Received: 20 September 2022 | Accepted: 4 June 2024

DOI: 10.1111/jfir.12429

The Journal of Financial Research

ORIGINAL ARTICLE

Time-Series Factor Modeling and Selection

Michael Michaelides

Department of Finance, Accounting and Management Science, Cyprus University of Technology, Limassol, Cyprus

Correspondence
Michael Michaelides, Cyprus University of Technology, Department of Finance, Accounting and Management Science, Limassol, Cyprus.
Email: michael.michaelidis@cut.ac.cy

Abstract

The article proposes a statistical time-series factor model that incorporates deterministic orthogonal trend polynomials. Such polynomials allow capturing variation in returns without initially identifying a set of robust time-series factors. This modeling approach can serve as a coherent basis for testing and selecting the most relevant factors among a set of possible ones. Additionally, it can help identify whether any factors are missing from a time-series asset pricing model. The use of the proposed model and empirical strategy is illustrated by two empirical applications from the literature, yielding results related to the Fama-French five-factor model and the factor zoo.

JEL CLASSIFICATION
C52, C58, G12

1 | INTRODUCTION

Following the pioneering study of Harvey, Liu, and Zhu (2016), there has been a consistent emergence of meta-analyses within the asset pricing literature. This growing body of literature aims to assess the reliability of the hundreds of existing factors, commonly referred to as the "zoo of factors" (Cochrane, 2011), that appear to explain the cross-sectional and time-series variation in stock returns. The general consensus is that the "zoo" comprises numerous spurious factors. For instance, Feng, Giglio, and Xiu (2020) and Bryzgalova, Huang, and Julliard (2023) found that only a small number of factors are robust explanators of returns. Within the literature, several explanations exist for why many existing factors are deemed spurious, including data dredging (Lo & MacKinlay, 1990; Harvey et al., 2016; Harvey, 2017; Linnainmaa & Roberts, 2018), inappropriate use of significance levels (Kim & Ji, 2015; Harvey et al., 2016; Michaelides, 2021), and model misspecification.

This article relates to the vast literature on factor inference under model misspecification. In a series of papers, Kan and Zhang (1999a, 1999b), Kan and Zhou (1999), Kan and Robotti (2008, 2009), Kan, Robotti, and Shanken (2013), and Gospodinov, Kan, and Robotti (2013, 2014, 2017, 2019) have shown that many conventional inference methods suffer from model misspecification, which distorts the model parameters and leads to erroneously identifying spurious factors as relevant. These studies, along with others such as Shanken and Zhou (2007), Kleibergen (2009), Burnside (2016), Kelly, Pruitt, and Su (2019), Chib, Zeng, and Zhao (2020), Feng et al. (2020), Freyberger, Neuhierl, and Weber (2020), Haddad, Kozak, and Santosh (2020), Lettau and Pelger (2020), Giglio and Xiu (2021), and Bryzgalova et al. (2023), provide inference methods that are robust to different types of misspecification.

The present study adds to this literature by proposing a fully parametric statistical model designed to account for all the systematic information exhibited by returns. The primary concern of the article is the potential bias that arises in time-series tests of asset pricing models from the departures of statistical assumptions imposed on the data. Statistical model misspecification is likely to give rise to unreliable inferences because any departures from the model assumptions can lead to significant discrepancies between the assumed and actual error probabilities of statistical hypotheses. For example, when a 0.05 significance level is used to test for the statistical significance of a new factor, but the actual Type I error probability is closer to 0.9, the inference results are likely to be erroneous. This may lead to a spurious factor being identified as significant.

Numerous studies have shown that stock returns exhibit a non-normal/heteroskedastic distribution,[1] along with departures from the assumptions of independence and identical distribution (i.i.d.), such as the presence of temporal dependence and several forms of nonstationarity.[2] Further, studies have shown that factor returns exhibit similar departures from normality and i.i.d. For instance, Chib and Zeng (2020) assumed the factor distribution to be the Student's t and showed that it is better supported than the normal distribution, while the best model under normality is quite different from the best-supported Student's t model. In a similar context, Chernov, Lochstoer, and Lundeby (2018) observed that the returns of many prominent factors exhibit surprising deviations from i.i.d.

Despite the concerns about departures from normality and i.i.d., it is common practice in much of the empirical asset pricing literature to assume that returns and regression residuals satisfy these assumptions (Linton, 2019 (p.290)).[3] When these standard assumptions are not used, empirical tests are usually constructed using generic methods that relax some or all of the assumptions.[4] To the best of my knowledge, no parametric time-series factor model exists in the literature that aims to describe all the relevant return features that could be accounted for. In the spirit of Hendry and Richard (1983, 1990) and Spanos (1990), this article proposes a statistical time-series factor model aimed at explicitly capturing all systematic information exhibited by returns. The proposed model assumes that returns are Student's t distributed, rendering the model heteroskedastic. Moreover, the model explicitly accounts for the non-i.i.d. features exhibited by returns through the incorporation of deterministic and lagged variables. The inclusion of these variables makes the model both time-dependent and dynamic.

The distinctive characteristic of the proposed model lies primarily in its modeling approach to explicitly capture variation in returns using deterministic orthogonal trend polynomials. As discussed in the empirical strategy section, this modeling approach allows for the explicit capture of variation in returns without directly identifying a set of robust time-series factors at the outset. This can serve as a coherent basis for testing and selecting the most relevant factors from a set of possible ones, as well as for indicating whether any factors are missing from a time-series asset pricing model. The general notion of the proposed empirical strategy is to first account for all the variation in returns using

[1]The normal distribution is the only one within the elliptically symmetric family of distributions (Fang, Kotz, & Ng, 1990) characterized by homoskedasticity; see Kelker (1970). Consequently, non-normality and heteroskedasticity are intertwined.

[2]Andreou, Pittis, and Spanos (2001) and Koundouri, Kourogenis, and Pittis (2016) provide excellent surveys regarding the empirical regularities exhibited by stock returns.

[3]Examples include the Fama and MacBeth (1973) two-pass regression, the GRS statistic of Gibbons, Ross, and Shanken (1989), and the Bayesian asset-pricing test of Barillas and Shanken (2018).

[4]A widely adopted generic method is the Generalized Method of Moments (GMM) framework, which relaxes the assumptions of distribution and temporal independence while assuming stationarity of returns (Hansen, 1982).

deterministic variables rather than observed factors, and then attempt to replace the deterministic variables with observed factors. To my knowledge, this study represents the first to propose such an empirical strategy.

Unlike recent developments in the literature (see, for example, Bryzgalova et al., 2023), the proposed model is not designed for handling high-dimensional inference generated by the factor zoo. Instead, it is intended to be used as a traditional low-dimensional time-series factor regression. Thus, as is typically the case with time-series asset pricing models (see, for example, Fama & French, 1993, 2015), the proposed model requires that a set of candidate factors be fully observable through previously established knowledge of the cross-section of average returns. Furthermore, while the emphasis of this article is on time-series factor modeling, the general version of the model could be applied more broadly in other empirical finance applications, provided that the data at hand exhibit similar features as returns.

To illustrate the use of the proposed model and empirical strategy, the article presents two empirical applications. The first application revisits the Fama and French (2015) five-factor model with the aim of evaluating the size, value, profitability, and investment factors in explaining the variation in average stock returns. The results indicate that, although all of the Fama-French factors explain variation in returns left unexplained by the Capital Asset Pricing Model (CAPM), the investment factor, *CMA*, becomes redundant when the value factor, *HML*, is also included in a time-series asset pricing model. Moreover, the results suggest that the Fama-French factors do not fully capture the variation in returns modeled at the outset using deterministic orthogonal trend polynomials. This implies the potential absence of relevant factors from the Fama-French five-factor model, at least for the particular data used.

The second empirical application expands upon the first by assessing 51 prominent factors proposed in prior literature. The results suggest that the vast majority of factors in the zoo are either useless or redundant. Among the factors considered, only the quality-minus-junk factor, *QMJ*, introduced by Asness, Frazzini, and Pedersen (2019), appears to capture strong variation in average returns that is left unexplained by the Fama-French factors. Nevertheless, it seems that there are still missing relevant factors from a time-series asset pricing model that combines the Fama-French and *QMJ* factors.

2 | THE STATISTICAL MODEL

The proposed statistical model is a parametric linear regression designed to explicitly capture all systematic information exhibited by returns. The model assumes a Student's t distribution for returns, rendering it heteroskedastic. Additionally, the model explicitly accounts for the non-i.i.d. features in returns by incorporating deterministic and lagged variables. Deterministic variables are used to account for any nonstationarity in the mean (including seasonal) and variance of returns. Lagged variables are used to account for any temporal dependence in returns. The inclusion of deterministic and lagged variables makes the model time-dependent and dynamic, respectively. To reflect its underlying distribution and characteristics of incorporating time-dependent and dynamic elements, the model is henceforth referred to as the "time-heterogeneous Student's t Dynamic Linear Regression (DLR) model".

2.1 | Time-series factor model

Since the primary focus of this article is on time-series factor modeling, the statistical model presented below has been adapted for this specific purpose. When tailored for time-series factor modeling, the time-heterogeneous Student's t DLR model[5] takes the following form:

[5]Appendix A provides further details on the general version of the statistical model, including its degrees of freedom and first two conditional moments. For additional insights into the basic structure of the model, refer to Heracleous and Spanos (2006).

$$R_{it} = \alpha_i + \boldsymbol{\beta}_i^\top \mathbf{f}_t + \sum_{j=1}^{m^*} \gamma_{1ij} d_{1jt} + \sum_{j=1}^{s-1} \gamma_{2ij} d_{2jt} + \sum_{j=1}^{p} \boldsymbol{\delta}_{ij}^\top \mathbf{Z}_{it-j} + \epsilon_{it},$$
$$\text{for } \epsilon_{it} \sim \text{St}(0, h_i^2(t); v + l^*), i = 1, 2, \ldots, k, t = 1, 2, \ldots, n, \quad (1)$$

where R_{it} is the return (or excess return) of stock or portfolio i for period t, α_i is the intercept[6] for stock or portfolio i, $\mathbf{f}_t := (f_{1t}, f_{2t}, \ldots, f_{lt})^\top$ is a set of l common factors at period t with factor loadings of $\boldsymbol{\beta}_i := (\beta_{1i}, \beta_{2i}, \ldots, \beta_{li})^\top$ for each stock or portfolio i, d_{1jt} denotes deterministic orthogonal trend polynomials[7] of degree $j = 1, 2, \ldots, m^*$, d_{2jt} denotes deterministic seasonal dummy variables for season $j = 1, \ldots, s - 1$, $\mathbf{Z}_{it} := (R_{it}, \mathbf{f}_t^\top)^\top$, and $\mathbf{Z}_{it-j}$ is a set of lagged variables of order $j = 1, 2, \ldots, p$. The error term, ϵ_{it}, is distributed $\text{St}(0, h_i^2(t); v + l^*)$, where $v + l^*$ denotes the degrees of freedom of the conditional Student's t distribution.

