(\theta)$ constitutes a local measure of goodness of fit for a particular quantile (and a particular month), rather than a global measure of the goodness of fit.

3.2. Adding new characteristics

We also aim to increase the model performance as measured by the quantile- R^2. Nevertheless, as emphasized in subsection 3.1 there is one R^2 per each quantile and per each period. We average the statistics in time using equally-weighted averages. But, given that we wish to emphasize the importance of considering the interest in different fragments of the distribution that a market participant may have, we use different weighting schemes across the quantile dimension. In particular, let $\overline{R^2(\theta)}$ be the time-average of the quantile θ. We consider five values of θ, namely $\theta_i \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$, for $i \in \{1, \dots, 5\}$, then our global goodness of fit statistic will be given by:

$$\text{Avg } R^2 = \sum_{i=1}^{5} a_i \, \overline{R^2(\theta_i)}, \tag{8}$$

Note that the parameter a_i reflects the importance of a given quantile for the global adjustment of the model. When a_i is small the associated quantile is not relatively important for us, while a large a_i mean that we assign a relatively high importance to the associated quantile. We consider five different weighting schemes, each of them with five components $(a_1, \dots, a_5)$, which aim to reflect different interests by the investor when deciding about a new variable to include in the asset-pricing model. We index each weighing scheme with a different subscript, so that we end up with five $\{w_1, \ldots, w_5\}$. Such that $w_1 = [0.5,0.3,0.15,0.05,0]$, $w_2 = [0.35,0.25,0.2,0.15,0.05]$, $w_3 = [0.2,0.2,0.2,0.2,0.2]$, $w_4 = [0.05,0.15,0.2,0.25,0.35]$ and $w_5 = [0,0.05,0.15,0.3,0.5]$, in this way, w_1 reflects a greater interest on the left tail of the cross-sectional distribution, while w_5 reflects a greater interest in the right tail. The other cases are in-between these two extremes.

4. Data

Our baseline specification uses the following definitions of variables: MC_{it} is the size of the firm, constructed as the natural log of the market capitalization of the firm *i* in period *t*, and it is not included in Chen and Zimmerman (2020) [CZ henceforth] data base, so we constructed it using CRSP monthly data. BM_{it} is the book-to-market ratio constructed as the annual book equity over market equity, and corresponds to "BM" in CZ database. INV_{it} refers to the growth rate of investment in one year, "grcapx1y" in CZ data-base. It was constructed as the growth between one-year lagged capital expenditures and two-year lagged capital expenditures. OP_{it} refers to operating profitability, which is our base line specification is measured as the gross profitability indicator employed by Novy-Marx (2013) calculated as revenue minus cost of goods sold, divided by total assets. This variable is "GP" in the CZ database, and it has the longest and most consistent information between the several profitability measures available in the database. MOM_{it} refers to momentum, and corresponds to Mom12m in CZ, which is constructed using stock return between months t-12 and t-1. $beta_{it}$ stands for the market beta of firm *i* in month *t*, and it corresponds to the simple CAPM beta ("beta" in CZ), constructed as the coefficient of a 60-month rolling window regression of monthly stock returns minus the risk-free rate on market return. LIQ_{it} refers to liquidity, which in our base-line specification corresponds to "illiquidity" in CZ, and is the (i)liquidity measure proposed Amihud (2002) constructed using the twelve month average of daily absolute value of the return divided by turnover.

Our data on characteristics have a monthly frequency, and we combined them with holding period returns in the US market, obtained via CRSP. This firm information database consists in 3.75M observations and 208 characteristics for 26,852 different firms between 1928 and 2019, besides, we have 4.61M returns of capital, that we combine altogether to generate our study database. A description of the data is provided in Table 1, where we present the mean, standard deviation, skewness and kurtosis averaged across time, for the statistics estimated on a monthly frequency for the cross section of characteristics. As it can be seen, there is a great deal of variability across characteristics. For this reason in our explanatory regressions we used standardized variables on the RHS of Equation 6, in order to make regression coefficient among the variables and among the regressions comparable (this is called in the statistical literature as a "beta-coefficient regression"). In this case, the intercepts are by construction set equal to zero, for the average scenario, but not for the quantiles (they become a rough estimate of the unconditional quantiles) and the slopes are directly measured in relative terms, which allow us to read their magnitude on top of their sign, across the variety of our specifications.

Table 1
Summary Statistics

	RET	Beta	BM	MC	LIQ	OP	INV	MOM
Mean	0.81	1.10	-0.72	12.88	0.00	0.26	2.30	0.12
Stand. Dev.	16.76	0.83	1.02	2.10	0.00	0.36	110.27	0.66
Skewness	4.02	2.05	-0.58	0.23	22.66	-4.06	38.09	5.26
Kurtosis	94.82	42.82	2.77	-0.21	821.78	151.16	2467.08	117.29

The table shows the time-averages of the cross-sectional mean, standard deviation, skewness and kurtosis calculated for each month from January 2000 to December 2019, for the market returns (RET), and the following characteristics: Beta, BM (Book to Market ratio), MV (size), LIQ (illiquidity), OP (Profitability), INV (Investment) and MOM (Momentum).

Table 2 provides the evolution in time of the number of firms in our sample. It summarizes the information of our main study period, which is available on a monthly frequency, from January 2000 to December 2019. It presents the information only for five months: January 2005, 2010, 2015 and December 2019. The rows represent the number of firms entering the sample in a particular date. For instance, at the beginning of the sample 7,040 firms entered the sample (because our main sample starts on that particular date), while, in January 2005, 37 firms entered the sample. In December 2019 15 new firms appeared in the sample. It is also possible to examine the permanency of a given set of firms during the sample period. For instance, of the 7,040 firms starting in January 2000, only 1,609 had information by the end of 2019. Of the 20 firms that entered the sample in January 2010, only 9 were still present by the end of 2019.

Table 2

Firms' Movement Triangle: January 2000-December 2019

	2000-01		2005-01		2010-01		2015-01		2019-12
2000-01	7040	...	3998	...	2707	...	2056	...	1609
		⋱	⋮		⋮		⋮		⋮
2005-01			37	...	26	...	14	...	9
				⋱	⋮		⋮		⋮
2010-01					20	...	15	...	9
						⋱	⋮		⋮
2015-01							21	...	12
								⋱	⋮
2019-12									15
TOTAL	7040		5204		4466		4273		4165

Note: The table shows the firms that enter the sample at each point of time (in rows), and how many of the ones that started in a particular date are still present in a latter month (columns). The original tables have a monthly frequency, and it is available from February 1928. We present a summarized version.

The above calculations allow us to highlight the infeasibility of constructing balanced-panels to estimate our pricing models at the tails. Quantiles require more information to be accurately predicted than means, so the loss of information due to balancing the panel structure is even more costly than in traditional empirical models that focus on the estimation of the conditional mean. For instance, if we were going to construct a panel for our main sample, we would end up with 1,609 firms that were present during the whole sample period. Our estimations would use, in such a hypothetical case, 1,609 observations per month. On the contrary if we conduct cross-sectional regressions as we do here, our estimates use 7,040 in 00-01, 5,204 in 05-01, …, 4165 in 19-12. In short, the balanced-panel structure requires dropping 69% of the working sample. But it is not only about the vast loss of information, but also the firms that remain in the sample across the years are likely the strongest ones, so we are facing a case of survivorship bias. Analyzing the effects of characteristics such as value, liquidity, size, etc. would be very inappropriate in the case that we only kept such an unrepresentative subset of the population.

Our main results regard the reduced sample January 2000- December 2019, but we also conduct an historical analysis to test the robustness of our claims, which uses all market information from 1928 to 2019. We focus our main analyses on the reduced and most recent sample. In this regard we follow the insights of past literature that has emphasized the different magnitude of the economic effects of characteristics such as value in recent years against older samples (Cohen et al., 2003; Fama and French, 2021). The same pattern has been established for most of the characteristics, which naturally tend to exert a greater impact on the returns in older samples when they were originally noticed, than in more recent samples (McLean, and Pontiff, J. 2016; Linnainmaa, and Roberts, 2018). Our intention is thus to provide timely and currently relevant results.

5. Results

Our results consist of six parts. In subsection 5.1 we report and describe our estimated quantile factors. We conduct cross-sectional quantile regressions each month and save the quantile slopes associated to each of the seven firms' characteristics in our model: Size (MC), Book to Market (BM), Operating Profitability (OP), Investment (INV), Momentum (MOM), market beta, and illiquidity (LIQ). We stack the quantile slopes associated to each characteristic in a vector, and we call this vector a quantile factor. In subsection 5.2 we evaluate the joint significance of the model across time. In 5.3 we report the individual significance of the characteristics in the pricing equation at different quantiles. These three subsections employ a dataset starting in January 2000 and ending in December 2019, so that we can focus our main analysis in more contemporaneous dynamics, still relevant today. In subsection 5.4 we provide a more historical perspective, enhancing the results from the previous subsections. Thus, we report the historical quantile factors from February 1928 to December 2019. This longer sample comes at a cost: We need to restrict our attention to only four traditional factors: beta, size, liquidity, and momentum. Data on profitability, investment and book-to-market, for the early years, are not as reliable and abundant as for the other four characteristics. In subsection 5.5 we evaluate the robustness of our claims across different economic states (i.e. crisis versus normal periods; before and after the Great Recession). Finally, in the last subsection, 5.6, we show that selecting new characteristics for our pricing equation depends on how we consider the various fragments of the cross-sectional distribution of market returns. In this subsection we consider a subset of 21 characteristics that are candidates to be added to our pricing equation and we define different weighting schemes that reflect a variety of interests of the investor when choosing characteristics. We show that if the maximum number of characteristics is small the pricing models are not necessarily the same for investors that assign more importance to some fragments of the returns' distribution than to others (e.g., left-tail, right-tail, center, etc.).

5.1. *Quantile factors based on characteristics*

We estimate a conditional quantile regression each month in our sample, where the explained (LHS) variable is the cross-sectional returns and the explanatory (RHS) variables are firms' characteristics. In our specification the factor loads are given by the characteristics (which change in time and across firms), not by the quantile slopes, which we stack in order to construct quantile factors series, common to all firms. Hence, quantile factors vary in time and across quantiles, but they do not vary across firms. Before estimation, all the characteristics were standardized to have unit variance and mean equal to zero each month, which eases the comparisons between the quantile slopes (i.e. factors) associated with different characteristics. Note that these transformations do not affect our conclusions because quantile regressions are robust to monotone transformations of original variables. This subsection 5.1 is organized in four parts: The first one shows the main statistics of the estimated quantile factors. The second describes how those effects change across quantiles. In the third part, we change our perspective and compare how fitted quantile factors vary across time. In the fourth and last one, we present and discuss the correlations between quantile factors.

A. Summary statistics of quantile factors

Table 3 shows the summary statistics of the various quantile factors that we estimate. The table shows the mean, median, standard deviation, maximum, minimum, skewness and kurtosis of the intercept of each regression (Alpha) and of the quantile factors associated with each characteristic. The largest effects in relative terms are produced by the variation of market beta across firms joint to the variation of size (MC), as can be observed panels of *Mean* and *Median* in Table 3. However, the distribution of the effects varies greatly across quantiles, and in unpaired ways for the considered characteristics. For instance, let us focus on the effect of market beta. It is positive for the winners' tail of the cross-sectional distribution of the returns reaching a maximum at θ=0.95 (mean=4.19, median=3.05), but it is negative for the losers' tail, again reaching a maximum (negative) effect at θ=0.95 (mean=-3.45, median=-2.89). We can compare it with the effect of profitability (OP), which is very stable across quantiles, and never flips its sign (the mean effect is 0.24 at θ=0.05 and 0.37 at θ=0.95). The sign's flipping also clearly occurs in the case of size (MC) -from mean=3.51 to mean=-4.51-, and illiquidity (LIQ), -from mean=-0.51 to mean=2.96-, at θ=0.05 and 0.95 respectively-. The reverse in the sign is ambiguous for momentum (MOM), book-to-market (BM) and investment (INV), because it clearly occurs just for the farther tail of the distribution when θ=0.95, but not for fragments closer to the center (e.g. θ=0.75).

Comparing *Maxima* and *Minima* in panels 4 and 5 of Table 3, helps us to document further heterogeneities across quantile factors. For instance, the maximum effects (largest values of quantile slopes) tend to occur at the right tail of the distribution for all characteristics (except for size), while the minima effects are more evenly divided in the two tails. The most notorious change in magnitude between the effects reported at the center of the distribution, θ=0.25, 0.5, 0.75 and the tails of the distribution, occurs for illiquidity (max=35.70) and investment (max=25.33) at the right tail, θ=0.95. This can be seen as an indicator of nonlinear effects across the effect of such characteristics on market return, which is particularly notorious for "outstanding" winners in the market.

In terms of *skewness* and *kurtosis* the panorama is not less heterogeneous. In general, factors at the lowest quantiles (associated to the losers' tail) present a negative skewness (except for investment). We emphasize that we are reporting a negative skewness in the *time* distribution of the factors. Thus, it means that the effects on the returns of the characteristics for the losers tend to be more frequently negative than otherwise. Some factors such as those constructed using BM, MC or MOM, are mainly characterized by a negative skewness independently on the quantile.

Finally, according to the kurtosis reported in the last panel of Table 3 (which measures how thick are the tails of the distribution), the factors more prone to outliers are beta and MC, especially at the right cross-sectional tail, and investment across the whole distribution. This coincides with the analysis of maxima and minima that we reported before. Investment effect on prices is mainly due through outliers in time and in the cross-section than as a product of a generalized pattern in the data.

Table 3

Summary Statistics of Quantile Factors

	Alpha	Beta	BM	MC	LIQ	OP	INV	MOM
				Mean				
θ=0.05	-18.07	-3.45	0.61	3.51	-0.51	0.24	-0.02	0.54
θ=0.25	-6.25	-1.97	0.42	1.35	-0.33	0.13	0.04	0.45
θ=0.5	0.10	-0.48	0.25	0.48	-0.09	0.23	0.02	0.32
θ=0.75	6.95	1.30	0.06	-0.65	0.30	0.43	-0.05	0.15
θ=0.95	22.60	4.19	-0.21	-4.51	2.96	0.37	0.25	-0.11
				Median				
θ=0.05	-16.69	-2.89	0.74	3.70	-0.36	0.25	0.09	0.63
θ=0.25	-5.05	-1.58	0.42	1.38	-0.33	0.15	0.00	0.32
θ=0.5	0.26	-0.31	0.21	0.52	-0.09	0.22	-0.04	0.30
θ=0.75	6.30	1.04	-0.09	-0.59	0.01	0.42	-0.11	0.23
θ=0.95	20.24	3.05	-0.45	-4.33	1.39	0.39	-0.35	-0.03
				Standard Deviation				

θ=0.05	7.22	2.75	1.38	1.64	1.83	1.12	1.20	1.35
θ=0.25	5.53	2.58	1.06	1.19	1.08	0.79	0.66	1.26
θ=0.5	5.11	2.74	1.11	1.15	0.87	0.75	0.63	1.34
θ=0.75	6.03	3.26	1.49	1.36	1.40	0.92	0.52	1.60
θ=0.95	10.39	4.98	2.76	2.17	5.92	1.81	2.90	2.13
				Maximun				
θ=0.05	-4.94	1.05	4.60	7.50	6.69	3.39	14.18	3.49
θ=0.25	5.04	6.18	4.48	5.23	4.79	3.27	8.91	4.24
θ=0.5	17.70	13.27	4.26	5.05	3.31	3.30	7.65	4.85
θ=0.75	34.15	21.02	6.65	5.71	7.15	3.83	5.08	7.06
θ=0.95	76.31	39.02	11.30	3.47	35.70	8.77	25.33	10.51
				Minimun				
θ=0.05	-52.07	-15.02	-5.60	-2.29	-9.32	-2.83	-4.40	-3.77
θ=0.25	-33.17	-13.05	-3.29	-3.76	-5.84	-2.05	-1.00	-5.11
θ=0.5	-21.15	-10.61	-6.88	-4.75	-4.16	-1.89	-0.96	-5.97
θ=0.75	-10.01	-8.05	-11.28	-7.02	-5.30	-2.63	-1.08	-7.24
θ=0.95	3.70	-3.55	-24.97	-18.36	-6.63	-8.05	-1.75	-9.21
				Skewness				
θ=0.05	-1.32	-1.51	-0.40	-0.56	-1.02	-0.02	6.38	-0.35
θ=0.25	-1.18	-1.15	-0.10	-0.51	-0.59	0.23	10.08	-0.22
θ=0.5	-0.31	0.14	-0.39	-0.06	0.15	0.08	8.52	-0.50
θ=0.75	0.85	1.62	-0.90	0.48	1.42	0.16	5.56	-0.63
θ=0.95	1.87	3.01	-2.13	-0.78	2.59	0.19	7.02	-0.25
				Kurtosis				
θ=0.05	2.56	2.74	1.30	0.97	5.19	0.09	80.85	0.13
θ=0.25	2.54	2.51	2.63	2.03	6.30	1.10	132.21	1.66
θ=0.5	1.69	3.41	8.06	3.30	4.04	1.20	95.51	3.08
θ=0.75	2.52	7.04	15.29	5.15	5.33	1.68	48.60	3.64
θ=0.95	6.03	13.93	27.98	7.21	9.06	3.44	53.37	3.90

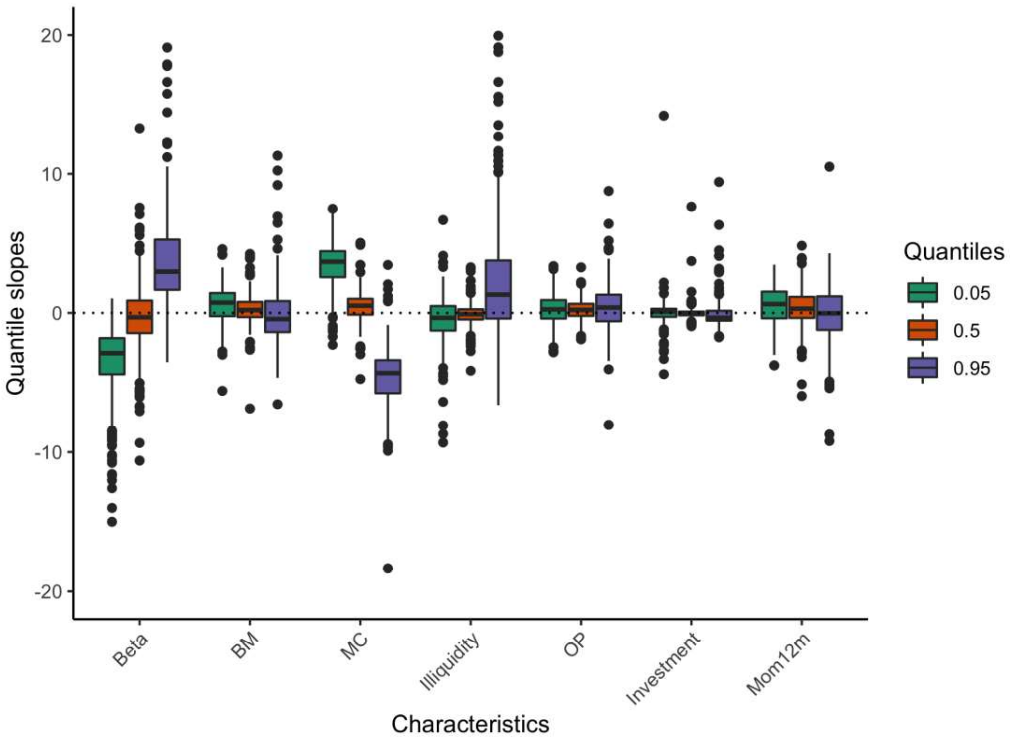
Note: The table shows the time-series average, standard error, skewness, kurtosis, maximum and minimum of the estimated quantile factors from January 2000 to December 2019.

B. *Effects of characteristics across quantiles*

In Figure 1 we plot the same information but this time we present it using a box-plot, which allows us to clearly identify patterns across quantiles, ignoring the time variation. The more notorious fact is that market beta and size are featured by opposite signs on the two tails of the cross-sectional distribution of the returns. That is, on the one hand, an increment in the market beta is associated with an increase of the winners' tail, because we report a positive quantile slope at $\theta = 0.95$, while it is associated to a decrease of the losers' tail, i.e. negative quantile slopes at $\theta = 0.05$. The effect in the median is somewhere in between, and for this reason is not significant for a large number of firms. On the other hand, an increment in size is associated to a reduction of the quantile of market returns at $\theta = 0.95$, and an increment of the quantile of market returns at $\theta = 0.05$. Putting this together means that an increment of beta enlarges the whole support of the unconditional market return cross-sectional distribution, while an increment of size shrinks this support. Another way to analyze this effect is by thinking of the support of the distribution as a measure of the dispersion of the returns, as it is clearly the case when we use the interquartile range as a proxy for volatility. In this case, while beta increases the cross-sectional volatility, size decreases it.

Some factors such as illiquidity, which is measured following Amihud (2002), as the twelve month average of daily return in absolute value, divided by the turnover, present a clearly more significant effect at the winners' tail than at the center or the losers' tail, while some others such as BM seen more important for the left tail than for the right tail.

Figure 1. Quantile Factors



Note: The figure shows the distribution from January 2000 to December 2019 of the quantile seven-factors model $\theta = \{0.05, 0.5, 0.95\}$. We plotted the values between -20 and 20, in order to make the effects more evident, so than some points for liquidity are not visible in the figure, but this does not alter our main conclusions at all.

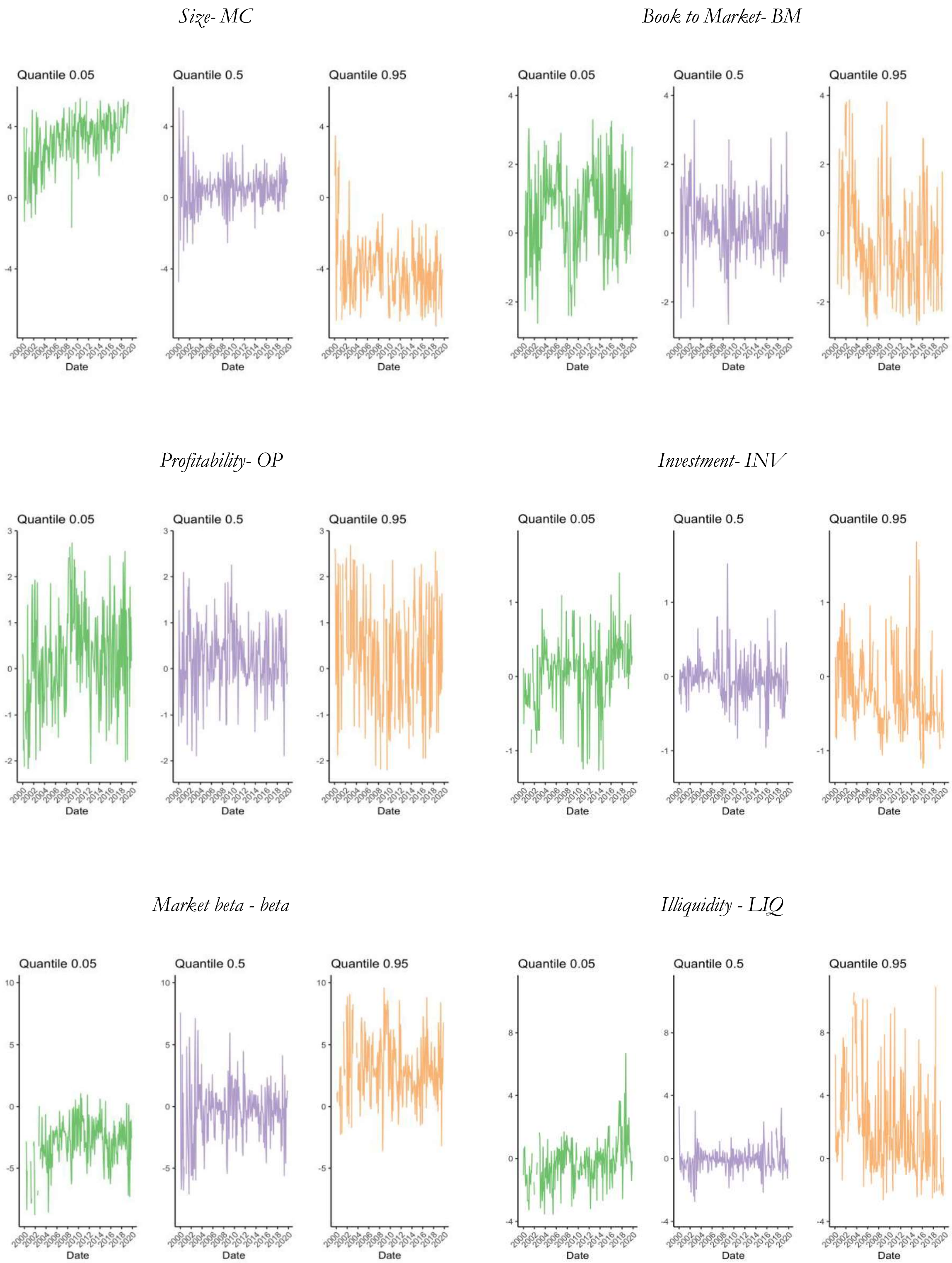
Profitability and investment show more stable effects across quantiles, at least for most of the months. Nevertheless, investment is characterized by an extremely high dispersion, with slope values of large magnitude at the right tail of the time distribution of the slopes (feature that is shared with the liquidity factor), so that it tends to exert its effect mainly through occasionally large shocks to the winners' tail of the cross-section of the returns. Momentum also tends to explain the cross-section at similar magnitudes across quantiles, but the dispersion between the effects (i.e. the monthly slopes of the quantiles) is larger at the tails, specially the right tail, than at the center of the distribution.