2.2 | Why a time-heterogeneous factor model?

In practice, it is common to work with portfolios rather than individual stocks since averaging returns is believed to address the problem of nonstationarity associated with individual stocks, thus improving the precision of inferences. However, even when using averages, the nonstationarity in the mean and variance of returns is not entirely eliminated.

Going back at least to Hagerman (1978), one concern associated with using averages is that much of the mean nonstationarity may be "hidden". Consider Figure 1, which plots the returns of Fama and French's (2015) extreme small size–low book-to-market equity ratio (B/M) portfolio. The red line illustrates the constant mean calculated under identical distribution (i.d.). In contrast, the blue line illustrates the non-constant trending-seasonal mean calculated after accounting for the hidden mean nonstationarity. Not accounting for mean nonstationarity may lead to deviations between the red and blue lines appearing in the residuals of a time-series factor model, potentially resulting in unreliable inferences and the identification of spurious factors as statistically significant.

More importantly, using averages does very little to mitigate variance nonstationarity. As evident from Figure 1, portfolio returns exhibit volatility clustering, characterized by extended periods of high volatility followed by extended periods of low volatility, and vice versa. These extended periods of high and low volatility also vary over time due to the differing magnitudes of influence that financial shocks have on volatility; this can be easily seen comparing the picture inside the frame in Figure 1. In the context of time-series factor modeling, volatility clustering can easily lead to the variance of the model residuals to vary not only with the factors included in the model (heteroskedasticity), but also with the time index (a departure from i.d.); the two are distinct sources of statistical model misspecification that must be accounted for separately. Failure to properly account for variance nonstationarity can almost certainly result in misleading inferences and the identification of spurious factors as statistically significant.

3 | EMPIRICAL STRATEGY

The proposed time-heterogeneous Student's t DLR model is not confined solely to enhancing the reliability of inference. The modeling approach to explicitly capture variation in returns using deterministic orthogonal trend polynomials is a distinctive characteristic of the model because it allows capturing all the variations in returns

[6]It should be pointed out that the intercept of the model is not the traditional Jensen's alpha (Jensen, 1968) because the model incorporates deterministic and lagged variables. The appropriate Jensen's alpha can be calculated by adding back to the estimated intercept the means of the deterministic and lagged variables; for additional information, see Appendix B.

[7]Several different types of orthogonal trend polynomials may be used to model the mean and variance nonstationarity in returns. The most potent among them are the Gram-Schmidt orthonormal polynomials, which had been shown to offer the flexibility to extend them up to higher degrees without giving rise to collinearity problems (Michaelides & Spanos, 2020); for additional information, see Appendix C. The Gram-Schmidt orthonormal polynomials are used later in the empirical applications.

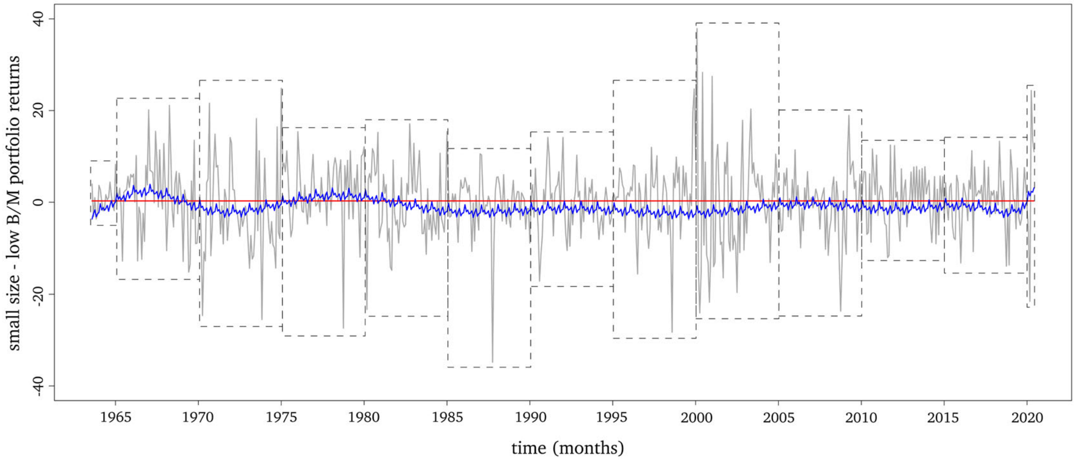

FIGURE 1 Small size-low B/M portfolio returns; July 1963 to June 2020, 684 months. The figure plots the returns of Fama and French's (2015) extreme small size-low B/M portfolio. The red line illustrates the constant mean calculated under the assumption that portfolio returns are i.d. over time. The blue line illustrates the non-constant trending-seasonal mean calculated after accounting for the (hidden) mean nonstationarity exhibited by portfolio returns. The frames illustrate a volatility-adjusted five-year rolling window. The deviations between the red and blue lines, and the drastic change of the picture inside the frame indicate the presence of mean and variance nonstationarity, respectively. This figure is used to illustrate features of portfolio returns that would suggest departures from nonstationarity. The figure is used primarily for illustrative purposes. It may not be representative of all portfolios. [Color figure can be viewed at wileyonlinelibrary.com]

without directly identifying a set of robust factors at the outset. This modeling approach can serve as a coherent basis for testing and selecting the most relevant factors from a set of possible ones, as well as for indicating whether any factors are missing from a time-series asset pricing model. A three-step empirical strategy is outlined below.

3.1 | Steps of the empirical strategy

Step I: CAPM estimation. In the first step of the empirical strategy, the CAPM is estimated using the time-heterogeneous Student's t DLR model in (1). This estimation takes the following form:

$$\begin{aligned} R_{if,t} &= \alpha_i + \beta_i R_{mf,t} + \\ &+ \sum_{j=1}^{m^*} \gamma_{1ij} d_{1jt} + \sum_{j=1}^{s-1} \gamma_{2ij} d_{2jt} + \sum_{j=1}^{p} \delta_{1ij} R_{if,t-j} + \sum_{j=1}^{p} \delta_{2ij} R_{mf,t-j} + \epsilon_{it}, \\ &\text{for } \epsilon_{it} \sim \text{St}(0, h_i^2(t); v + l^*), i = 1, 2, \ldots, k, t = 1, 2, \ldots, n, \end{aligned} \tag{2}$$

where $R_{if,t} = (R_{it} - R_{ft})$ is the excess return of stock or portfolio i for period t, $R_{mf,t} = (R_{mt} - R_{ft})$ is the excess return of the market portfolio for period t − for R_{ft} denoting the risk-free interest rate for period t −, d_{1jt} are deterministic orthogonal trend polynomials of degree $j = 1, 2, \ldots, m^*$, d_{2jt} are deterministic seasonal dummy variables for season $j = 1, \ldots, s - 1$, and $R_{if,t-j}$ and $R_{mf,t-j}$ are the lagged variables of $R_{if,t}$ and $R_{mf,t}$, respectively, of order $j = 1, 2, \ldots, p$.

Step II: Factor testing and selection. In the second step of the empirical strategy, the CAPM estimation from step I is reestimated by incorporating the candidate factors (whatever those might be) and their lags into the model. This step aims to test for the significance of the candidate factors by posing the general question of omitted relevant variables. The statistical significance of the candidate factors is considered trustworthy since the estimation in step I accounts for all the systematic information exhibited by the observed stock or portfolio i.

Step III: Identification of missing factors. In the third step of the empirical strategy, the objective is to assess whether the candidate factors effectively capture all the variation in returns exhibited by stock or portfolio *i*. If the set of candidate factors from step II fully explains the variation in returns of stock or portfolio *i*, the deterministic orthogonal trend polynomials included in the CAPM estimation from step I should become statistically insignificant. Conversely, if any of the polynomials (with degrees less than or equal to m^*) remain statistically significant, that would indicate potential missing relevant factors, as the variation in returns captured by these polynomials is not fully explained by the considered set of candidate factors.

4 | EMPIRICAL APPLICATION: FAMA-FRENCH FIVE-FACTOR MODEL

This section applies the proposed empirical strategy on the Fama and French (2015) five-factor model. The factors used in this empirical application are the market risk premium (*MKT* or $R_m - R_f$, used interchangeably henceforth), size (*SMB*), value (*HML*), profitability (*RMW*), and investment (*CMA*). The left-hand-side portfolios are the 25 *Size-B/M*, 25 *Size-OP*, and 25 *Size-Inv* value-weighted portfolios.[8] The sample spans a 684-month period, from July 1963 to June 2020.

4.1 | Empirical strategy

For this empirical application, the empirical strategy discussed in Section 3 can be summarized as follows.

Step I. Estimate the CAPM using the time-heterogeneous Student's *t* DLR model.

Step II. Test for the significance of the Fama-French factors by reestimating the model from step I that incorporates the factors.

Step III. Evaluate whether the Fama-French factors capture all the variation in returns by observing the change in statistical significance of the deterministic orthogonal trend polynomials from step I to step II.

4.2 | Portfolio non-i.i.d. features

The left-hand-side portfolios do not uniformly exhibit identical non-i.i.d. features. First, the smaller size, lower B/M, lower operating profitability (OP), and higher investment (Inv) portfolios display more intricate forms of mean and variance nonstationarity compared to the bigger size, higher B/M, higher OP, and lower Inv portfolios; for simplicity we could refer to the former portfolios as "distressed" and the latter as "undistressed".[9] Second, the average January returns of the smaller size, higher B/M, lower OP, and lower Inv portfolios are larger relative to those of the bigger size, lower B/M, higher OP, and higher Inv portfolios. Third, smaller size portfolios appear to exhibit positive first-order temporal dependence, whereas the bigger size portfolios exhibit no temporal dependence at all.[10]

[8]The factors and portfolios are obtained from Kenneth French's online data library. Details regarding the factors, portfolios, and their construction can be found in the online data library of Kenneth French and in Fama and French (2015).

[9]The difference in mean nonstationarity might be attributed to distressed portfolios, on average, encompassing a larger number of stocks compared to undistressed portfolios; the averaging of this larger number of stocks into portfolios is likely to generate more complicated forms of mean nonstationarity. The difference in variance nonstationarity possibly arises because distressed portfolios are generally more volatile than undistressed portfolios, leading to more complicated forms of variance nonstationarity.