C. *Quantile factors across time*

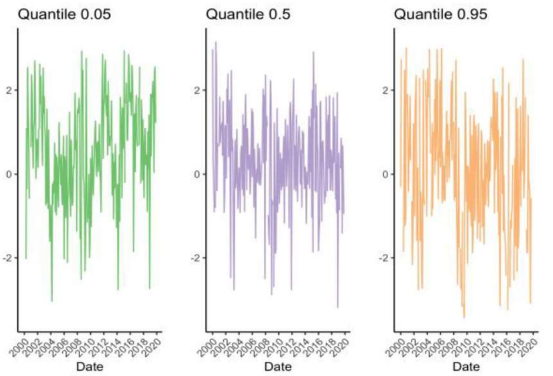
Figure 2 shows the evolution of the quantile factors from January 2000 to December 2019, which can be interpreted as the time-varying effects of a given characteristic on a specific quantile of the cross-sectional distribution of stock returns. All the quantile-factors are very volatile in time, which highlights the importance of considering time-varying slopes in the estimation process. Quantile factors are based on returns, in the same way that cross-sectional factors coming from the slopes of linear regressions, as in Fama and French (2019), are based on returns, which explains the high volatility observed in Figure 2. In short, as it is always the case with pricing factors in asset pricing, the average *in time* of the factors is dominated by its standard deviation.

Note that in our model the factor loads in the asset pricing equation are time-varying (because are given by the characteristics) and the slopes of the quantile regression are time-varying as well. This gives a great deal of flexibility to our estimates, which is necessary because, for instance, the fact that factor loads are time-varying has been noticed at least since Rosenberg (1974), and more recently revisited by authors such as Avramov and Chordia (2006) or Fama and French (2019). Our findings help to establish the same fact also across quantiles. That is, not only the effects of characteristics at the mean of the distribution of the returns are described by time-varying factors loads but this occurs as well at the quantile level.

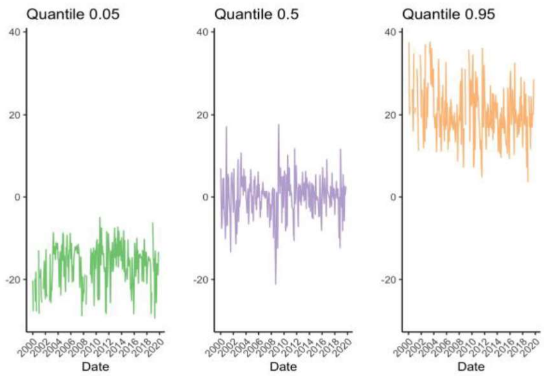
Figure 2. Evolution of Quantile Factors Jan 2000- Dec 2019



Momentum 12 - MOM



Quantile Intercepts



Note: The figure shows the evolution of the quantile factors for the seven-factor model estimated at $\theta = \{0.05, 0.5, 0.95\}$ from January 2000 to December 2019. In the vertical axis is shown the effect of a given characteristic on the cross-sectional quantile of market returns, while the horizontal effect is the corresponding month. All the characteristics were normalized to have zero mean and unitary variance each month.

D. Correlation between the quantile factors

In Table 4 we show the correlation between the quantile factors within characteristics. As highlighted by Chen et al (2021) a low correlation across different quantile factors means that the information convey by such factors is dissimilar, and therefore, linear regressions, which condense information at the center of the distribution of the LHS variables are inadequate, or analogously that PCA is inappropriate, because it focuses exclusively on the average scenario. The evidence reported in this respect is conclusive. Namely, for all factors the correlation steadily decreases when the distance between the quantiles increases. In other words, the effects of asset characteristics are very different across the conditional distribution of the returns. In some cases, such as liquidity and investment, this correlation even changes its sign for the farther away factors (ρ=-0.2 between LIQ at θ=0.05 and θ=0.95 and INV at the same quantiles). Factors are more similar if they are closer. For instance, regarding beta, the correlation between the beta-factor at θ=0.05 and θ=0.25 and between θ=0.95 and θ=0.75 is 0.9, while it reduces to ρ= 0.5 between the factors at θ=0.05 and θ=0.75, and even further to ρ=0.3 between the factors at θ=0.05 and θ=0.95. The same pattern is detected for all the variables, so Table 4 provides solid evidence about the heterogeneity of the effects of asset characteristics on the cross-sectional returns distribution, and of course, on the presence of non-redundant quantile factors in asset prices.

Table 4

Correlations (ρ) 'within' Factors at Different Quantiles

	Beta					BM				
	θ=0.05	θ=0.25	θ=0.5	θ=0.75	θ=0.95	θ=0.05	θ=0.25	θ=0.5	θ=0.75	θ=0.95
θ=0.05	1.0					1.0				
θ=0.25	0.9	1.0				0.8	1.0			
θ=0.5	0.7	0.9	1.0			0.7	0.9	1.0		
θ=0.75	0.5	0.8	0.9	1.0		0.5	0.8	0.9	1.0	
θ=0.95	0.3	0.6	0.8	0.9	1.0	0.3	0.5	0.7	0.9	1.0
	MC					LIQ				
	θ=0.05	θ=0.25	θ=0.5	θ=0.75	θ=0.95	θ=0.05	θ=0.25	θ=0.5	θ=0.75	θ=0.95
θ=0.05	1.0					1.0				
θ=0.25	0.9	1.0				0.5	1.0			
θ=0.5	0.6	0.9	1.0			0.1	0.5	1.0		
θ=0.75	0.4	0.7	0.9	1.0		0.0	0.3	0.6	1.0	
θ=0.95	0.0	0.3	0.5	0.8	1.0	-0.2	-0.1	0.3	0.6	1.0
	OP					INV				
	θ=0.05	θ=0.25	θ=0.5	θ=0.75	θ=0.95	θ=0.05	θ=0.25	θ=0.5	θ=0.75	θ=0.95
θ=0.05	1.0					1.0				
θ=0.25	0.8	1.0				0.6	1.0			
θ=0.5	0.6	0.9	1.0			0.3	0.7	1.0		
θ=0.75	0.4	0.7	0.9	1.0		0.1	0.4	0.8	1.0	
θ=0.95	0.2	0.4	0.6	0.8	1.0	-0.2	0.0	0.3	0.4	1.0
	MOM									
	θ=0.05	θ=0.25	θ=0.5	θ=0.75	θ=0.95					
θ=0.05	1.0									
θ=0.25	0.8	1.0								
θ=0.5	0.6	0.9	1.0							
θ=0.75	0.5	0.8	0.9	1.0						
θ=0.95	0.2	0.5	0.7	0.8	1.0					

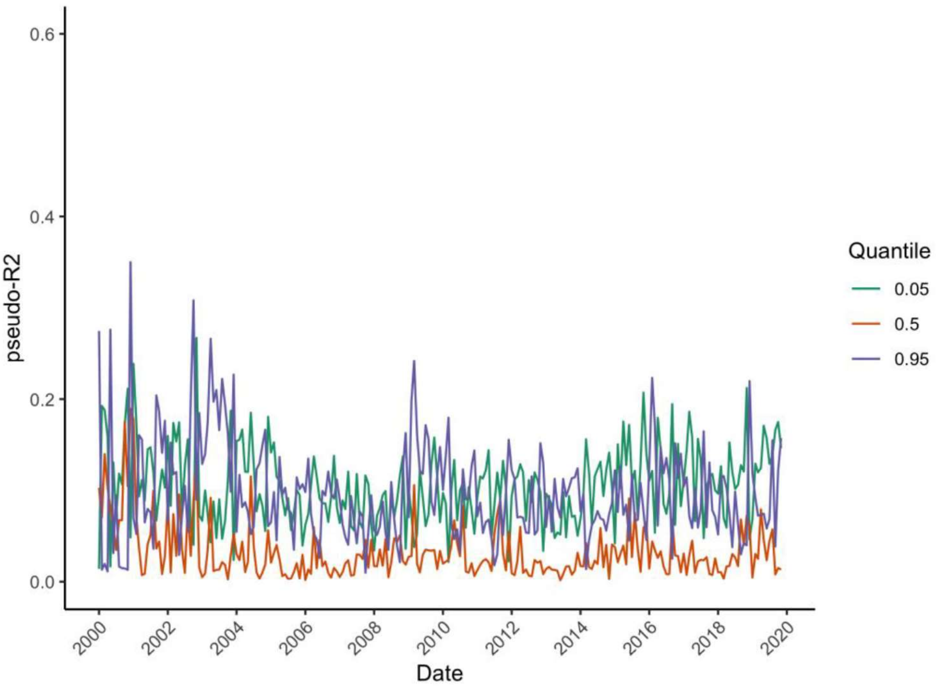
Note: The table shows Pearson's correlation within factors. That is, the correlation of the factors at different quantiles for the same characteristic. The sample starts in January 2000 and ends in December 2019.

5.2. *Analysis of the joint significance in time*

Figure 3 shows the evolution in time of the (pseudo) R-square defined in Equation 7, for the sample running from January 2000 until December 2019. The cross-sectional approach that we follow enables us to use the R^2 as a local measure of the goodness of fit of pricing characteristics, because we are able to estimate one statistic for each quantile, for each month. We restrict our attention again to the two tails of the distribution located at θ=0.05 and θ=0.95 and to the median, θ=0.50.

Intermediate results are consistent with the ones reported in this section and they are available for θ=0.25 and θ=0.75, upon request.

Figure 3. Time Dynamics of Pseudo- R^2 by Quantile



Note: The figure shows the evolution from January 2000 to December 2019 of the pseudo-R^2 given by $R^2(\theta) = 1 - \hat{V}(\theta)/\tilde{V}(\theta)$, where $\tilde{V}(\theta)$ is the square residuals of a restricted model that only considers an intercept, specific to the θ quantile, and $\hat{V}(\theta)$ is the residual variation left unexplained the conditional quantile asset pricing model. By construction $R^2(\theta)$ lies between zero and one. We have plotted in the figure the corresponding statistics for $\theta = \{0.05, 0.5, 0.95\}$.

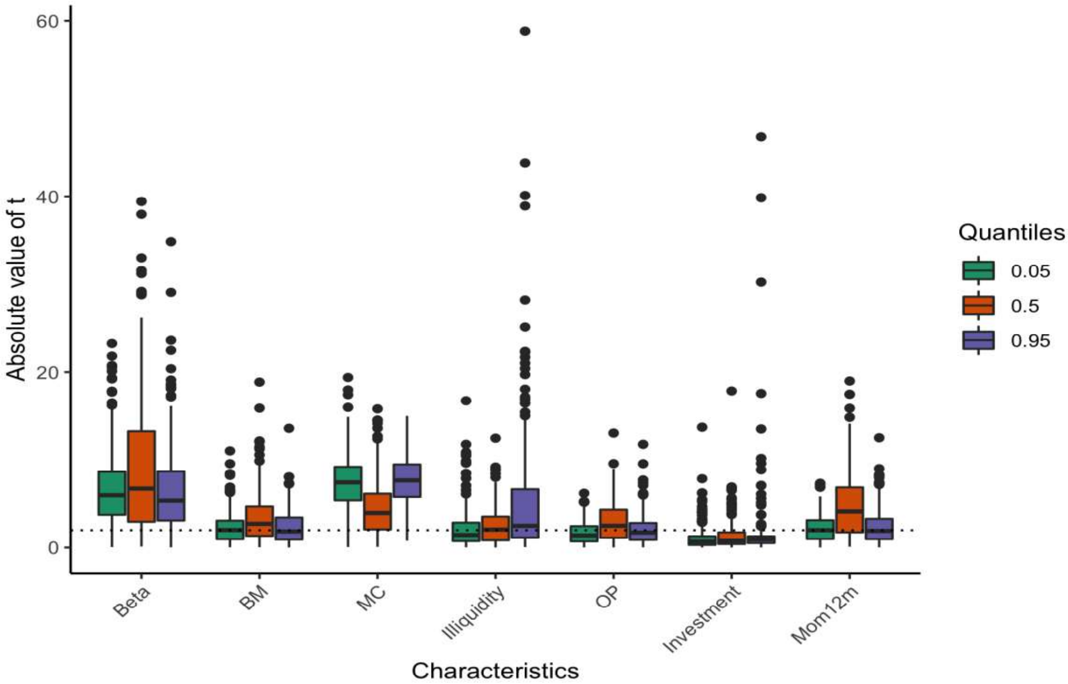
The first thing that can be noticed from examining the figure is that the time variation of the statistic is larger than its quantile-variation, nevertheless both sources of heterogeneities in the model adjustment are important. There are not clear trends or cycles in the model adjustment, which does not seem particularly affected by the 2008-2009 crisis either. Regarding the variation in time, the minimum value of the R^2 for θ=0.05 is 0.01 and it was recorded in January 2000, while the maximum value is 0.26, recorded in November 2002, amounting to a difference between the two of 0.25. When we focus on the highest quantile, θ=0.95, the minimum R^2 is 0.01 and the maximum 0.35, recorded in October 2007 and December 2000, respectively (with a difference of 0.34). For the median these numbers are 0.19 (December 2000) and 0.001 (June 2013), amounting to a difference of 0.19. With respect to the variation of R^2 between the quantiles, the model does a much better job explaining the tails of the distribution (average R^2 for θ=0.05 is 0.104 and R^2 for θ=0.95 is 0.098) than the center of the distribution (average R^2 for θ=0.50 equals 0.033). This large gap in the explanatory power remains relatively stable across the sample, as the model almost all the time explains better the tails than the center.