[10]The dissimilarity in temporal dependence might be attributed to the fact that stocks included in smaller size portfolios are generally less liquid and traded less frequently than their bigger size counterparts.

4.3 | Incorporation of deterministic and lagged variables

As the left-hand-side portfolios do not uniformly exhibit identical non-i.i.d. features, the incorporation of deterministic and lagged variables for each set of 25 portfolios is based on the most "extreme" portfolio in terms of non-i.i.d. features. In step I of the empirical strategy, the CAPM estimation includes the following deterministic and lagged variables: (i) Gram-Schmidt orthonormal trend polynomials of degree 12 for the estimations of the 25 *Size-B/M* and 25 *Size-OP* portfolios, and of degree 13 for the estimations of the 25 *Size-Inv* portfolios; (ii) monthly dummy variables (excluding January); and (iii) lagged variables of order one.[11]

4.4 | Empirical results

This subsection presents the estimation results from the time-heterogeneous Student's *t* DLR model. The following time-series asset pricing models are considered: CAPM that includes only *MKT*; Fama-French three-factor model that combines *MKT*, *SMB*, and *HML*; three four-factor models that combine *MKT*, *SMB*, and pairs of *HML*, *RMW*, and *CMA*; and Fama-French five-factor model that combines *MKT*, *SMB*, *HML*, *RMW*, and *CMA*. The dependent variables in each set of regressions are the monthly excess returns on the 25 *Size-B/M*, 25 *Size-OP*, and 25 *Size-Inv* portfolios.

Tables 1–3 report the estimated coefficients and associated *t*-statistics for *HML*, *RMW*, and *CMA*. The estimation results for *MKT*, *SMB*, and lagged variables are not presented. The regression coefficients of *MKT* are always close to 1, and the regression coefficients of *SMB* are always decreasing monotonically from strongly positive for small size portfolios to slightly negative for big size portfolios – size effect (see Fama & French, 2015). *MKT* and *SMB* are always highly statistically significant with large *t*-statistics. The lagged variables of order one of the excess returns of the portfolios and *MKT* are statistically significant only for the smaller size portfolios. The lagged variables of *SMB*, *HML*, *RMW*, and *CMA* were not statistically significant with no exceptions; they were dropped completely from the models for simplicity.[12] The discussion below is primarily related to the statistical significance of the factors; their estimated coefficients are reported but not discussed extensively since their interpretation is beyond the scope of this article.

Table 4 reports the estimated Jensen's alphas for the various time-series asset pricing models. These Jensen's alphas are used to compare the relative performance among the competing models considered. Table 5 reports the binary statistical significance of the Gram-Schmidt orthonormal trend polynomials incorporated into the Fama-French five-factor estimations. The significance of these polynomials provides the basis for identifying if any relevant factors are missing from the Fama-French five-factor model.

4.4.1 | 25 Size-B/M portfolios

Table 1 presents the estimated coefficients and *t*-statistics of *HML*, *RMW*, and *CMA* for the three four-factor models that combine *MKT*, *SMB*, and pairs of *HML*, *RMW*, and *CMA*, as well as for the Fama-French five-factor model that combines all factors. The dependent variables in each set of 25 regressions are the monthly excess returns on the 25 *Size-B/M* portfolios. The primary coefficients and *t*-statistics of interest in this set of regressions are those of *RMW* and *CMA*. The *HML* slopes are highly statistically significant and increasing monotonically from strongly

[11]Concerning the degrees of freedom parameter of the Student's *t* distribution, ν, it is set equal to 1 in this empirical application. Hence, the degrees of freedom of the conditional Student's *t* distribution ($\nu + l^*$) for the CAPM and Fama-French five-factor models are 4 and 12, respectively, while for the models in between, they vary accordingly.

[12]As mentioned in the previous subsection, smaller size portfolios exhibit positive first-order temporal dependence. For robustness purposes, lagged variables up to order three were included in different specifications of the estimated models. The empirical results (not presented) are not affected when additional lags are added to the models.

TABLE 1 Coefficients and t-statistics from the time-heterogeneous Student's t DLR for the 25 Size-B/M portfolios.

Size\B/M	Low	2	3	4	High	Low	2	3	4	High
Panel A: *MKT,SMB,HML,RMW*										
	h_i					$t(h_i)$				
Small	−0.43	−0.10	0.13	0.30	0.52	−16.90	−4.75	7.75	18.82	30.80
2	−0.54	−0.05	0.22	0.44	0.65	−28.77	−2.93	13.94	27.61	38.37
3	−0.53	−0.01	0.28	0.47	0.64	−27.31	−0.84	15.93	27.30	29.54
4	−0.40	0.04	0.30	0.45	0.71	−21.97	2.07	15.27	22.35	28.19
Big	−0.34	0.03	0.24	0.58	0.85	−24.85	1.57	11.99	31.86	29.90
	r_i					$t(r_i)$				
Small	−0.39	−0.28	−0.10	−0.03	0.03	−10.47	−9.42	−4.23	−1.41	1.04
2	−0.18	−0.02	0.09	0.05	0.02	−6.49	−0.64	4.06	2.29	0.83
3	−0.06	0.06	0.11	0.02	0.03	−2.10	2.35	4.38	0.80	0.97
4	−0.03	−0.01	0.03	−0.03	−0.06	−1.09	−0.48	1.07	−1.13	−1.47
Big	0.21	0.04	−0.06	−0.02	−0.19	11.13	1.48	−2.10	−0.68	−4.34
Panel B: *MKT,SMB,HML,CMA*										
	h_i					$t(h_i)$				
Small	−0.40	−0.06	0.13	0.27	0.44	−11.74	−2.28	5.96	13.80	20.89
2	−0.46	−0.02	0.22	0.41	0.62	−18.53	−1.16	11.10	20.53	29.11
3	−0.42	−0.01	0.27	0.47	0.62	−17.65	−0.37	11.67	20.84	22.48
4	−0.34	0.00	0.29	0.45	0.74	−15.25	0.20	11.57	17.49	22.94
Big	−0.33	−0.03	0.18	0.59	1.05	−18.52	−1.29	7.25	25.16	28.90
	c_i					$t(c_i)$				
Small	0.07	0.04	0.05	0.06	0.16	1.52	0.98	1.54	2.23	5.51
2	−0.08	−0.03	−0.04	0.05	0.04	−2.32	−1.05	−1.36	1.76	1.29
3	−0.21	−0.02	0.00	0.03	0.07	−6.29	−0.53	0.08	1.04	1.76
4	−0.12	0.12	0.03	0.05	−0.01	−3.87	3.66	0.99	1.38	−0.29
Big	−0.09	0.13	0.18	0.03	−0.35	−3.48	4.33	5.21	0.86	−7.01
Panel C: *MKT,SMB,RMW,CMA*										
	r_i					$t(r_i)$				
Small	−0.37	−0.28	−0.10	−0.04	−0.00	−8.63	−8.77	−3.98	−1.76	−0.14
2	−0.16	−0.02	0.06	0.01	−0.04	−4.93	−0.83	2.28	0.50	−1.10
3	−0.06	0.06	0.08	−0.03	−0.02	−1.85	2.12	2.76	−1.07	−0.39
4	−0.03	0.01	0.01	−0.07	−0.13	−0.88	0.51	0.18	−2.06	−2.71
Big	0.25	0.05	−0.04	−0.08	−0.36	10.86	1.98	−1.43	−2.20	−6.19

TABLE 1 (Continued)

Size\B/M	Low	2	3	4	High	Low	2	3	4	High
	c_i					$t(c_i)$				
Small	−0.41	−0.11	0.14	0.30	0.59	−10.45	−3.90	5.82	12.67	20.77
2	−0.57	−0.07	0.19	0.44	0.63	−18.21	−3.00	7.90	16.46	19.12
3	−0.63	−0.03	0.27	0.46	0.62	−20.38	−1.27	9.94	16.04	16.03
4	−0.46	0.10	0.28	0.44	0.64	−16.06	3.86	9.43	13.40	14.71
Big	−0.34	0.11	0.32	0.54	0.53	−15.58	4.27	11.09	16.60	10.09
Panel D: *MKT,SMB,HML,RMW,CMA*										
	h_i					$t(h_i)$				
Small	−0.41	−0.08	0.11	0.27	0.45	−12.83	−3.16	5.41	13.66	21.35
2	−0.48	−0.04	0.23	0.42	0.62	−20.08	−1.96	11.50	20.73	29.03
3	−0.43	−0.01	0.27	0.46	0.62	−17.96	−0.41	11.94	20.66	22.50
4	−0.35	−0.01	0.28	0.43	0.73	−15.39	−0.36	11.39	17.07	22.71
Big	−0.32	−0.03	0.17	0.58	1.03	−18.41	−1.55	6.86	25.09	29.20
	r_i					$t(r_i)$				
Small	−0.40	−0.29	−0.09	−0.02	0.05	−10.31	−9.28	−3.65	−0.71	2.06
2	−0.20	−0.01	0.09	0.07	0.03	−7.13	−0.60	3.71	2.96	1.25
3	−0.10	0.06	0.12	0.02	0.05	−3.66	2.16	4.37	0.73	1.51
4	−0.06	0.01	0.03	−0.03	−0.06	−2.26	0.30	1.12	−0.93	−1.46
Big	0.22	0.05	−0.03	−0.02	−0.27	11.01	1.96	−0.97	−0.81	−6.23
	c_i					$t(c_i)$				
Small	−0.04	−0.04	0.03	0.07	0.18	−0.86	−1.07	1.04	2.39	6.05
2	−0.13	−0.03	−0.02	0.06	0.05	−3.78	−1.01	−0.68	2.19	1.77
3	−0.24	−0.02	0.02	0.04	0.06	−6.92	−0.49	0.75	1.28	1.61
4	−0.14	0.11	0.03	0.05	−0.03	−4.35	3.31	0.89	1.24	−0.75
Big	−0.05	0.14	0.17	0.00	−0.42	−1.83	4.42	4.82	0.10	−8.22

Note: The table presents the coefficients and t-statistics for the 25 *Size-B/M* portfolios from the time-heterogeneous Student's t DLR model; July 1963 to June 2020, 684 months. The right-hand side of the regressions includes the Gram-Schmidt orthonormal trend polynomials, the monthly dummy variables for the months of February through December, *MKT*, *SMB*, *HML*, *RMW*, *CMA*, and the lagged variables of the respective excess portfolio return and *MKT*. Panel A reports estimation results of a four-factor model that combines *MKT*, *SMB*, *HML*, and *RMW*. Panel B reports estimation results of a four-factor model that combines *MKT*, *SMB*, *HML*, and *CMA*. Panel C reports estimation results of a four-factor model that combines *MKT*, *SMB*, *RMW*, and *CMA*. Panel D reports estimation results of the Fama-French five-factor model that combines *MKT*, *SMB*, *HML*, *RMW*, and *CMA*. The first column of the table reports the estimated regression coefficients. The second column reports their associated t-statistics.

negative for low B/M portfolios to strongly positive for high B/M portfolios – value effect (see Fama & French, 2015). This is not surprising given that these portfolios are formed on B/M ratios.

The slopes of *RMW* are statistically significant for about half of the portfolios, regardless of what other factors are included on the right-hand side of the regressions. For the smaller size portfolios, the *RMW* slopes are increasing

TABLE 2 Coefficients and t-statistics from the time-heterogeneous Student's t DLR for the 25 Size-OP portfolios.

Size\OP	Low	2	3	4	High	Low	2	3	4	High
Panel A: *MKT,SMB,HML,RMW*										
	h_i					$t(h_i)$				
Small	−0.05	0.22	0.25	0.22	0.13	−2.60	12.49	14.01	11.35	6.30
2	−0.18	0.15	0.18	0.17	0.06	−9.30	9.38	12.12	10.09	3.12
3	−0.15	0.14	0.17	0.13	0.01	−6.63	7.74	10.22	7.63	0.41
4	−0.01	0.15	0.15	0.11	−0.01	−0.56	8.39	8.11	6.18	−0.58
Big	0.11	0.15	0.11	0.00	−0.12	5.03	9.12	6.21	0.04	−9.11
	r_i					$t(r_i)$				
Small	−0.52	0.01	0.11	0.24	0.25	−17.66	0.52	4.22	8.44	8.23
2	−0.70	0.01	0.09	0.32	0.45	−25.34	0.45	4.24	12.47	16.38
3	−0.76	−0.08	0.13	0.28	0.46	−22.76	−2.85	5.29	10.80	15.41
4	−0.74	−0.26	0.06	0.20	0.31	−20.25	−9.82	2.36	7.77	11.06
Big	−0.85	−0.37	−0.09	0.14	0.39	−25.97	−14.93	−3.71	7.65	19.33
Panel B: *MKT,SMB,HML,CMA*										
	h_i					$t(h_i)$				
Small	−0.07	0.20	0.24	0.20	0.10	−2.44	9.03	10.28	7.58	3.82
2	−0.16	0.13	0.21	0.20	0.05	−5.26	6.29	10.82	8.40	1.80
3	−0.07	0.19	0.17	0.10	0.03	−2.11	7.86	8.13	4.43	0.95
4	0.05	0.19	0.15	0.08	0.02	1.36	7.72	6.50	3.49	0.73
Big	0.22	0.15	0.08	−0.03	−0.14	6.10	6.52	3.74	−2.00	−6.93
	c_i					$t(c_i)$				
Small	0.23	0.03	0.00	−0.04	−0.04	5.80	1.13	0.04	−1.02	−1.12
2	0.24	0.07	−0.06	−0.16	−0.14	5.62	2.44	−2.35	−4.66	−3.84
3	0.13	−0.03	−0.05	−0.03	−0.20	2.67	−0.90	−1.58	−1.05	−5.16
4	0.13	0.04	−0.01	−0.00	−0.14	2.42	1.18	−0.21	−0.14	−3.88
Big	0.07	0.16	0.08	0.01	−0.11	1.39	4.78	2.47	0.26	−4.06
Panel C: *MKT,SMB,RMW,CMA*										
	r_i					$t(r_i)$				
Small	−0.49	−0.00	0.10	0.23	0.25	−16.18	−0.11	3.48	7.61	7.94
2	−0.67	0.01	0.06	0.30	0.45	−22.69	0.29	2.47	10.84	15.65
3	−0.77	−0.11	0.09	0.28	0.43	−22.18	−3.73	3.61	10.46	14.19
4	−0.77	−0.29	0.04	0.19	0.29	−20.35	−10.08	1.44	7.22	10.24
Big	−0.88	−0.35	−0.09	0.16	0.39	−25.90	−13.51	−3.44	8.45	18.58

TABLE 2 (Continued)

Size\OP	Low	2	3	4	High	Low	2	3	4	High
	c_i					$t(c_i)$				
Small	0.02	0.22	0.24	0.22	0.15	0.79	8.51	9.29	7.96	4.88
2	−0.12	0.18	0.14	0.12	0.03	−4.39	7.31	6.07	4.50	1.06
3	−0.18	0.11	0.13	0.13	−0.06	−5.57	4.04	5.54	4.93	−2.00
4	−0.05	0.13	0.13	0.12	−0.06	−1.32	4.70	4.74	4.81	−2.29
Big	0.02	0.19	0.14	0.03	−0.14	0.57	7.84	5.69	1.93	−7.11
Panel D: *MKT,SMB,HML,RMW,CMA*										
	h_i					$t(h_i)$				
Small	−0.10	0.20	0.24	0.20	0.12	−4.07	8.97	10.45	8.31	4.63
2	−0.21	0.12	0.20	0.21	0.08	−8.54	5.93	10.71	9.73	3.21
3	−0.12	0.17	0.17	0.12	0.05	−4.34	7.24	8.27	5.41	1.98
4	0.01	0.17	0.14	0.09	0.03	0.32	7.20	6.26	3.98	1.08
Big	0.18	0.12	0.08	−0.02	−0.10	6.24	5.76	3.67	−1.49	−6.04
	r_i					$t(r_i)$				
Small	−0.50	0.02	0.12	0.24	0.26	−16.87	0.84	4.49	8.42	8.50
2	−0.69	0.02	0.08	0.31	0.45	−24.18	1.00	3.75	12.12	15.96
3	−0.79	−0.09	0.11	0.28	0.43	−23.13	−3.09	4.60	10.99	14.41
4	−0.76	−0.27	0.06	0.20	0.29	−20.30	−9.76	2.08	7.81	10.44
Big	−0.87	−0.35	−0.08	0.16	0.38	−26.46	−13.96	−3.19	8.40	18.78
	c_i					$t(c_i)$				
Small	0.11	0.04	0.03	0.04	0.03	3.16	1.22	1.04	1.06	0.82
2	0.06	0.07	−0.05	−0.09	−0.04	1.90	2.17	−1.72	−2.87	−1.20
3	−0.06	−0.05	−0.02	0.02	−0.10	−1.46	−1.54	−0.76	0.66	−2.70
4	−0.06	−0.03	0.00	0.04	−0.08	−1.30	−0.75	0.07	1.40	−2.26
Big	−0.15	0.07	0.05	0.05	−0.04	−3.62	2.35	1.78	2.41	−1.69

Note: The table presents the coefficients and t-statistics for the 25 *Size-OP* portfolios from the time-heterogeneous Student's t DLR model; July 1963 to June 2020, 684 months. The right-hand side of the regressions includes the Gram-Schmidt orthonormal trend polynomials, the monthly dummy variables for the months of February through December, *MKT*, *SMB*, *HML*, *RMW*, *CMA*, and the lagged variables of the respective excess portfolio return and *MKT*. Panel A reports estimation results of a four-factor model that combines *MKT*, *SMB*, *HML*, and *RMW*. Panel B reports estimation results of a four-factor model that combines *MKT*, *SMB*, *HML*, and *CMA*. Panel C reports estimation results of a four-factor model that combines *MKT*, *SMB*, *RMW*, and *CMA*. Panel D reports estimation results of the Fama-French five-factor model that combines *MKT*, *SMB*, *HML*, *RMW*, and *CMA*. The first column of the table reports the estimated regression coefficients. The second column reports their associated t-statistics.

monotonically from strongly negative for low B/M portfolios to slightly positive for high B/M portfolios. For the big size portfolios, however, the effect is reversed; the *RMW* slopes are decreasing monotonically from strongly positive for low B/M portfolios to strongly negative for high B/M portfolios. These slopes survive, no matter what is included on the right-hand side of the regressions.

TABLE 3 Coefficients and t-statistics from the time-heterogeneous Student's t DLR for the 25 Size-Inv portfolios.

Size\Inv	Low	2	3	4	High	Low	2	3	4	High
Panel A: *MKT,SMB,HML,RMW*										
	h_i					$t(h_i)$				
Small	0.12	0.28	0.25	0.15	−0.17	5.81	16.30	14.38	8.32	−8.73
2	0.21	0.29	0.25	0.16	−0.30	11.14	17.10	15.10	10.28	−18.10
3	0.22	0.29	0.21	0.11	−0.30	9.67	15.61	12.78	6.28	−15.13
4	0.28	0.29	0.21	0.07	−0.28	11.98	15.86	12.78	3.75	−12.61
Big	0.21	0.21	0.12	−0.05	−0.35	8.29	12.75	8.89	−3.58	−17.46
	r_i					$t(r_i)$				
Small	−0.39	−0.02	0.06	0.00	−0.23	−12.56	−0.82	2.35	0.17	−8.30
2	−0.28	0.04	0.07	0.18	−0.15	−9.78	1.60	2.93	7.57	−6.23
3	−0.15	−0.00	0.13	0.16	−0.04	−4.35	−0.11	5.34	6.35	−1.42
4	−0.18	−0.03	0.04	0.08	−0.07	−5.28	−1.00	1.63	3.02	−2.22
Big	−0.15	−0.06	0.07	0.15	0.09	−3.91	−2.62	3.50	7.23	2.96
Panel B: *MKT,SMB,HML,CMA*										
	h_i					$t(h_i)$				
Small	−0.00	0.20	0.20	0.14	−0.04	−0.11	9.14	9.14	6.05	−1.75
2	0.02	0.23	0.20	0.21	−0.12	0.83	10.41	9.59	10.52	−5.76
3	0.04	0.18	0.18	0.19	−0.07	1.56	7.53	8.19	8.43	−3.20
4	0.07	0.19	0.20	0.12	−0.07	2.49	8.36	9.42	5.49	−2.67
Big	−0.13	−0.02	0.09	0.04	−0.09	−4.77	−1.19	5.32	2.57	−4.13
	c_i					$t(c_i)$				
Small	0.43	0.19	0.07	0.02	−0.16	11.32	6.29	2.37	0.71	−4.71
2	0.51	0.15	0.10	−0.16	−0.32	15.76	4.96	3.32	−5.70	−11.07
3	0.44	0.25	0.05	−0.21	−0.45	11.07	7.47	1.50	−6.78	−14.01
4	0.55	0.23	0.05	−0.14	−0.43	14.43	7.37	1.55	−4.37	−11.32
Big	0.79	0.53	0.05	−0.25	−0.60	21.70	20.61	2.01	−10.38	−18.76
Panel C: *MKT,SMB,RMW,CMA*										
	r_i					$t(r_i)$				
Small	−0.33	−0.01	0.07	−0.00	−0.26	−10.90	−0.52	2.67	−0.15	−9.02
2	−0.21	0.05	0.08	0.14	−0.19	−7.77	1.75	3.04	5.37	−8.18
3	−0.09	0.02	0.12	0.12	−0.12	−2.81	0.74	4.76	4.37	−4.59
4	−0.10	−0.02	0.03	0.04	−0.15	−3.14	−0.59	1.10	1.45	−4.72
Big	0.00	0.04	0.08	0.10	0.01	0.04	1.65	3.65	4.86	0.21