The difference in the joint explanatory power of the variables in our specification across the sample period may obey the omission of some factors (characteristics) that appear to be relevant only under specific economic states (and perhaps are short-lived). On the other hand, the relative stability of the explanatory power in the two tails of the distribution which is considerably larger than the explanatory power at the center, means that the explanatory power of the individual characteristics considered in the model significantly changes across the distribution of extreme market returns, and that there exist notable increments in the explanatory power of some factors in the tails compared to the center of the distribution. From these results we still do not know if the relevance of the factors increases symmetrically from the median towards the two tails of the distribution, or if by the contrary, the increments in the explanatory power of some factors are compensated by reductions in the explanatory power of other factors, according to the tail of the distribution that we aim to emphasize. But, in any case, the explanatory power of the characteristics considered in our model changes significantly across the distribution of the returns. In short, the results in Figure 3 point out to the presence of more powerful factors in the tails with respect to the center of distribution.

5.3. Analysis of individual significance

Figure 4 shows the t-statistics at different quantiles, associated to the individual characteristics included in our model. The figure consists of three stacked box-plots for the cases $\theta = \{0.05, 0.5, 0.95\}$ for each characteristic. The horizontal dotted line in the figure marks a t-statistic equal to two. The characteristics that tend to be significant most of the time across all specifications are beta and size. The least significant is investment. Nevertheless, the plot also documents important asymmetries of significance at the quantile level. Beta presents larger t-statistics at the center of the distribution (which are also more dispersed) than at the tails, but they are significant alongside the whole distribution, while size shows the opposite behavior. Momentum presents larger t-statistics at the center of the distribution than at the tails, which means that momentum has more relevance for the median stocks than for the stocks with returns located at the tails of the distribution. Profitability is not as significant as momentum, but it also shows a greater number of statistics above two in the center of the distribution, compared to its tails. On its side, liquidity concentrates a larger number of t-statistics above 2 at the winners' tail of the distribution, compared to the center and the left tail (and also a larger number of outliers in the same tail).

Figure 4. Individual Significance of Characteristics Across Quantiles

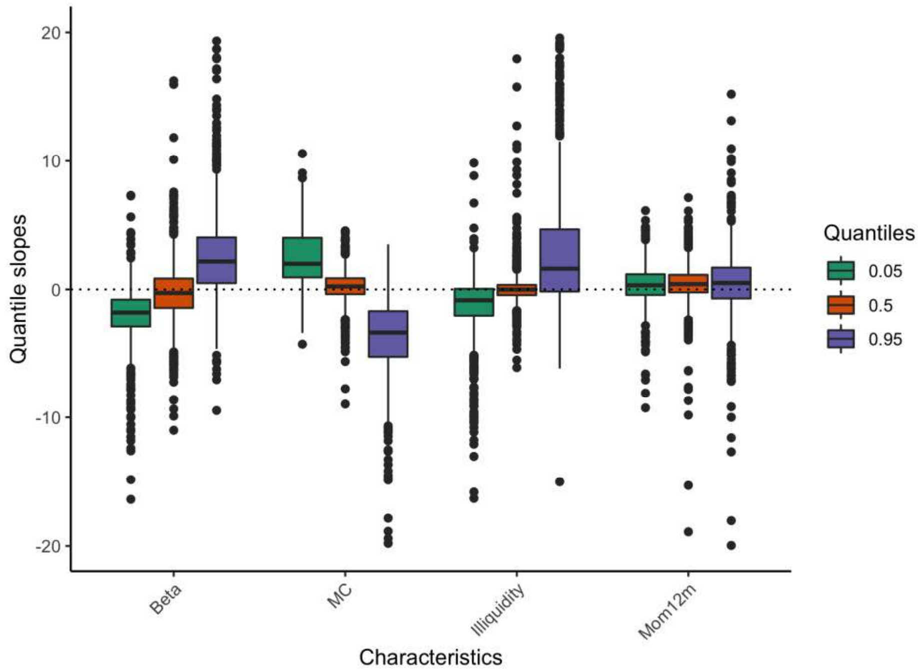


Note: The figure shows the absolute value of the t-statistics from January 2000 to December 2019, associated to the explanatory characteristics in our baseline specification: beta, size, book-market, profitability, investment, liquidity and momentum, for the cases $\theta = \{0.05, 0.5, 0.95\}$. The height of each box is the interquartile range, so that within the box are found 50% of the statistics, while the solid black line is the time-series-median of the statistics. The horizontal dotted line that crosses the figure is the critical value 2. Each t-statistic (in absolute value) above 2 corresponds to the combination of a month and a quantile at which the characteristic was significant to explain the market returns.

5.4. Historical perspective on quantile factors

Figure 5 depicts the evolution of the quantile factors for size, market beta, liquidity and momentum, for which there is enough data to conduct cross-sectional quantile regression starting in February 1928. The asymmetric patterns that we have documented before, regarding the way in which characteristics impact the cross-sectional distribution of the returns, are confirmed by Figure 5. In particular, beta exerts a significant effect on the returns, negative for the left tail and positive for the right tail, while size shows the opposite dynamics. Liquidity is more of a right-tail factor, which gains relevance (and presents the largest variability) for the winners' tail of the distribution. Finally, momentum seems to exert a more modest and stable effect across quantiles.

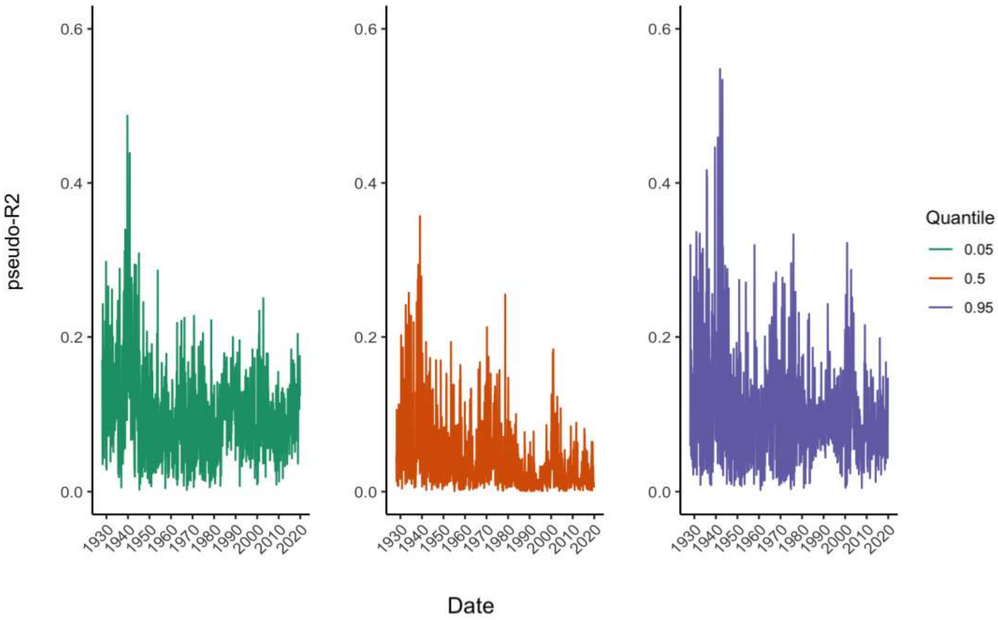
Figure 5. Quantile Effects February 1928 – December 2019



Note: The figure shows the evolution of the quantile four-factor model estimated at $\theta = \{0.05, 0.5, 0.95\}$ from February 1928 to December 2019. In the vertical axis is shown the effect of a given characteristic on the cross-sectional quantile of market returns, while the horizontal effect is the corresponding month. The factors included in the model were selected based on data availability for the longest time span in the sample.

In terms of model adjustment we also confirm our previous analyses. Figure 6 shows the evolution of the quantile R^2 for the two tails of the distribution and the median. During this long sample, which spans almost a century, the two tails are easier to explain than the center of the distribution. It also seems that, in general, model fit was larger in the first half of the 20th century than at the end of the period. It seems as well that there are periods in which model adjustment increases, and some other in which it decreases, describing a cyclical behavior, and the volatility of the statistic was greater at the beginning of the sample, which is expected as the number of firms listed at the market was lower. We can also observe that none of the tails is easier to explain than the other, in a general sense, but instead it depends on the period.