TABLE 3 (Continued)

Size\Inv	Low	2	3	4	High	Low	2	3	4	High
	c_i					$t(c_i)$				
Small	0.33	0.36	0.28	0.15	−0.28	11.79	14.71	10.90	5.69	−10.26
2	0.47	0.37	0.30	0.08	−0.50	19.11	14.71	12.44	3.23	−22.05
3	0.46	0.42	0.24	−0.02	−0.57	14.61	15.25	9.87	−0.72	−22.10
4	0.55	0.39	0.22	−0.02	−0.55	18.11	15.09	9.08	−0.91	−18.30
Big	0.67	0.52	0.16	−0.19	−0.69	22.46	25.11	8.10	−9.56	−26.98
Panel D: *MKT,SMB,HML,RMW,CMA*										
	h_i					$t(h_i)$				
Small	−0.02	0.20	0.20	0.14	−0.07	−0.98	9.20	9.36	6.01	−2.70
2	0.01	0.22	0.20	0.21	−0.14	0.56	10.32	9.70	10.72	−7.14
3	0.04	0.18	0.19	0.19	−0.08	1.46	7.54	8.77	8.83	−3.74
4	0.05	0.19	0.20	0.12	−0.08	1.69	8.26	9.49	5.54	−2.99
Big	−0.14	−0.02	0.10	0.05	−0.09	−5.43	−1.17	5.64	3.07	−3.83
	r_i					$t(r_i)$				
Small	−0.34	0.01	0.08	0.00	−0.27	−11.28	0.38	3.26	0.10	−9.53
2	−0.20	0.08	0.09	0.16	−0.21	−7.49	3.08	3.80	6.71	−9.17
3	−0.09	0.04	0.14	0.12	−0.13	−2.60	1.34	5.55	4.83	−4.86
4	−0.09	0.00	0.04	0.05	−0.15	−2.85	0.11	1.67	1.88	−4.88
Big	−0.00	0.03	0.09	0.11	0.00	−0.03	1.23	4.16	5.54	0.18
	c_i					$t(c_i)$				
Small	0.35	0.19	0.10	0.02	−0.22	9.62	6.23	3.19	0.67	−6.54
2	0.46	0.17	0.11	−0.12	−0.37	14.33	5.42	3.76	−4.17	−13.10
3	0.42	0.26	0.07	−0.19	−0.49	10.33	7.54	2.33	−6.15	−14.94
4	0.52	0.23	0.04	−0.13	−0.47	13.36	6.94	1.46	−4.07	−12.25
Big	0.80	0.54	0.06	−0.23	−0.60	21.29	20.54	2.55	−9.59	−18.37

Note: The table presents the coefficients and t-statistics for the 25 *Size-Inv* portfolios from the time-heterogeneous Student's t DLR model; July 1963 to June 2020, 684 months. The right-hand side of the regressions includes the Gram-Schmidt orthonormal trend polynomials, the monthly dummy variables for the months of February through December, *MKT*, *SMB*, *HML*, *RMW*, *CMA*, and the lagged variables of the respective excess portfolio return and *MKT*. Panel A reports estimation results of a four-factor model that combines *MKT*, *SMB*, *HML*, and *RMW*. Panel B reports estimation results of a four-factor model that combines *MKT*, *SMB*, *HML*, and *CMA*. Panel C reports estimation results of a four-factor model that combines *MKT*, *SMB*, *RMW*, and *CMA*. Panel D reports estimation results of the Fama-French five-factor model that combines *MKT*, *SMB*, *HML*, *RMW*, and *CMA*. The first column of the table reports the estimated regression coefficients. The second column reports their associated t-statistics.

The slopes and statistical significance of *CMA* are not similar across different models. In regressions that include both *CMA* and *HML* (Panels B and D), the slopes of *CMA* are statistically significant for about one-third of the portfolios. *CMA* exhibits high statistical significance only for the four-factor model that excludes *HML* (Panel C). In this case, the *CMA* slopes reveal the value effect, indicating that the average *HML* return is captured by its exposure to *CMA*.

TABLE 4 Jensen's alphas from the time-heterogeneous Student's *t* DLR for the 25 Size-B/M, 25 Size-OP, and 25 Size-Inv portfolios.

Panel A: 25 *Size-B/M*

Size\B/M	α_i^J Low	2	3	4	High	α_i^J Low	2	3	4	High
	1. *MKT*					2. *MKT,SMB,HML*				
Small	−0.82	−0.20	−0.05	0.14	0.31	−0.54	−0.09	−0.05	0.12	0.14
2	−0.38	−0.10	0.17	0.21	0.24	−0.21	−0.02	0.09	0.08	−0.03
3	−0.23	0.08	0.07	0.27	0.25	−0.07	0.10	−0.01	0.10	−0.01
4	−0.13	−0.07	0.09	0.19	0.17	0.02	−0.09	−0.02	0.06	−0.11
Big	−0.00	0.01	0.10	0.06	0.07	0.14	−0.03	0.04	−0.16	−0.19
	3. *MKT,SMB,HML,RMW*					4. *MKT,SMB,HML,CMA*				
Small	−0.40	0.01	−0.01	0.13	0.14	−0.54	−0.08	−0.07	0.11	0.12
2	−0.15	−0.01	0.06	0.05	−0.04	−0.19	−0.01	0.11	0.06	−0.03
3	−0.04	0.07	−0.06	0.05	−0.04	−0.03	0.10	−0.01	0.07	−0.01
4	0.04	−0.09	−0.06	0.06	−0.11	0.05	−0.09	−0.03	0.06	−0.11
Big	0.06	−0.06	0.04	−0.17	−0.13	0.16	−0.05	−0.01	−0.17	−0.14
	5. *MKT,SMB,RMW,CMA*					6. *MKT,SMB,HML,RMW,CMA*				
Small	−0.43	0.04	−0.02	0.15	0.17	−0.39	0.03	−0.03	0.12	0.11
2	−0.15	−0.02	0.10	0.07	0.05	−0.11	−0.01	0.07	0.03	−0.05
3	−0.04	0.07	−0.03	0.09	0.03	0.02	0.08	−0.06	0.04	−0.04
4	0.05	−0.11	−0.03	0.10	0.01	0.08	−0.10	−0.05	0.05	−0.10
Big	0.05	−0.09	0.01	−0.10	0.05	0.07	−0.08	−0.00	−0.18	−0.05

Panel B: 25 *Size-OP*

Size\OP	α_i^J Low	2	3	4	High	α_i^J Low	2	3	4	High
	1. *MKT*					2. *MKT,SMB,HML*				
Small	−0.45	0.09	0.08	0.10	−0.12	−0.37	0.06	0.02	0.11	−0.08
2	−0.35	−0.03	0.03	−0.02	0.11	−0.25	−0.06	0.04	−0.03	0.14
3	−0.21	0.08	0.10	0.04	0.14	−0.20	0.00	0.01	0.04	0.14
4	−0.16	−0.01	0.05	0.04	0.06	−0.18	−0.07	0.01	0.02	0.09
Big	−0.21	−0.07	0.03	0.08	0.09	−0.27	−0.15	−0.02	0.08	0.20
	3. *MKT,SMB,HML,RMW*					4. *MKT,SMB,HML,CMA*				
Small	−0.20	0.07	−0.03	0.02	−0.18	−0.39	0.06	0.01	0.11	−0.07
2	−0.01	−0.07	−0.00	−0.15	−0.03	−0.26	−0.06	0.04	−0.03	0.17
3	0.06	0.00	−0.05	−0.05	−0.02	−0.20	0.00	0.02	0.03	0.18
4	0.09	0.01	−0.03	−0.05	−0.02	−0.16	−0.05	−0.00	0.02	0.10
Big	0.05	−0.02	0.01	0.01	0.05	−0.25	−0.17	−0.02	0.07	0.21

TABLE 4 (Continued)

Panel B: 25 *Size-OP*

Size\OP	α_i^J Low	2	3	4	High	α_i^J Low	2	3	4	High
	5. *MKT,SMB,RMW,CMA*					6. *MKT,SMB,HML,RMW,CMA*				
Small	−0.22	0.09	0.00	0.03	−0.16	−0.22	0.07	−0.03	0.02	−0.17
2	−0.03	−0.05	0.01	−0.12	−0.03	−0.01	−0.08	0.00	−0.14	−0.01
3	0.05	0.03	−0.02	−0.05	0.02	0.08	0.01	−0.04	−0.07	0.02
4	0.10	0.03	−0.01	−0.05	−0.00	0.10	0.03	−0.03	−0.05	−0.01
Big	0.09	−0.03	0.00	0.01	0.05	0.08	−0.04	0.00	−0.00	0.06

Panel C: 25 *Size-Inv*

Size\Inv	α_i^J Low	2	3	4	High	α_i^J Low	2	3	4	High
	1. *MKT*					2. *MKT,SMB,HML*				
Small	−0.05	0.32	0.17	0.10	−0.64	−0.06	0.17	0.12	0.03	−0.50
2	0.05	0.17	0.17	0.17	−0.42	−0.05	0.11	0.09	0.11	−0.23
3	0.16	0.32	0.18	0.16	−0.25	0.04	0.15	0.10	0.10	−0.15
4	0.10	0.13	0.06	0.05	−0.21	−0.06	−0.02	−0.01	0.05	−0.06
Big	0.20	0.06	0.06	0.01	−0.17	0.04	−0.01	0.02	0.09	0.06
	3. *MKT,SMB,HML,RMW*					4. *MKT,SMB,HML,CMA*				
Small	0.08	0.17	0.11	0.02	−0.40	−0.10	0.13	0.11	0.02	−0.47
2	0.01	0.10	0.07	0.05	−0.19	−0.11	0.08	0.09	0.13	−0.20
3	0.09	0.14	0.04	0.03	−0.14	−0.01	0.12	0.09	0.11	−0.10
4	0.01	−0.02	−0.03	0.01	−0.04	−0.12	−0.03	−0.01	0.06	−0.01
Big	0.08	−0.00	−0.01	0.03	0.04	−0.07	−0.08	−0.00	0.11	0.14
	5. *MKT,SMB,RMW,CMA*					6. *MKT,SMB,HML,RMW,CMA*				
Small	0.02	0.16	0.12	0.03	−0.36	0.03	0.14	0.10	0.03	−0.36
2	−0.07	0.08	0.07	0.09	−0.15	−0.06	0.06	0.05	0.07	−0.14
3	0.02	0.12	0.05	0.07	−0.06	0.03	0.10	0.03	0.05	−0.05
4	−0.07	−0.02	−0.02	0.04	0.05	−0.07	−0.04	−0.03	0.03	0.05
Big	−0.06	−0.10	−0.02	0.07	0.12	−0.06	−0.10	−0.03	0.06	0.14

Note: The table presents Jensen's alphas for the 25 *Size-B/M*, 25 *Size-OP*, and 25 *Size-Inv* portfolios from the time-heterogeneous Student's *t* DLR model; July 1963 to June 2020, 684 months. The right-hand side of the regressions includes the Gram-Schmidt orthonormal trend polynomials, the monthly dummy variables for the months of February through December, *MKT*, *SMB*, *HML*, *RMW*, *CMA*, and the lagged variables of the respective excess portfolio return and *MKT*. Panel A reports Jensen's alphas for the 25 *Size-B/M* portfolios. Panel B reports Jensen's alphas for the 25 *Size-OP* portfolios. Panel C reports Jensen's alphas for the 25 *Size-Inv* portfolios. Each panel reports Jensen's alphas for the CAPM that includes only *MKT*; the Fama-French three-factor model that combines *MKT*, *SMB*, and *HML*; the three four-factor models that combine *MKT*, *SMB*, and pairs of *HML*, *RMW*, and *CMA*; and the Fama-French five-factor model that combines *MKT*, *SMB*, *HML*, *RMW*, and *CMA*. The reported Jensen's alphas are calculated by adding back the means of the associated deterministic and lagged variables to the estimated intercept of each regression (see Appendix B).

The Journal of Financial Research

TABLE 5 Binary statistical significance of Gram-Schmidt orthonormal trend polynomials from the time-heterogeneous Student's *t* DLR for the 25 Size-B/M, 25 Size-OP, and 25 Size-Inv portfolios.

	Significance at the 5% level					Significance at the 1% level					Significance at the 0.1% level				
Panel A: 25 *Size-B/M*															
Size\B/M	**Low**	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**
Small	√	√	√	√	√	√	√	√	√	√	√	√	√	√	
2	√	√	√	√		√	√	√			√	√	√		
3	√	√	√	√	√	√	√	√	√		√	√	√	√	
4	√	√	√	√	√	√	√	√	√	√	√	√		√	
Big	√	√	√	√	√	√	√	√	√		√				
Panel B: 25 *Size-OP*															
Size\OP	**Low**	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**
Small	√	√	√	√	√	√	√	√	√	√	√	√	√	√	
2	√	√	√	√	√	√	√	√	√	√	√	√	√	√	
3	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
4	√	√	√	√	√	√	√	√	√		√	√			
Big	√	√	√	√	√	√	√		√	√	√			√	
Panel C: 25 *Size-Inv*															
Size\Inv	**Low**	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**
Small	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
2	√	√	√	√	√		√	√	√	√			√	√	√
3	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√

TABLE 5 (Continued)

Panel C: 25 *Size-Inv*															
Size\Inv	Low	2	3	4	High	Low	2	3	4	High	Low	2	3	4	High
4	√	√	√	√	√			√		√			√		√
Big	√	√	√	√	√		√	√	√	√				√	√

Note: The table presents the binary statistical significance of the Gram-Schmidt orthonormal trend polynomials for the 25 *Size-B/M*, 25 *Size-OP*, and 25 *Size-Inv* portfolios from the time-heterogeneous Student's t DLR model; July 1963 to June 2020, 684 months. The right-hand side of the regressions includes the Gram-Schmidt orthonormal trend polynomials, the monthly dummy variables for the months of February through December, *MKT*, *SMB*, *HML*, *RMW*, *CMA*, and the lagged variables of the respective excess portfolio return and *MKT*. Panel A reports the binary statistical significance of the Gram-Schmidt orthonormal trend polynomials for the 25 *Size-B/M* portfolios. Panel B reports the binary statistical significance of the Gram-Schmidt orthonormal trend polynomials for the 25 *Size-OP* portfolios. Panel C reports the binary statistical significance of the Gram-Schmidt orthonormal trend polynomials for the 25 *Size-Inv* portfolios. Each panel reports the binary statistical significance of the Gram-Schmidt orthonormal trend polynomials at the significance levels of 5%, 1%, and 0.1% for the Fama-French five-factor model that combines *MKT*, *SMB*, *HML*, *RMW*, and *CMA*. Check marks indicate statistical significance at the respective significance level. Absence of check marks denotes statistical insignificance.

4.4.2 | 25 Size-OP portfolios

Table 2 presents the estimated coefficients and *t*-statistics of *HML*, *RMW*, and *CMA* for the three four-factor models that combine *MKT*, *SMB*, and pairs of *HML*, *RMW*, and *CMA*, as well as for the Fama-French five-factor model that combines all factors. The dependent variables in each set of 25 regressions are the monthly excess returns on the 25 *Size-OP* portfolios. The primary coefficients and *t*-statistics of interest in this set of regressions are those of *HML* and *CMA*. The *RMW* slopes are highly statistically significant and increasing monotonically from strongly negative for low OP portfolios to strongly positive for high OP portfolios – profitability effect (see Fama & French, 2015). This is not surprising given that these portfolios are formed on profitability.

The slopes of *HML* are highly statistically significant for all portfolios, irrespective of the factors included on the right-hand side of the regressions. Regarding their estimated coefficients, the *HML* slopes mirror the pattern observed in the *RMW* slopes from the 25 *Size-B/M* estimations. Specifically, for the smaller size portfolios, the *HML* slopes are increasing monotonically from negative for low OP portfolios to positive for high OP portfolios, whereas for the big size portfolios, the slopes are decreasing monotonically from positive for low OP portfolios to negative for high OP portfolios. Similar to the *RMW* slopes from the 25 *Size-B/M* estimations, the slopes of *HML* survive, regardless of what is included on the right-hand side of the regressions.

The slopes of *CMA* do not reveal any consistent patterns, while their statistical significance is highly sensitive to what is included on the right-hand side of the regressions. For the four-factor model that excludes *RMW* (Panel B), the *CMA* slopes are statistically significant for about half of the portfolios, with significance mostly observed for the extreme low and high OP portfolios. On the other hand, for the four-factor model that excludes *HML* (Panel C), while the statistical significance of *CMA* is considerably improved, it is mostly evident for the medium OP portfolios. The *CMA* slopes are statistically significant for only a small number of portfolios when the regressions include all *HML*, *RMW*, and *CMA* (Panel D).

4.4.3 | 25 Size-Inv portfolios

Table 3 presents the estimated coefficients and *t*-statistics of *HML*, *RMW*, and *CMA* for the three four-factor models that combine *MKT*, *SMB*, and pairs of *HML*, *RMW*, and *CMA*, as well as for the Fama-French five-factor model that combines all factors. The dependent variables in each set of 25 regressions are the monthly excess returns on the 25 *Size-Inv* portfolios. The primary coefficients and *t*-statistics of interest in this set of regressions are those of *HML* and *RMW*. The *CMA* slopes are highly statistically significant and decreasing monotonically from strongly positive for low Inv portfolios to strongly negative for high Inv portfolios – investment effect (see Fama & French, 2015). In this instance, it comes as no surprise that the *CMA* slopes are highly statistically significant since these portfolios are formed on investment.

The slopes of *HML* and *RMW*, especially *HML*, maintain high statistical significance, regardless of what factors are included on the right-hand side of the regressions. Similar to the 25 *Size-B/M* estimations, when the regressions exclude *CMA* (Panel A), the *HML* slopes reveal the investment effect, indicating that the average *CMA* return is captured by its exposure to *HML*.

4.4.4 | Jensen's alphas

Table 4 presents Jensen's alphas for the CAPM that includes only *MKT*; the Fama-French three-factor model that combines *MKT*, *SMB*, and *HML*; the three four-factor models that combine *MKT*, *SMB*, and pairs of *HML*, *RMW*, and *CMA*; and the Fama-French five-factor model that combines all factors. The dependent variables in each set of 25 regressions are the monthly excess returns on the 25 *Size-B/M* (Panel A), 25 *Size-OP* (Panel B), and 25 *Size-Inv* (Panel C) portfolios.

The absolute Jensen's alphas of models incorporating *RMW* (subpanels 3, 5, and 6) are significantly smaller than those of models excluding *RMW* (subpanels 1, 2, and 4). This clearly indicates that *RMW* captures strong variation in returns not accounted for by other factors. Interestingly, *RMW* appears to capture variation in returns equally well for the set of regressions on the 25 *Size-B/M* and 25 *Size-Inv* portfolios (formed based on B/M ratios and investment, respectively) as it does for the set of regressions on the 25 *Size-OP* portfolios (formed based on profitability).

Moreover, the similarity in Jensen's alphas between the Fama-French three-factor model (subpanels 2) and the four-factor model that adds *CMA* (subpanels 4), as well as between the Fama-French five-factor model (subpanels 6) and the two four-factor models that combine *MKT*, *SMB*, *RMW*, and either *HML* or *CMA* (subpanels 3 and 5), suggests that the contribution of either *HML* or *CMA* in capturing variation in returns becomes marginal when the other factor is included in the regressions.

4.4.5 | HML vs. CMA redundancy

In their original U.S. study, Fama and French (2015) show that *HML* has no information about average returns that is not in other factors, leading to the conclusion that *HML* is redundant. The Jensen's alphas in Table 4 indicate that the Fama-French five-factor model has similar performance to a four-factor model that excludes either *HML* or *CMA*. This is perhaps not surprising, given the high correlation between *HML* and *CMA*; high B/M stocks tend to invest conservatively, while low B/M stocks tend to invest aggressively (see Fama & French, 2015).

However, the question remains: which of these two factors, *HML* or *CMA*, should find a place in a time-series asset pricing model? As observed in Tables 1–3, *HML* is always highly statistically significant, irrespective of what other factors are included in the regressions. Even when *CMA* is part of the regressions, including those for the 25 *Size-Inv* portfolios, the statistically significant explanatory power of *HML* does not seem to diminish. On the other hand, *CMA* exhibits high statistical significance exclusively for regressions that exclude *HML* or those of the 25 *Size-Inv* portfolios. When *HML* is included in the regressions, the statistically significant explanatory power of *CMA* appears to diminish considerably. With *HML* emerging as the dominant factor in terms of statistical significance, one may contend that it is *CMA*, not *HML*, that is redundant.