Figure 6. Time Dynamics of Pseudo- R^2 by Quantile

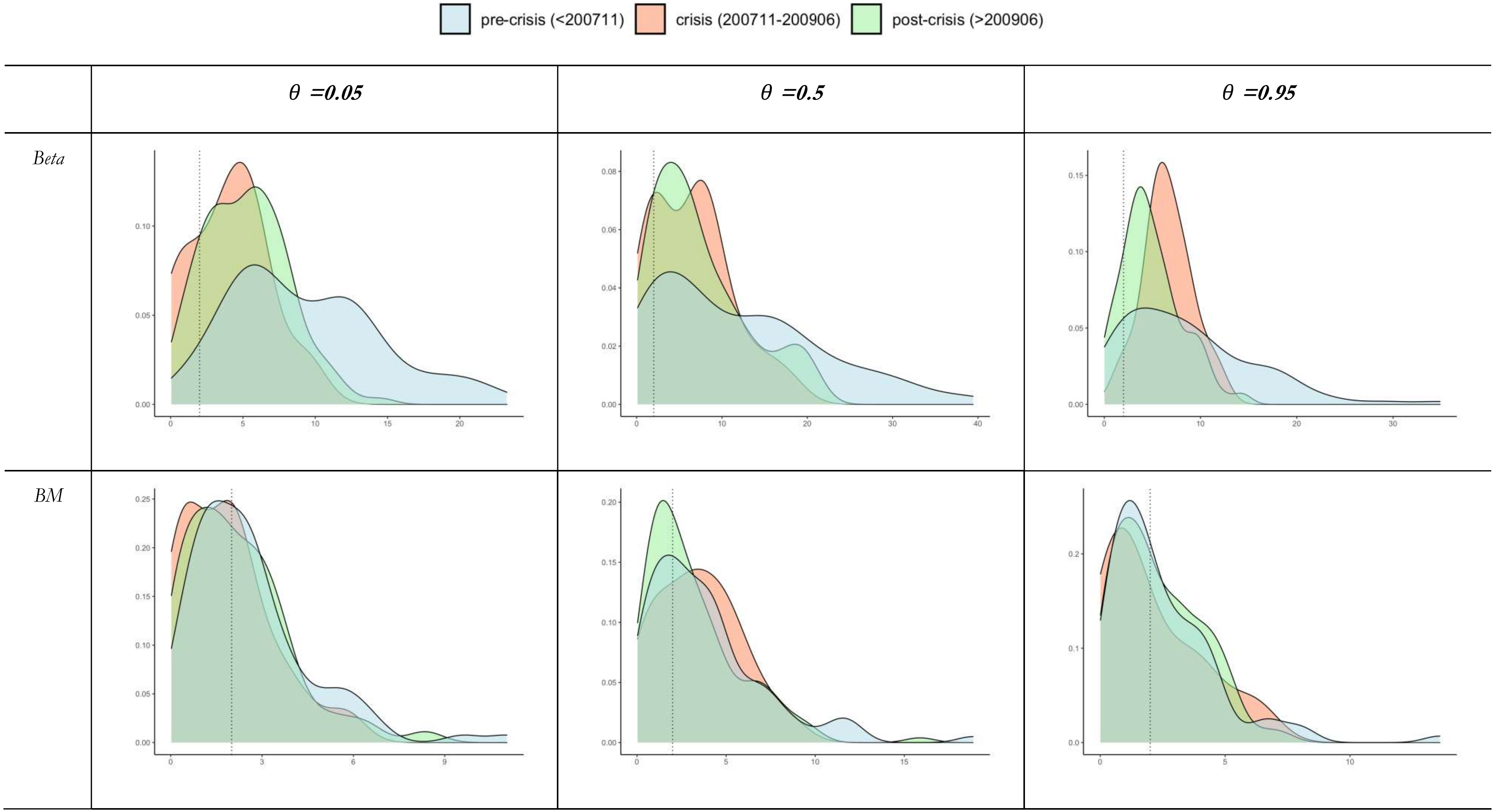


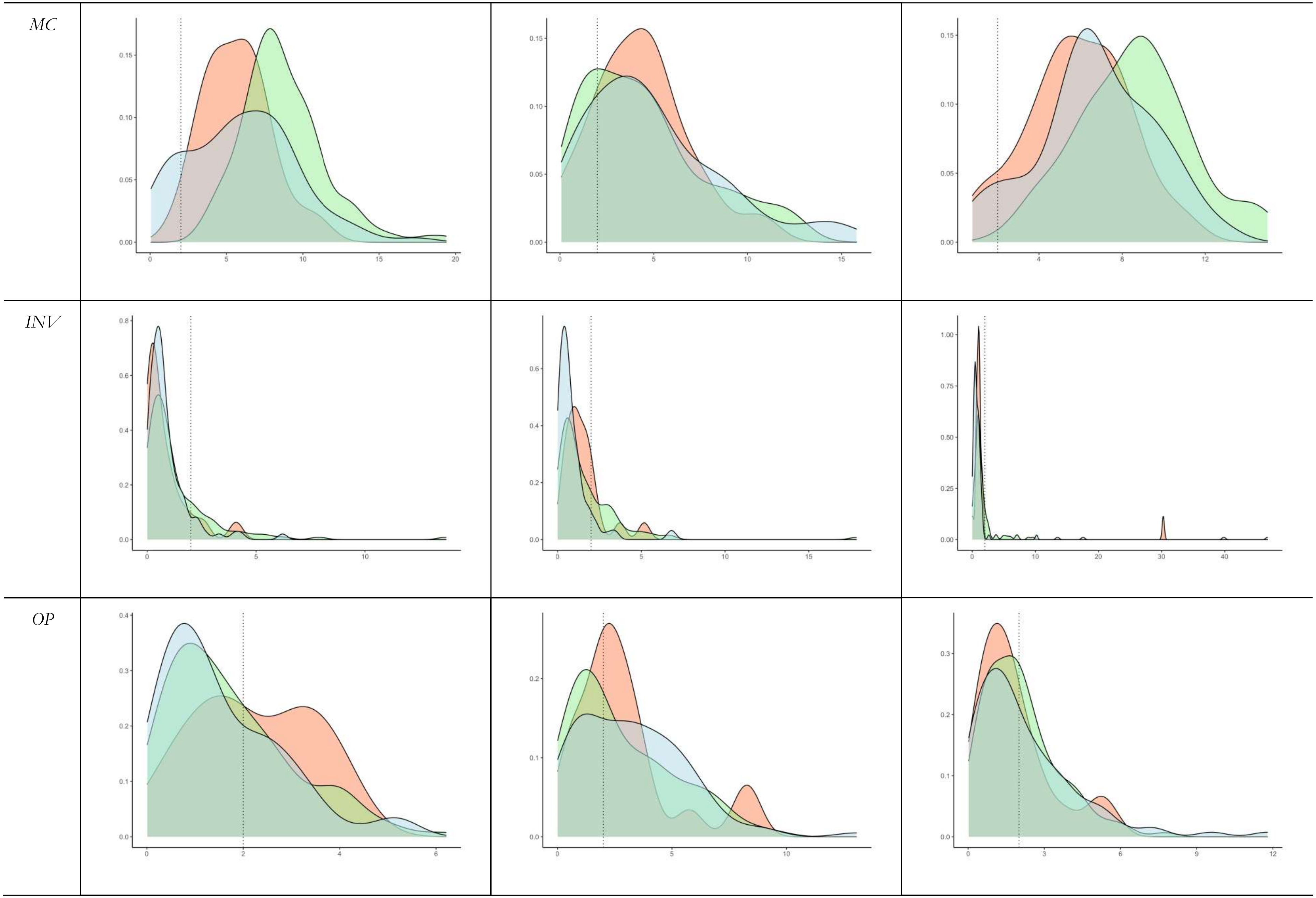
Note: The figure shows the evolution from February 1928 to December 2019 of the pseudo-R^2 given by $R^2(\theta) = 1 - \hat{V}(\theta)/\tilde{V}(\theta)$, as it is defined in equation 7. By construction $R^2(\theta)$ lies between zero and one. We have plotted in the figure the corresponding statistics for $\theta = \{0.05, 0.5, 0.95\}$.

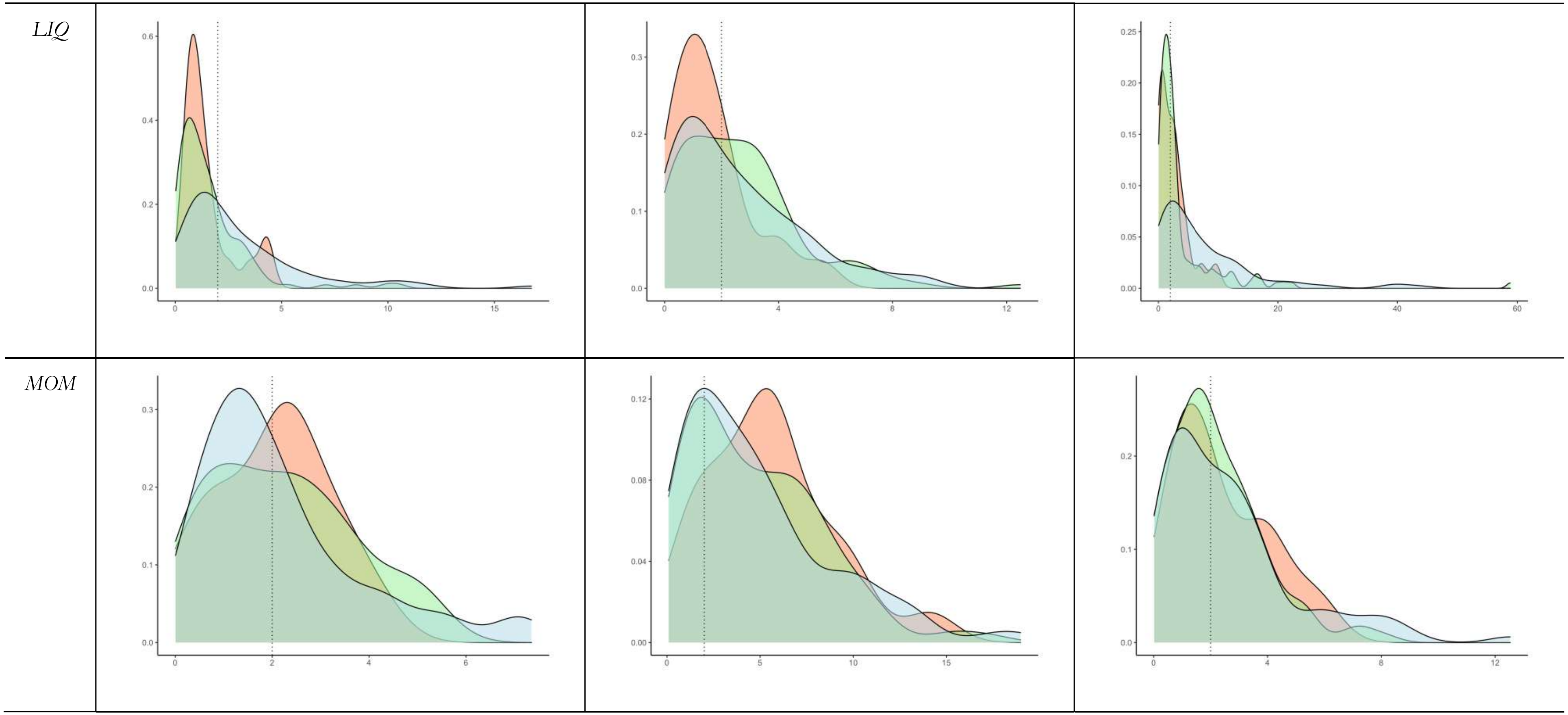
5.5. Comparison under different economic states

In Figure 7 we present differently the information regarding the absolute value of the t-statistics. Namely, it divides our main sample in three different periods: pre-crisis from January 2000 to October 2007, crisis from November 2007 to June 2009, and post-crisis July 2009 to December 2019. We then plotted the kernel estimated densities of the statistics for each subsample at the three quantiles, $\theta = \{0.05, 0.5, 0.95\}$. Each subplot in the figure corresponds to a given characteristic at a given quantile. If the densities differ significantly, it means that the relevance (assimilated to statistical significance in the figure) has changed during the sample period, for instance as a consequence of a changing macroeconomic state, which could be associated to new emerging factors or with a reduction (or increment) of the influence of the old factors included in the model. On the other hand, if the densities look alike for the different subsamples, but they do not across the columns of the table, means that there are asymmetries across the quantiles, additional to those documented in the previous subsections.

Figure 7. T-statistics before the crisis (Jan 00- Oct 07), during the crisis (Nov 07- Jun 09) and after the crisis (Jul 09- Dec 19)




MC
INV
OP



Note: The figure shows the densities of the absolute value of t-statics for each explanatory characteristic -top to bottom- at three different quantiles, $\theta = \{0.05, 0.5, 0.95\}$ -left to right-. The densities correspond to three periods within the sample: pre-crisis from January 2000 to October 2007, crisis from November 2007 to June 2009, and post-crisis July 2009 to December 2019. The dotted vertical line in each subplot corresponds to the critical value of 2.

Regarding the market beta, in the first row of the figure, the crisis and post-crisis periods present densities relatively similar, in contrast, before the crisis the t-statistics tend to be larger, so the distribution is characterized by a longer right tail in the pre-crisis period. There are also differences across the quantiles, notably the statistic density depicts a larger tail towards the right of the time distribution for ***θ =0.5*** and ***θ =0.95***. In all the cases the peak of the densities is recorded well above the value of 2, which attests to the significance of market beta at explaining cross-sectional differences of the returns.

With respect to the characteristic associated to value, BM in the figure, the three densities are very similar for ***θ =0.05*** and ***θ =0.95***, while the density of the crisis period in the median scenario housed larger t-statistics, meaning that during the crisis, value gained statistical power for the stocks generating returns close to the center of the cross-sectional distribution. Compared to beta, BM is less significant. We can notice as well that the t-statistics are generally larger when ***θ =0.5*** than otherwise. Interestingly, we do not find any reduction in the individual significance of value during the sample period for ***θ =0.05*** and ***θ =0.95***. Only when θ =0.5 one could say that the importance has been reduced after the crisis. Although we cannot directly compare our results with the previous literature that documents the decreasing importance of value as a factor (see references in Fama and French (2021)), because we have focused on a recent sample in this subsection, we can exert that at least in the last twenty years BM remains equally important across time, especially at the two tails of the cross-sectional distribution of the returns. The importance has decreased only at the center of the distribution compared to the crisis period.

Now we turn our attention to size, MC in the figure. Size shows heterogeneities at the two levels that we are analyzing. At ***θ =0.05***, left column of the figure, size has gained relevance since 2000. Indeed, this characteristic became more important to explain returns, during the crisis (compared with the previous 8 years), and even more relevant after the crisis, as witnessed by a peak of the distribution shifted to the right in time. At ***θ =0.50***, central column of the figure, the three densities of the subsamples are more similar, and indeed it seems that size gained some relevance during the crisis that disappeared afterwards. At ***θ =0.95***, the right-column, the densities of the pre-crisis and crisis periods look alike, while the peak of the density after the crisis has shifted to the right, meaning that the size characteristic has gained some power to explain the winners' tail of the cross-section across the years.

Investment tells us a different history. It is clearly less significant than all of the aforementioned characteristics, as can be noticed observing the peaks of the densities, located in all the cases before the dotted lines that mark a critical value equal to 2. When investment is significant, it has very long tails, and this pattern is present alongside the three quantiles, plotted in the figure, especially at $\boldsymbol{\theta = 0.95}$. An important asymmetry with respect to investment appears when we compare the median case with the tails. For the median, the statistical significance of investment decreased during the crisis, and it has increased since then, reaching similar pre-crisis levels for the third subsample (2008-2019), while it remained stable for the tails during the last decades.

Profitability gained relevance during the crisis, particularly for the worst performing assets, when $\boldsymbol{\theta = 0.05}$, but since then it has returned to its previous levels. In general, profitability is not as significant as the most traditional characteristics in our model (i.e. beta, size), but it is more relevant than investment for the cross-section of the returns. It also exerts a stable impact on the cross-section at different quantiles and times.

Liquidity is a very especial factor. In short, the higher the quantile the larger the effects of liquidity on the returns. We can observe that liquidity houses very high values for winners' stocks, which confirms our classification of liquidity as a right-tail factor. That liquidity is a nonlinear factor can be rationalized by theoretical studies on the limits of arbitrage, that point out to a nonlinear relationship between liquidity and prices.

Finally, regarding momentum, at both $\boldsymbol{\theta = 0.05}$ and $\boldsymbol{\theta = 0.50}$, it gained some additional relevance during the crisis. It also seems that momentum has gained relevance for the lower and central quantiles of the distribution of the returns in recent years, as witnessed by the fact that the density of the t-statistics has shifted to the right for $\boldsymbol{\theta = 0.05}$ and $\boldsymbol{\theta = 0.50}$. Nevertheless, for at $\boldsymbol{\theta = 0.95}$, the three densities are stable in time.

All in all, as it can be perceived after a general examination of the subplots in the figure, that the divergence across the quantiles' patterns documented in the previous subsections dominate the changes experienced by individual characteristics in terms of statistical significance across macroeconomic states.

5.6. Competing characteristics, weighting the tails

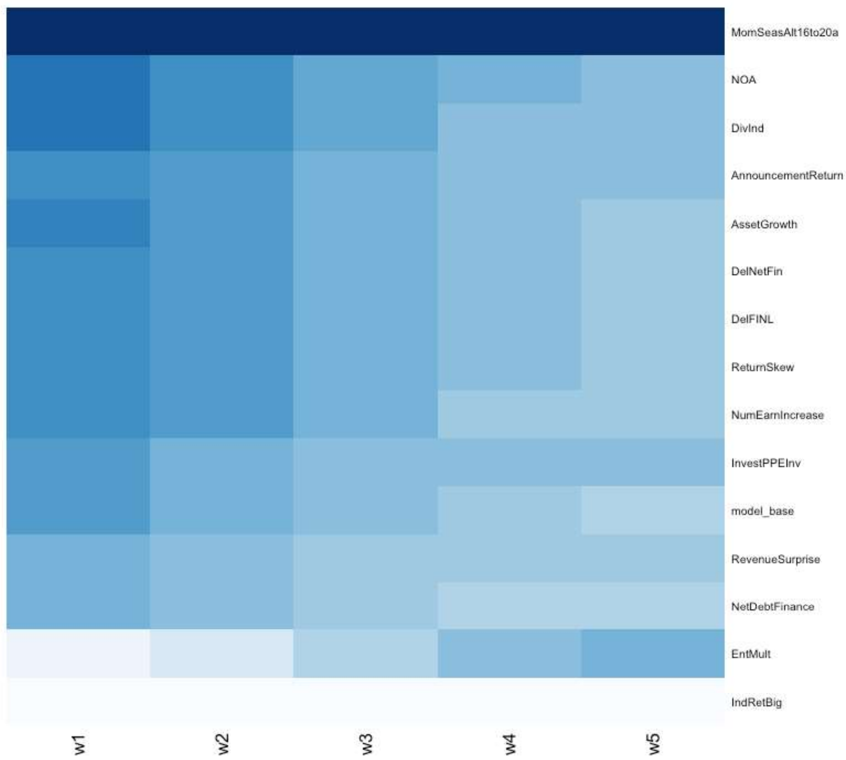
In this section we illustrate the importance of considering the various motivations that an investor, regulator or market analyst may have when they study the cross-sectional distribution of asset returns. These motivations can make market participants more interested in some fragments of the cross-sectional distribution than in others. For instance, a regulator may be more interested in evaluating the expected performance of the left tail of the market returns, which consists of the stocks with the most negative performance, and therefore, is related to the market's downside risk. On the contrary, an investor may be more interested in evaluating the future performance of the right tail of the market returns, the so-called upside risk, if he holds a short position against a market index. It could also be the case that, given the nature of the stocks in our portfolio, we could expect variations of the returns of our own portfolio closer to the mean of the market distribution and in this case, we will not be equally interested in the market tails than in the center of the distribution. In all the cases above, the interest could be in principle connected with the attitude towards risk of the agents, which in turn, has an effect on the risk statistic that they set to minimize in their optimization program, i.e., variance, left- or right high- quantiles, left- or right- low quantiles, etc. All this information is lost in traditional empirical asset pricing exercises that exclusively focus on the average return.

There is a large set of variables to choose from for adding to our baseline specification. Chen and Zimmerman (2020) compare the performance of individual clear predictors of the cross-section, by reporting the associated t-statistic of each predictor. They built upon 210 firm-level predictors to conduct their calculations. Although these authors are only interested in the prediction of the average returns, their work serves us as a starting point. We select the predictors with highest t-statistic among the 210 predictors considered by Chen and Zimmerman (2020), and then we exclude from this subset similar predictors to the ones already included in our baseline specification. For instance, we only consider one additional momentum characteristic, on top of the momentum predictor that we already have included in Equation 6. Table A1 in the Appendix reports the variables' name, acronym and description, joint to the t-statistics reported by CZ, which we consider in this section as candidates to be included in our model.

As stated in the methodology, which reflect various interests by the investor when deciding about the asset-pricing equation. We index each weighing scheme with a subscript, so that we end up with five $\{w_1, \dots, w_5\}$. Such that $w_1 = [0.5, 0.3, 0.15, 0.05, 0]$, $w_2 = [0.35, 0.25, 0.2, 0.15, 0.05]$, $w_3 = [0.2, 0.2, 0.2, 0.2, 0.2]$, $w_4 = [0.05, 0.15, 0.2, 0.25, 0.35]$ and $w_5 = [0, 0.05, 0.15, 0.3, 0.5]$. w_1 reflects a larger interest on the left tail of the cross-sectional distribution, while w_5 reflects a greater interest in the right tail. The other cases are in-between these two extremes.

Figure 8 presents a ranking of the variables according to the $\text{Avg } R^2$ of the model when they are included in the asset pricing equation. The intensity of the color of each cell indicates the value of the statistic, the darker the cell the greater the explanatory power of the associated model. It is clear that the ranking depends on the weighting scheme. That is evident because the colors are not uniformly distributed across the figure. For instance, let us compared scheme w_1 and w_5 which are the more extreme ways to assign the weight to each quantile. According to these two schemes (and all the schemes in between) the first variable that should be added to our baseline specification is momentum seasonality 16-20 (MomSeasAlt16to20a) proposed among other forms of momentum by Heston and Sadka (2008). In the case of w_1 the second variable in the ranking is Net Operating Assets (NOA) proposed by Hirshleifer et al. (2004), while the second variable in the ranking for w_5 is the enterprise multiple (EntMult) proposed by Loughran and Wellman (2011). There are some variables like Industry return of big firms (IndRetBig) proposed by Hou (2007) which are placed at the bottom of the ranking for all the models. Nevertheless, in order to consider the marginal effect of a new variable, we need to go one step forward (see Figures 9 and 10).

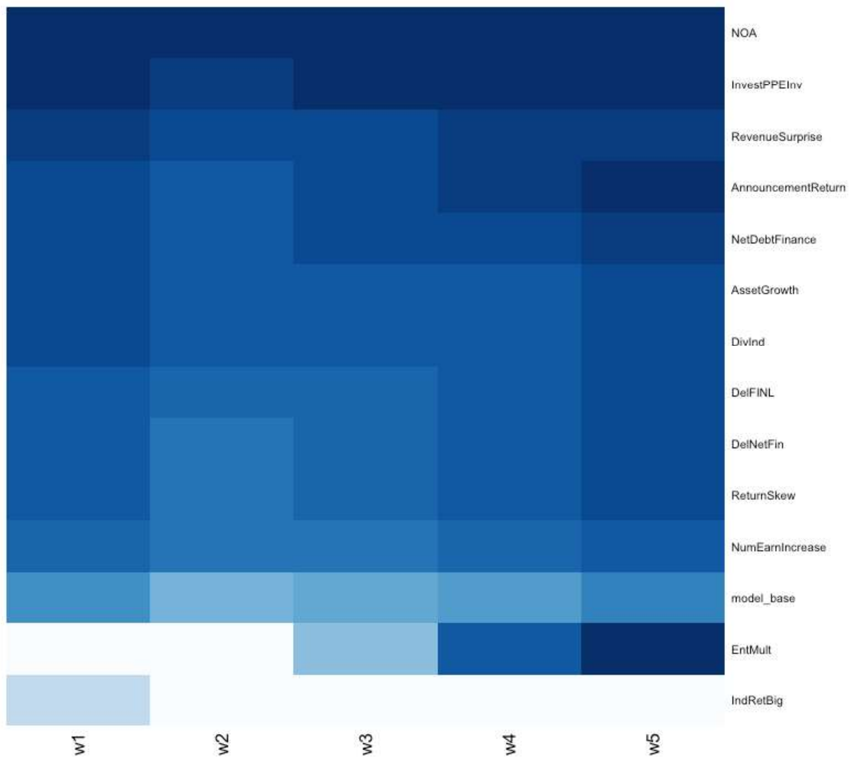
Figure 8. Ranking of New Variables (Round One)



Note: The figure shows the ranking of the characteristics according to different weighting schemes using as a criterion the Avg. R^2 reported in Equation 8 of the main text. The darker the cell the higher the model goodness of fit when a particular variable on the RHS of the figure is added to our baseline specification. In particular we estimate our models for $w_1 = [0.5,0.3,0.15,0.05,0]$, $w_2 = [0.35,0.25,0.2,0.15,0.05]$, $w_3 = [0.2,0.2,0.2,0.2,0.2]$, $w_4 = [0.05,0.15,0.2,0.25,0.35]$ and $w_5 = [0,0.05,0.15,0.3,0.5]$. Moving from w_1 towards w_5 reflects a decreasing interest in the left tail of the distribution, and correspondingly increased interest in the right tail of the distribution. Our sample starts in January 2000 and ends in November 2019.

Now we wonder what if we were interested in considering a second additional variable, on top of MomSeasAlt16to20a. It would not be appropriate just selecting the next variable in the ranking provided in Figure 7, because the estimations that support Figure 8 do not take into account the correlation between two possible added variables, instead, they only consider their marginal effects compared to the baseline model. In this case, what one should do is to construct again the same figure, but taking as a baseline the new updated model that includes MomSeasAlt16to20a. This is what Figure 9 reports. Note that this time we have excluded MomSeasAlt16to20a because it was already incorporated in the first round of our analysis. Now NOA seems to be the best election independently on the weighting scheme.

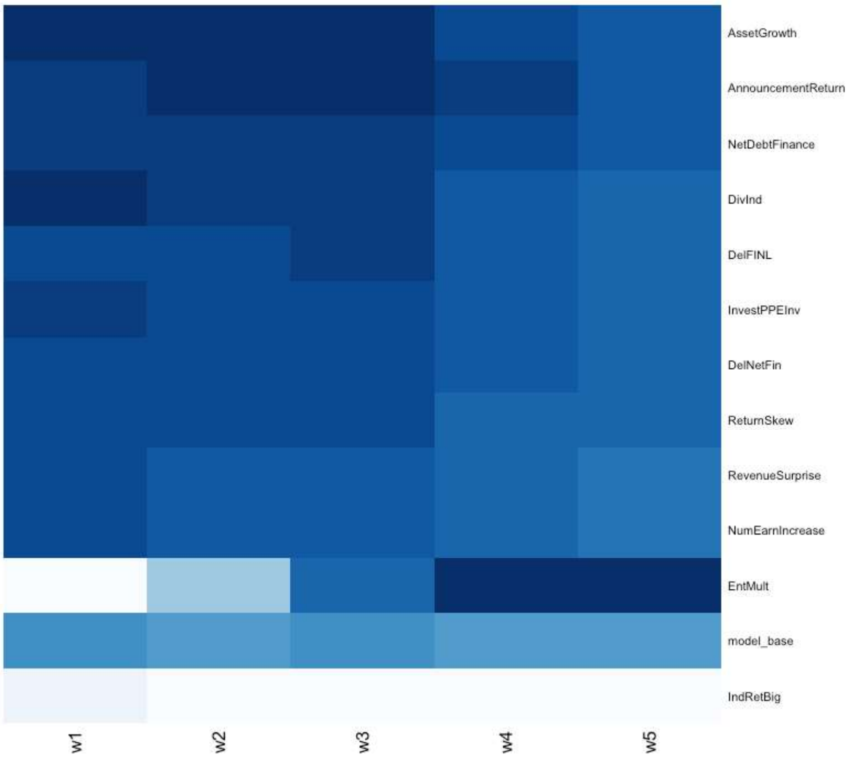
Figure 9. Ranking of New Variables (Round Two)



Note: The figure shows the ranking of the characteristics according to different weighting schemes using as a criterion the Avg. R^2 reported in Equation 8 of the main text. The darker the cell the higher the model goodness of fit when a particular variable on the RHS of the figure is added to our baseline specification. In particular we estimate our models for $w_1 = [0.5,0.3,0.15,0.05,0]$, $w_2 = [0.35,0.25,0.2,0.15,0.05]$, $w_3 = [0.2,0.2,0.2,0.2,0.2]$, $w_4 = [0.05,0.15,0.2,0.25,0.35]$ and $w_5 = [0,0.05,0.15,0.3,0.5]$. Moving from w_1 towards w_5 reflects a decreasing interest in the left tail of the distribution, and correspondingly increased interest in the right tail of the distribution. Our sample starts in January 2000 and ends in November 2019.