4.4.6 | Missing factors

Table 5 reports the binary statistical significance of the Gram-Schmidt orthonormal trend polynomials included in the Fama-French five-factor model that combines *MKT*, *SMB*, *HML*, *RMW*, and *CMA*. The dependent variables in each set of 25 regressions are the monthly excess returns on the 25 *Size-B/M* (Panel A), 25 *Size-OP* (Panel B), and 25 *Size-Inv* (Panel C) portfolios. In each panel, the binary statistical significance of the polynomials is reported for statistical significance of 5%, 1%, and 0.1%. Check marks (√) indicate statistical significance at the respective significance level, while the absence of check marks denotes statistical insignificance.

The Gram-Schmidt orthonormal trend polynomials, initially incorporated into the CAPM estimations to capture the variation in returns, remained statistically significant after the inclusion of the Fama-French factors into the regressions. Notably, even at a conservative significance level of 0.1%, the polynomials retained their statistical significance. The only exception is observed for the more extreme undistressed portfolios (bigger size–higher B/M, bigger size–higher OP, and bigger size–lower Inv), where the Fama-French factors appear to fully account for the variation in returns. Overall, the presence of unexplained variation in returns, captured by the polynomials but not accounted for by the Fama-French factors, suggests a potential absence of relevant factors from the Fama-French five-factor model, at least for the particular data used in this application.[13]

5 | EMPIRICAL APPLICATION: FACTOR ZOO

This empirical application expands upon the previous one by evaluating 51 prominent factors, including those from the Fama-French five-factor model and an additional 46 factors; these factors are identical to the ones used in Bryzgalova et al. (2023). Table 6 presents a detailed list of the 51 factors.[14] The factors are categorized into "accounting", "financial", "macro", "behavioral", "microstructure", and "other" following the taxonomy of Harvey and Liu (2019). The left-hand-side portfolios (25 *Size-B/M*, 25 *Size-OP*, 25 *Size-Inv*) remain consistent with those used in the previous application, preserving the same empirical strategy, as well as the same incorporation of deterministic and lagged variables. The sample spans a 519-month period, from October 1973 to December 2016.

5.1 | Empirical results

Table 7 presents estimation results from the time-heterogeneous Student's t DLR model. The following time-series asset pricing models are considered: two-factor models (F_2) that combine *MKT* and one more factor, henceforth referred to as the "additional" factor, f^*; four-factor models (F_4) that combine *MKT*, f^*, and the two "formation" factors whose characteristics are used to form the left-hand-side portfolios (for example, *SMB* and *HML* for the 25 *Size-B/M* portfolios); five-factor models (F_5) that combine *MKT*, f^*, the formation factors, and either *HML* or *RMW*, depending on which of the two is missing from the four-factor models; and six-factor models (F_6) that combine *MKT*, f^*, the formation factors, and both *HML* and *RMW*.[15] Each row in Table 7 reports the number of statistically significant portfolios corresponding to the respective additional factor, f^*, in each set of 25 regressions. For each model, the number of portfolios with t-statistics greater than two and three is reported in separate columns. The table does not report the number of statistically significant portfolios for *MKT*, the formation factors, *HML*, and *RMW*. The statistical significance of these factors does not change considerably from the previous application, while it is not sensitive to the inclusion of other factors in the regressions.

5.1.1 | Useless and redundant factors

The results presented in Table 7 suggest that the vast majority of factors are either useless or redundant. A significant number of factors, including all macroeconomic, most behavioral, some financial, and the prominent *TLIQ* and *UMD*, seem to lack statistically significant explanatory power entirely. These factors are essentially deemed

[13]In light of the aforementioned empirical results highlighting the redundancy of *CMA*, it is pertinent to also examine the binary statistical significance of the Gram-Schmidt orthonormal trend polynomials from the four-factor model that combines *MKT*, *SMB*, *HML*, and *RMW*, in order to check whether *CMA* plays a role in capturing variation in returns of the more extreme undistressed portfolios. Although not presented independently, the binary statistical significance of the polynomials from the four-factor model (excluding *CMA*) mirrors exactly that of the Fama-French five-factor model in Table 5. This consistent result further reinforces the redundancy of *CMA*.

[14]All factors are obtained from publicly accessible data sources. The data libraries from where the factors are obtained are as follows (number of factors obtained in parentheses): Robert Stambaugh (14); Kenneth French (8); Kent Daniel (7); FRED-MD (6); AQR (3); Sydney Ludvigson (3); Bureau of Economic Analysis (BEA, 2); Dashan Huang (2); Lu Zhang (2); Bureau of Labor Statistics (BLS, 1); Board of Governors of the Federal Reserve System (BGFRS, 1); Hugues Langlois (1); Asaf Manela (1). Details regarding the factors and their construction can be found in the references in Table 6, the references therein, the online data libraries from where the factors are obtained, and in Bryzgalova et al. (2023).

[15]Six-factor models are only estimated for the 25 *Size-Inv* portfolios. These are the only estimations that include *CMA*, as one of the two formation factors, since the 25 *Size-Inv* portfolios are formed on investment. Given that *CMA* was identified as a redundant factor in the previous empirical application, six-factor models are not estimated for the 25 *Size-B/M* and 25 *Size-OP* portfolios.

TABLE 6 List of factors.

Risk factor (ID)	Reference	Source
Fama-French factors		
Market return (MKT)	Sharpe (1964), Lintner (1965)	Kenneth French
Size (SMB)	Fama and French (1992; 1993)	−‖−
Value (HML)	−‖−	−‖−
Profitability (RMW)	Fama and French (2015)	−‖−
Investment (CMA)	−‖−	−‖−
Accounting		
Accruals (ACCR)	Sloan (1996)	Robert Stambaugh
Net stock issues (SISSUE)	Loughran and Ritter (1995)	−‖−
Composite equity issues (CISSUE)	Daniel and Titman (2006)	−‖−
Net operating assets (NOA)	Hirshleifer, Hou, Teoh, and Zhang (2004)	−‖−
Gross profitability (GPROF)	Novy-Marx (2013)	−‖−
Asset growth (ASSETG)	Cooper, Gulen, and Schill (2008)	−‖−
Return on assets (ROA)	Chen, Novy-Marx, and Zhang (2011)	−‖−
Investment-to-assets (INVASSETS)	Titman, Wei, and Xie (2004)	−‖−
HML Devil (HMLd)	Asness and Frazzini (2013)	AQR
Quality-minus-junk (QMJ)	Asness, Frazzini, and Pedersen (2019)	−‖−
Q investment, to-assets (qIA)	Hou, Xue, and Zhang (2015)	Lu Zhang
Q profitability, return on equity (qROE)	−‖−	−‖−
SMB hedged (SMBh)	Daniel, Mota, Rottke, and Santos (2020)	Kent Daniel
HML hedged (HMLh)	−‖−	−‖−
RMW hedged (RMWh)	−‖−	−‖−
CMA hedged (CMAh)	−‖−	−‖−
Financial		
Distress (DISTR)	Campbell, Hilscher, and Szilagyi (2008)	Robert Stambaugh
Ohlson O-score (OSCORE)	Ohlson (1980)	−‖−
Betting against beta (BAB)	Frazzini and Pedersen (2014)	AQR
Term structure (TERM)	Chan, Chen and Hsieh (1985)	FRED-MD
Credit premium (CREDIT)	−‖−	−‖−
Yield curve slope change (YC)	Ferson and Harvey (1991)	−‖−
Dividend yield on index (DY)	Campbell (1996)	−‖−
Price-earnings ratio on index (PE)	Basu (1977)	−‖−
Primary dealer capital ratio (PDCR)	He, Kelly and Manela (2017)	Asaf Manela
Systematic skewness (SSKEW)	Langlois (2020)	Hugues Langlois
MKT hedged (MKTh)	Daniel, Mota, Rottke, and Santos (2020)	Kent Daniel

(Continues)

TABLE 6 (Continued)

Risk factor (ID)	Reference	Source
Macro		
Macro uncertainty index (MUNC)	Jurado, Ludvigson and Ng (2015)	Sydney Ludvigson
Real uncertainty index (RUNC)	−‖−	−‖−
Financial uncertainty index (FUNC)	−‖−	−‖−
Industrial production growth (IP)	Chan, Chen and Hsieh (1985)	BGFRS
Oil prices growth (OIL)	Chen, Roll and Ross (1986)	BLS
Unemployment rate (URATE)	Gertler and Grinols (1982)	FRED-MD
Consumption growth, nondurables (NDCON)	Breeden, Gibbons, and Litzenberger (1989)	BEA
Consumption growth, services (SCON)	−‖−	−‖−
Behavioral		
BW sentiment index (BWSENT)	Baker and Wurgler (2006)	Dashan Huang
HJTZ PLS sentiment index (HJTZSENT)	Huang, Jiang, Tu, and Zhou (2015)	−‖−
Short-horizon behavioral (PEAD)	Daniel, Hirshleifer, and Sun (2020)	Kent Daniel
Long-horizon behavioral (FIN)	−‖−	−‖−
Microstructure		
Traded liquidity (TLIQ)	Pástor and Stambaugh (2003)	Robert Stambaugh
Nontraded liquidity (NTLIQ)	−‖−	−‖−
Other		
Momentum (UMD)	Carhart (1997)	Kenneth French
Short-term reversal (STREV)	Jegadeesh and Titman (1993)	−‖−
Long-term reversal (LTREV)	De Bondt and Thaler (1985)	−‖−
Mispricing, management (MGMT)	Stambaugh and Yuan (2017)	Robert Stambaugh
Mispricing, performance (PERF)	−‖−	−‖−

useless. On the other hand, a substantial number of factors appear to be redundant. While these factors exhibit significant explanatory power when included as additional factors in the CAPM, their statistical significance diminishes, or even vanishes, with the addition of the formation, *HML*, and/or *RMW* factors into the models. Notably, the majority of accounting and financial factors, comprising various versions of profitability and investment, along with other prominent factors such as *NTLIQ*, *STREV*, *LTREV*, and *PERF*, can be identified as redundant.