If we repeat the same exercise, as we do in Figure 10 one more time, in our third round, something interesting is uncovered. It totally depends on the weighting scheme what variable should be added next to increase our model performance. If we assign a higher importance to the left-tail realizations of the market we would include asset growth, while if we assign more weight to the right-tail realization we should include the enterprise multiple.

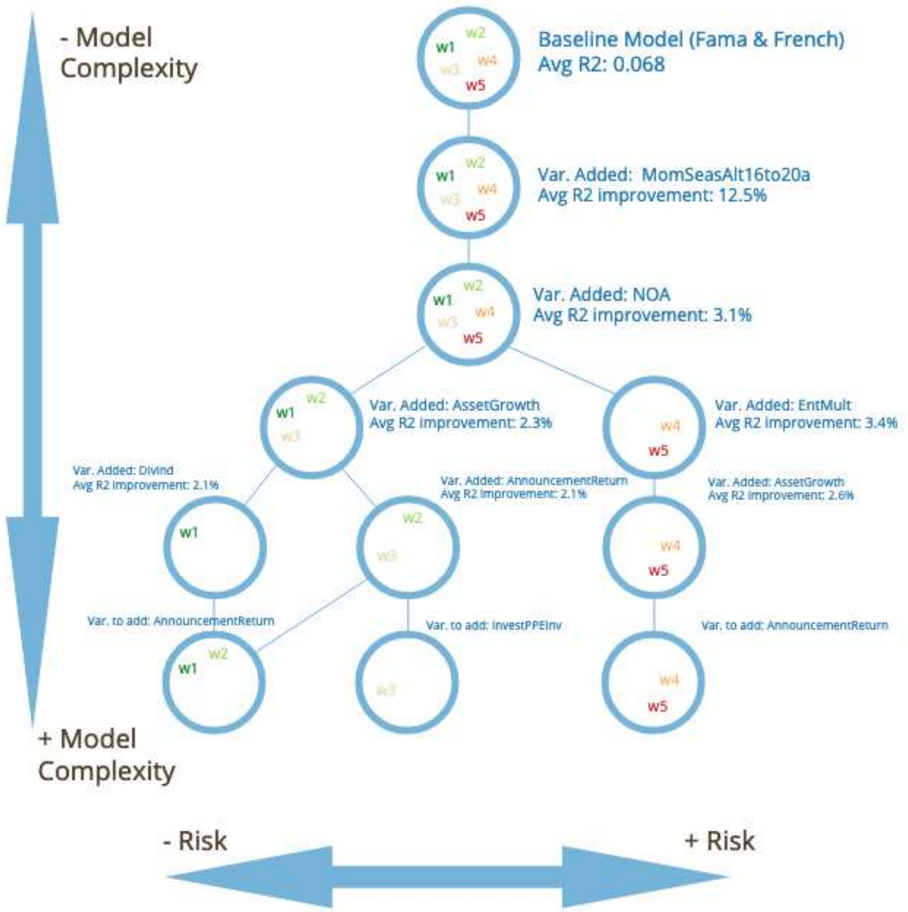
Figure 10. Ranking of New Variables (Round Three)



Note: The figure shows the ranking of the characteristics according to different weighting schemes using as a criterion the Avg. R^2 reported in Equation 8 of the main text. The darker the cell the higher the model goodness of fit when a particular variable on the RHS of the figure is added to our baseline specification. In particular we estimate our models for $w_1 = [0.5,0.3,0.15,0.05,0]$, $w_2 = [0.35,0.25,0.2,0.15,0.05]$, $w_3 = [0.2,0.2,0.2,0.2,0.2]$, $w_4 = [0.05,0.15,0.2,0.25,0.35]$ and $w_5 = [0,0.05,0.15,0.3,0.5]$. Moving from w_1 towards w_5 reflects a decreasing interest in the left tail of the distribution, and correspondingly increased interest in the right tail of the distribution. Our sample starts in January 2000 and ends in November 2019.

What if we were to include an additional variable on top of the last three? In this case, we should plot again the same figure, excluding the variables already incorporated in our pricing model. However, our pricing model would be different depending on the weighting scheme. Figure 11 illustrates this idea up to eight rounds of iteration of the same reasoning. Notice that the decision at each node of the plot from up to down is conditional on our previous decision, so that the final model after three or four rounds of adding a variable would be significantly different. In principle we cannot compare across these models, even using our quantile pseudo R^2. Indeed, different models would be optimal for different agents. This is a point overlook by the literature and traditional asset pricing exercises, which implicitly assume either that the investor only cares about the center of the distribution, so that $w = [0,0,1,0,0]$ or that there are not heterogeneities in the model adjustment (or equivalently in the pricing errors) across the distribution of the returns. Both assumptions are wrong, as illustrated by Figure 10.

Figure 11. Variable Selection Graph



Note: The figure shows the decision three that an hypothetical market participant would face, if she was to decide on what new variable to include in her model, at the top of the figure the investors starts with our baseline model (which we take as granted), in each subsequent node, from top to down, the investor would select a new variable to add to the model, but the decision would depend on the weighting scheme that she uses, which reflects the interest of that particular investor in the different fragments of the cross-sectional distribution of the returns. The tree is not hypothetical but it was constructed using the Avg. R^2 that we report in Equation 8 with five different weighting schemes $w_1 = [0.5,0.3,0.15,0.05,0]$, $w_2 = [0.35,0.25,0.2,0.15,0.05]$, $w_3 = [0.2,0.2,0.2,0.2,0.2]$, $w_4 = [0.05,0.15,0.2,0.25,0.35]$ and $w_5 = [0,0.05,0.15,0.3,0.5]$. Moving from w_1 towards w_5 reflects a decreasing interest in the left tail of the distribution, and a correspondent increasing interest in the right tail of the distribution.

In the limiting case in which we end up adding all possible variables to the model (210 in our case), the reasoning sketched in this subsection would be mainly irrelevant. But in real life the number of variables in the model does matter. For instance, Molavi et al. (2021) has recently explained the way in which model complexity found in real life applications, measured by the maximum number of factors that agents can add to their pricing models, is likely an explanation of the returns' predictability patterns found in asset returns. This fact emphasizes the importance of considering what to explain in the cross-section of the returns, which do not have to necessarily be the cross-sectional mean.

6. Conclusions

We propose a new way to analyze cross-sectional market returns using traditional characteristics. Our model is able to generate quantile factors series that can be interpreted as time-varying effects of traditional characteristics on the various cross-sectional quantiles of stock returns, or alternatively, as underlying time-varying systemic factors that determine the joint dynamics of the full conditional distribution of asset returns.

Factor series vary greatly across time, which highlights the necessity of considering time-varying factors loads explicitly in the asset-pricing model formulation. Moreover, the joint fit of the model also varies in time. Indeed, as always occurs with asset pricing factors, the time-series volatility of the factor dominates their time-series averages. The correlations within the quantiles-factors of a given characteristic are low in absolute value. The lowest correlation across the two tails (quantiles at $\theta = 0.05$ and 0.95) are recorded for investment and liquidity (-0.2), and the highest for beta (0.3), which witnesses to the important drawback of only considering factors that explain the center of the distribution, as it is done by the extant literature on empirical asset pricing.

Not all characteristics explain the cross-section at the same level alongside its various quantiles. Indeed, the winners' tail of the distribution is better explained by factors such as investment or liquidity, the losers' tail by the book-to-market ratio, and the center of the distribution by momentum or profitability. In general, even factors such as beta and size that exert a significant effect on the whole cross-sectional distribution of the returns, present opposite signs in the two tails. Beta's effect is positive at the winners' tail and negative at the losers' tail, while the opposite holds for size. In general, the tails of the distribution are better explained than the center by the traditional asset pricing equation that we studied, which points out to the existence of additional statistical power of the factors in the model at the tails of the distribution compared to the median.

Our results are robust to expanding our sample size as early as February 1928, since we have reliable information in our database to assess the quantile effects of beta, size, liquidity and momentum. We confirm in this longer sample that beta and size exert opposite effects on the left and right tails of the cross-section, while momentum impacts it in a stable fashion. Liquidity in this longer sample also exerts a different effect on winners and losers, which is different than for the shorter sample. Indeed, the effect of an increase in illiquidity is negative on the left of the market and positive on the right tail, but it is greater in magnitude on the right tail, thus confirming liquidity as a right tail factor. In short, liquidity is a very particular factor, the larger the quantile of the cross-section the longer the right tail of the liquidity effects. That liquidity is a nonlinear factor can be rationalized by theoretical studies on the limits of arbitrage, that point out to a nonlinear relationship between liquidity and prices, for instance due to binding risk-taking constraints, or to the surpassing of certain market volatility thresholds. Nevertheless, the asymmetric impact of liquidity, which is especially large for the tail of the winners, calls for a closer examination for instance of the relationship between central bank liquidity and individual stock market liquidity. It could be the case that a greater provision of liquidity impacts the strong stocks in the market in a greater magnitude to what it affects the non-winners' stocks.

This analysis is only possible because by construction our model does not rely on a balanced panel structure, which would reduce the sample to a point of making unfeasible the estimations. We also examine the significance of the quantile effects under the light of different economic states. In particular, we compare the t-statistics at the two tails and the center of the distribution, before, during and after the Great Recession in the US and we document relatively stable patterns in terms of significance in time, which might have changed during the crisis, but after the crisis came back to the previous levels.

These results are of evident interest for portfolio managers, who usually decide to diversify their portfolios according to some 'smart beta' strategy based on characteristics. Our insight is that depending on the portfolio, whether it tends to generate returns closer to the average or to the tails, the inclusion of a new stock based on characteristics must be informed by the ability of such characteristic to diversify risk at the particular fragment of the distribution that the manager cares more about (usually the left tail).

There is an important issue derived from our analysis that we approach from a conservative perspective. If the effect of characteristics is different alongside the various fragments of the cross-section distribution of the returns, pricing metrics based on alphas, or model residuals constructed targeting the mean in their loss function, are not entirely adequate to measure the performance of pricing models. The implications of this fact are important because potential investors could be more interested in the left tail of the distribution than in the center, if for instance they are particularly risk-averse. Conversely, more aggressive investors would be interested in explaining better the right tail of the distribution than the center. From this perspective, the information on the level of risk-aversion of practitioners, or in general on the interest of the modelers, which is ignored by traditional alpha-tests, becomes fundamental. We propose to complement traditional pricing-error statistics with the quantile-R^2, and average global quantile-R^2 calculated using different schemes to weight across quantiles. We show empirically that indeed heterogeneities in the quantile dimension matter when selecting the optimal pricing model. In short, not all factors serve all purposes. More research is needed in this respect, especially to aggregate the local adjustment metrics that we report into a global measure, easier to use and interpret, but also seeking to incorporate in the quantile dimensions the recent progresses made in the field to perform effective (and unbiased) intensive search for factors, but considering the whole cross-sectional of the returns and not only the average cases. This same reasoning that we have developed for the traditional 7-factor model in the literature could be expanded even to test the robustness to such a traditional approach.

References

Adrian, T., Boyarchenko, N., Giannone, D. (2019) Vulnerable Growth. *The American Economic Review* 109: 1263-1289.

Adrian, T., Bruneermeier, M. (2016) CoVaR. *The American Economic Review* 106: 1705-1741.

Amengual, D., Sentana, E. (2018) Is a Normal Copula the Right Copula? *Journal of Business & Economic Statistics* 38: 350-366.

Amihud, Y. (2002) Illiquidity and Stock Returns: Cross-Section and Time-Series Effects. *Journal of Financial Markets* 5: 31-56.

Amihud, Y., Mendelson, H. (1986) Asset Pricing and the Bid-Ask Spread. *Journal of Financial Economics* 17: 223-249.

Ando, T., Bai, J. (2019) Quantile Co-Movement in Financial Markets: A Panel Quantile Model with Unobserved Heterogeneity. *Journal of the American Statistical Association* 115: 266-279.

Andreini P., Izzo, C., Ricco, G. (2020) Deep Dynamic Factor Models. Working Paper Available at arXiv:2007.11887v1 [econ.EM]. [Accessed 21 Feb 2021].

Avramov, D., Chordia, T. (2006). Asset Pricing Models and Financial Market Anomalies. *The Review of Financial Studies* 19: 1001–1040.

Back, K., Kapadia, N., Ostdiek, B. (2013). Slopes as Factors: Characteristic Pure Plays. Capital Markets: *Asset Pricing & Valuation eJournal*. DOI: 10.2139/ssrn.2295993. [Accessed 21 Feb 2021].

Balakrishnan, K., Bartov, E. Faurel, L. (2010) Post Loss/Profit Announcement Drift. *Journal of Accounting and Economics* 50: 20–41.

Bali, T., Engle, R., Murray, S. (2016) *Empirical asset pricing: The cross section of stock returns*. John Wiley & Sons.

Ball, R., Gerakos, J., Linnainmaa, J. T., Nikolaev, V. (2016) Accruals, Cash Flows, and Operating Profitability in the Cross Section of Stock Returns. *The Journal of Financial Economics* 121: 28-45.

Bansal, R., Yaron, A. (2004) Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles. *The Journal of Finance* 59: 1481–1509.

Barillas, F., Shanken, J. (2018) Comparing Asset Pricing Models. *The Journal of Finance* 73: 715- 754.

Belloni, A., Chen, M., Madrid-Padilla, O.H., Wang, Z. (2019) High Dimensional Latent Panel Quantile Regression with an Application to Asset Pricing. Working Paper Harvard University Available at https://j.mp/2OL3CbO [Accessed 21 Feb 2021].

Bradshaw M., Richardson, S., Sloan, R. (2006) The Relation Between Corporate Financing Activities, Analysts' Forecasts and Stock Returns. *Journal of Accounting and Economics* 42: 53–85.

Campbell, J.Y. (2018) *Financial Decisions and Markets: A Course in Asset Pricing*. Princeton University Press, Princeton New Jersey.

Campbell, J.Y., Cochrane, J.H. (1999) By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior. *Journal of Political Economy* 107: 205–251.

Chamberlain G., Rothschild, M. (1986) Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets. *Econometrica* 51: 1281-1304.

Chan, L., Jegadeesh, N., Lakonishok, J. (1996) Momentum Strategies. *The Journal of Finance* 51: 1681-1713.

Chen, A.Y., Zimmermann, T. (2020) Open Source Cross-Sectional Asset Pricing. Working paper Available at SSRN: https://ssrn.com/abstract=3604626 [Accessed 21 Feb 2021].

Chen, L., Dolado, J.J., Gonzalo, J. (2021) Quantile Factor Models. *Econometrica* Forthcoming Papers 89(1). Working paper available at arXiv:1911.02173v2 [econ.EM]. [Accessed 21 Feb 2021].

Chen, L., Pelger, M., Zhu, J. (2019) Deep Learning in Asset Pricing. Working Paper Available at arXiv:1904.00745 [q-fin.ST]. [Accessed 21 Feb 2021].

Chib, S., Zeng, X., Zhao, L. (2020) Comparing Asset Pricing Models. *The Journal of Finance* 75: 551- 577

Cohen, R. B., C. Polk, Vuolteenaho, T. (2003) The Value Spread. *The Journal of Finance* 78: 609–642.

Connor G., Korajczyk, R. (1988) A Risk and Return in an Equilibrium APT: Application of a New Test Methodology. *Journal of Financial Economics* 21: 255–289.

Connor, G., Hagmann, M., Linton, O. (2012). Efficient Semiparametric Estimation of The Fama–French Model and Extensions. *Econometrica* 80: 713–754.

Cooper, M.J, Gulen, H., Schill, M.J. (2008) Asset Growth and The Cross-Section of Stock Returns. *The Journal of Finance* 63: 1609– 1651.

Fama, E. (1976). *Foundations of Finance.* Basic Books: New York.

Fama, E., French, K. (1993) Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33: 3–56.

Fama, E., French, K. (2006). Profitability, Investment and Average Returns. *Journal of Financial Economics* 82: 491–518.

Fama, E., French, K. (2015) A Five-Factor Asset Pricing Model. *Journal of Financial Economics* 116: 1-22.

Fama, E., French, K. (2019) Comparing Cross-Section and Time-Series Factor Models. *The Review of Financial Studies* 33: 1891-6.

Fama, E., French, K. (2021) The Value Premium. *The Review of Asset Pricing Studies* 11: 105-121.

Fama, E., MacBeth, J. (1973) Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy* 81: 607–36.

Fan, J., Liao, Y., Wang, W. (2016). Projected Principal Component Analysis in Factor Models. *The Annals of Statistics* 44: 219–254.

Feng, G., Giglio, S., Xiu, D. (2020) Taming the Factor Zoo: A Test of New Factors. *The Journal of Finance* 75: 1327-1370.

Feng, G., He, J., Polson, N. (2018) Deep Learning for Predicting Asset Returns. Working Paper Available at arXiv:1804.09314 [stat.ML]. [Accessed 21 Feb 2021].

Giglio, S., Liao, Y., Xiu, D. (2020) Thousands of Alpha Tests. *The Review of Financial Studies* https://doi.org/10.1093/rfs/hhaa111.

Giglio, S., Xiu, D. (2019), Asset Pricing with Omitted Factors. Chicago Booth Research Paper No. 16-21, Available at SSRN: https://ssrn.com/abstract=2865922 . [Accessed 21 Feb 2021].

Gu, S., Kelly, B., Xiu, D. (2020) Autoencoder Asset Pricing Models. *Journal of Econometrics* https://doi.org/10.1016/j.jeconom.2020.07.009.

Gu, S., Kelly, B., Xiu, D. (2020) Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies* 33: 2223-2273.

Hartzmark, S. Solomon, D. (2013) The Dividend Month Premium. *Journal of Financial Economics* 109: 640-660.

He, Z., Krishnamurthy, A. (2013) Intermediary Asset Pricing. *American Economic Review* 103: 732–70.

Heston, S. L Sadka, R (2008) Seasonality in the Cross-Section of Stock Returns. *Journal of Financial Economics* 87: 418–445.

Hirshleifer, D., Hou, K Hong-Teoh, S., Zhang, Y. (2004) Do Investors Overvalue Firms With Bloated Balance Sheets? Finance 0412001, University Library of Munich, Germany. [Accessed 21 Feb 2021].

Hou, K. (2007) Industry Information Diffusion and the Lead-lag Effect in Stock Returns. *The Review of Financial Studies* 20: 1113–1138.

Jegadeesh, N., Livnat, J. (2006) Revenue Surprises and Returns. *Journal of Accounting Economics* 41: 147-171

Kelly, B., Malamud, S., Pedersen, L.H. (2020) Principal portfolios. Working paper 27388, NBER. Available at http://www.nber.org/papers/w27388. [Accessed 21 Feb 2021].

Kelly, B., Pruitt, S., Su, Y. (2019) Characteristics are Covariances: A Unified Model of Risk and Return. *The Journal of Financial Economics* 134: 501-524.

Koenker, R., Bassett, J. (1978) Regression Quantiles, *Econometrica* 46: 33-50.

Koenker, R., Machado, J. (1999) Goodness of Fit and Related Inference Processes for Quantile Regression. *Journal of the American Statistical Association* 94: 1296-1310.

Koenker, R.W. (2005). *Quantile Regression*. Cambridge University Press, Cambridge.

Lettau, M., Pelger, M. (2020) Factors that Fit the Time Series and Cross-Section of Stock Returns. *The Review of Financial Studies* 33: 2274-2325.

Linnainmaa, J. T., Roberts, M. R. (2018) The History of the Cross-Section of Stock Returns. *The Review of Financial Studies* 31: 2606-2649.

Loh, R. Warachka, M. (2012) Streaks in Earnings Surprises and the Cross-Section of Stock Returns. *Management Science* 58: 1305-1321.

Loughran, T., Wellman, J. (2011) New Evidence on the Relation Between the Enterprise Multiple and Average Stock Returns. *Journal of Financial and Quantitative Analysis* 46: 1629–1650.

Lyandres, E., Sun, L., Zhang, L. (2008) The New Issues Puzzle: Testing the Investment-Based Explanation. *Review of Financial Studies* 21: 2825-2855.

Ma, S., Linton, O., Gao, J. (2020) Estimation and Inference in Semiparametric Quantile Factor Models. *Journal of Econometrics* Available at https://doi.org/10.1016/j.jeconom.2020.07.003.

McLean, R. D., Pontiff, J. (2016) Does Academic Research Destroy Stock Return Predictability? *The Journal of Finance* 71: 5–31.

Molavi, P., Tahbaz-Salehi, A., Vedolin, A. (2021) Model Complexity, Expectations, and Asset Prices. Working paper 28408, NBER. Available at http://www.nber.org/papers/w28408. [Accessed 21 Feb 2021].

Novy-Marx, R. (2013). The Other Side of Value: the Gross Profitability Premium. *The Journal of Financial Economics* 108: 1–28.

Richardson, S., Sloan, R., Soliman, M., Tuna, I. (2005) Accrual Reliability, Earnings Persistence and Stock Prices. *Journal of Accounting and Economics* 39: 437-485.

Uribe, J.M., Guillen M. (2020) *Quantile Regression for Cross-Section and Time Series Data: An Application to Energy Markets Using* R. Springer, Germany.

Appendix

Table A1.

Candidate Variables

Authors	Name (Acronym)	Description	t-stat. (linear)
Bradshaw et al. (2006)	Net debt financing (NetDebtFinance)	Long-term debt issuance minus long-term debt reduction minus current debt changes, scaled by the average total assets in years t-1 and t.	8.44
Chan et al. (1996)	Earnings forecast revisions (REV6)	REV6 is the sum of revisions from months t-6 to t. Revisions are the change in the mean earnings estimate for the next quarter from month t-1 to t, scaled by stock price in month t-1.	6.98
Chan et al. (1996)	Earnings announcement return (AnnouncementReturn)	Sum of the firm return from one day before earnings announcement to 2 days after the announcement.	12.98
Cooper et al. (2008)	Asset Growth (AssetGrowth)	Annual growth rate of total assets (at)	7.80
Hartzmark and Salomon (2013)	Dividends (DivInd)	Dummy variable equal to 1 if return with dividends is greater than return without dividends 11 months ago or 2 months ago, and 0 otherwise	6.06
Hirshleifer et al. (2004)	Net Operating Assets (NOA)	Difference between operating assets and operating liabilities, scaled by lagged total assets.	7.71
Hou (2007)	Industry return of big firms (IndRetBig)	Average monthly return of the 30% largest companies by market value of equity in Fama-French 48 industry portfolios. The 30% largest companies were excluded.	9.52

Jegadeesh and Livnat (2006)	Revenue Surprise (RevenueSurprise)	Yearly change in revenue per share minus the average yearly change in revenue per share, over the previous two years, scaled by the standard deviation.	7.23
Loh and Warachka (2012)	Number of consecutive earnings increases (NumEarnIncrease)	Number of four-quarter net income increases over the previous 2 years.	14.24
Loughran and Wellman (2011)	Enterprise Multiple (EntMult)	Market value of equity plus long-term debt, plus debt in current liabilities, plus deferred charges, minus cash and short-term investment, divided by operating income.	6.36
Lyandres et al. (2008)	Change in property, plants and equipment, and investment assets, (InvestPPEInv)	One-year change in property, plants and equipment plus one-year change in inventory, scaled by one-year lagged assets.	9.19
Richardson et al. (2005)	Change in financial liabilities (DelFINL)	Financial liabilities and current liabilities minus preferred stock, between years t-1 and t, scaled by average total assets in years t-1 and t.	11.88
Richardson et al. (2005)	Change in net financial assets (DelNetFin)	Short-term investments plus investments and advances minus the sum of long-term debt, debt in current liabilities and preferred stock capital. All Divided by the difference between the current and one-year lagged sum by total assets.	9.09
Heston and Sadka (2008)	Return seasonality years 11 to 15 (MomSeasAlt16to20a)	Returns seasonality years 16 to 20	5.07
Bali et al. (2016)	Skewness of daily returns (ReturnSkew)	Skewness of daily returns over previous month.	6.40

Note: The table shows the candidate variables to be added to our pricing equation (Eq. 6), the proponent authors, description and linear t-statistic. All information was taken from CZ (2020), where more information details may be retrieved.

UBIREA
Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics
Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona

WEBSITE: www.ub.edu/irea/ • **CONTACT:** irea@ub.edu