5.1.2 | Potentially useful factors

None of the additional factors under consideration remain highly statistically significant after the inclusion of the formation, *HML*, and/or *RMW* factors into the two-factor models. However, there are a few factors that could be identified as potentially useful. Specifically, *SISSUE*, *QMJ*, *OSCORE*, *SSKEW*, and *MGMT* are highly statistically significant when included as additional factors in the CAPM, while their significance does not diminish to the same extent as of other factors when the formation, *HML*, and/or *RMW* factors are added into the models. These factors retain statistical significance for at least half of all 75 portfolios with t-statistics greater than two, and for at least one third of all 75 portfolios with t-statistics greater than three. The only other additional factors meeting this "criterion" are *HML* and *RMW*. *MKT* and *SMB* also meet the criterion, but they are purely essential and formation factors, respectively. *CMA* passes the criterion only as a formation but not as an additional factor.

5.1.3 | Jensen's alphas

Similar to Table 4, Table 8 presents Jensen's alphas for the time-series asset pricing models that combine *MKT*, the formation factors, *HML*, *RMW*, and either one of the identified potentially useful factors or all of them together. The reported Jensen's alphas indicate that *QMJ* is the only one among these factors capturing strong variation in returns (subpanels 2 and 6). The absolute Jensen's alphas of models that exclude *QMJ* (subpanels 1, 3, 4, and 5) are comparable to those produced by the four- and five-factor models that included the *RMW* factor (Table 4, subpanels 3, 5, and 6). This suggests that *SISSUE*, *OSCORE*, *SSKEW*, and *MGMT* might be more redundant than useful, as they do not appear to capture strong variation in returns.

5.1.4 | Missing factors

Similar to Table 5, Table 9 reports the binary statistical significance of the Gram-Schmidt orthonormal trend polynomials included in the time-series asset pricing model that combines *MKT*, the formation factors, *HML*, *RMW*, and *QMJ*. As can be seen in Table 9, the polynomials remained statistically significant after the inclusion of *QMJ* into the regressions, suggesting that there are still missing relevant factors from a time-series asset pricing model that combines the Fama-French and *QMJ* factors. However, upon closer inspection of Table 9 and a comparison with Table 5, it becomes apparent that *QMJ* adds significant explanatory power to the Fama-French five-factor model. Specifically, *QMJ* appears to account for previously unexplained variation in returns, particularly for undistressed and medium portfolios. Nevertheless, the model does not fully capture the variation in returns of the more extreme distressed portfolios (smaller size–lower B/M, smaller size–lower OP, and smaller size–higher Inv). This observation raises the possibility that distressed portfolios may require exposure to a larger number of factors compared to undistressed portfolios.[16]

6 | CONCLUDING REMARKS

This article proposes a statistical time-series factor model aimed at explicitly capturing all systematic information exhibited by returns. The model's distinctive characteristic lies in its incorporation of deterministic orthogonal trend polynomials. In the context of time-series factor models, these polynomials offer the flexibility to capture variation in returns without initially identifying a set of robust factors. This modeling approach can serve as a basis for testing

[16]As part of a robustness check, the binary statistical significance of the Gram-Schmidt orthonormal trend polynomials was also observed for the time-series asset pricing models that combine *MKT*, the formation factors, *HML*, *RMW*, and either one of the other identified potentially useful factors or all of them together. Although not presented separately, the binary statistical significance of the polynomials in models incorporating one of the identified potentially useful factors closely resembles that of the Fama-French five-factor model in Table 5, while for the model combining all identified potentially useful factors, it is nearly identical to that reported in Table 9. This consistency in the significance of the polynomials further confirms that *SISSUE*, *OSCORE*, *SSKEW*, and *MGMT* do not capture strong variation in returns, rendering them more redundant than useful factors.

TABLE 7 Number of statistically significant portfolios from the time-heterogeneous Student's *t* DLR for the 25 Size-B/M, 25 Size-OP, and 25 Size-Inv portfolios.

Additional factor (f^*)	Panel A: 25 *Size-B/M*						Panel B: 25 *Size-OP*					
	F_2		F_4		F_5		F_2		F_4		F_5	
t-statistics	2	3	2	3	2	3	2	3	2	3	2	3
Fama-French factors	# of portfolios with *t*-statistic >											
MKT	25	25	25	25	25	25	25	25	25	25	25	25
SMB	25	25	25	25	25	25	25	25	25	25	25	25
HML	21	20	22	21	22	22	18	15	21	19	21	19
RMW	24	21	12	10	12	10	20	18	21	20	23	22
CMA	20	19	10	6	9	7	16	12	20	16	12	6
Accounting												
ACCR	19	15	5	1	9	5	21	20	6	1	12	8
SISSUE	24	21	14	9	15	6	24	24	16	11	19	11
CISSUE	19	17	12	4	14	3	23	21	17	12	12	7
NOA	13	12	14	9	14	8	16	12	12	6	15	6
GPROF	20	19	14	10	12	6	14	10	23	22	10	4
ASSETG	20	20	12	5	10	6	18	12	16	13	8	5
ROA	23	23	9	6	7	3	22	20	13	6	12	6
INVASSETS	20	16	10	4	10	6	12	6	13	8	6	2
HMLd	22	20	8	6	10	5	16	12	21	19	5	0
QMJ	24	21	14	7	14	10	20	19	15	9	15	8
qIA	22	18	7	6	10	6	11	6	19	14	12	4
qROE	22	20	13	12	10	6	20	19	10	4	15	8
SMBh	25	23	13	6	14	9	25	25	12	9	14	8
HMLh	20	19	7	2	4	1	21	16	19	17	5	2
RMWh	20	19	13	8	5	2	21	17	8	5	10	1
CMAh	20	17	15	7	11	6	15	11	12	7	11	6
Financial												
DISTR	24	23	16	7	10	3	24	24	17	12	11	3
OSCORE	24	23	14	9	15	10	24	23	13	10	12	7
BAB	21	18	15	12	14	7	13	5	13	12	11	4
TERM	16	6	3	1	3	0	18	4	0	0	0	0
CREDIT	0	0	1	0	1	0	0	0	0	0	1	0
YC	13	6	2	0	2	0	18	7	2	0	0	0
DY	0	0	4	1	5	2	6	1	4	1	6	2
PE	5	0	2	1	2	1	3	1	3	1	4	1

TABLE 7 (Continued)

Additional factor (f^*)	Panel A: 25 *Size-B/M*						Panel B: 25 *Size-OP*					
	F_2		F_4		F_5		F_2		F_4		F_5	
t-statistics	2	3	2	3	2	3	2	3	2	3	2	3
PDCR	13	6	4	0	2	0	7	3	9	4	4	0
SSKEW	25	24	14	6	18	6	24	23	13	10	12	11
MKTh	23	21	10	6	8	4	25	23	10	9	9	6
Macro												
MUNC	1	0	3	1	2	2	0	0	0	0	0	0
RUNC	1	0	1	0	2	0	0	0	2	0	1	0
FUNC	5	0	1	0	3	1	0	0	1	0	3	0
IP	4	0	4	0	4	1	5	1	6	1	5	0
OIL	11	2	3	0	2	0	14	5	4	1	5	1
URATE	0	0	2	0	2	0	2	0	2	0	2	0
NDCON	0	0	0	0	0	0	1	1	1	0	3	0
SCON	3	0	5	0	5	1	0	0	2	0	1	0
Behavioral												
BWSENT	0	0	2	0	2	0	0	0	1	0	2	0
HJTZSENT	0	0	0	0	1	0	1	0	3	0	3	0
PEAD	6	1	1	0	1	0	3	0	2	1	1	0
FIN	21	16	13	9	11	8	22	20	19	16	7	2
Microstructure												
TLIQ	6	1	10	6	9	6	4	1	7	3	10	2
NTLIQ	14	5	3	0	2	0	15	7	5	1	3	0
Other												
UMD	9	4	6	1	7	0	5	2	11	2	2	0
STREV	19	12	4	2	5	1	23	21	6	1	6	0
LTREV	22	18	13	7	12	9	20	19	11	5	11	3
MGMT	22	19	11	9	11	9	16	13	21	15	17	11
PERF	19	18	10	6	9	1	16	12	18	13	3	1

Additional factor (f^*)	Panel C: 25 *Size-Inv*									
	F_2		F_4		F_{5h}		F_{5r}		F_6	
t-statistics	2	3	2	3	2	3	2	3	2	3
Fama-French factors	# of portfolios with *t*-statistic >									
MKT	25	25	25	25	25	25	25	25	25	25
SMB	25	25	25	25	25	25	25	25	25	25
HML	21	19	20	17	20	17	21	17	21	17

(Continues)

TABLE 7 (Continued)

Additional factor (f^*)	Panel C: 25 *Size-Inv*									
	F_2		F_4		F_{5h}		F_{5r}		F_6	
t-statistics	2	3	2	3	2	3	2	3	2	3
RMW	22	20	15	13	16	14	15	13	16	14
CMA	22	21	23	22	24	22	23	22	23	22
Accounting										
ACCR	20	17	11	7	11	8	8	2	9	5
SISSUE	25	24	15	10	16	11	16	10	15	9
CISSUE	21	19	12	7	12	8	9	6	7	5
NOA	12	8	10	5	7	4	11	7	8	4
GPROF	20	13	19	14	11	7	18	14	2	0
ASSETG	22	21	12	6	11	5	12	7	11	6
ROA	23	23	15	11	18	11	11	7	12	6
INVASSETS	21	17	7	3	9	2	10	6	10	9
HMLd	22	17	18	14	9	5	20	13	7	3
QMJ	23	23	19	14	18	18	14	12	16	12
qIA	23	21	5	2	5	2	6	1	6	0
qROE	22	18	14	10	20	14	6	3	14	6
SMBh	25	24	13	8	15	7	13	6	15	5
HMLh	20	19	15	8	14	6	17	12	7	3
RMWh	19	16	11	5	15	6	3	1	3	1
CMAh	21	18	17	12	17	13	13	6	13	7
Financial										
DISTR	24	24	13	9	14	9	12	6	8	3
OSCORE	24	23	16	11	14	12	14	8	15	11
BAB	15	11	18	16	18	12	17	9	15	9
TERM	19	3	2	0	0	0	1	0	1	0
CREDIT	2	0	1	1	2	1	2	0	2	1
YC	15	6	1	0	0	0	1	0	0	0
DY	1	0	5	1	6	1	4	2	7	2
PE	3	1	3	2	4	2	2	2	4	2
PDCR	5	3	8	0	5	1	6	1	2	0
SSKEW	25	25	17	11	19	13	15	9	16	10
MKTh	23	21	16	13	16	12	15	12	14	11
Macro										
MUNC	1	0	1	0	1	0	2	0	2	0

TABLE 7 (Continued)

Additional factor (f^*)	Panel C: 25 *Size-Inv*									
	F_2		F_4		F_{5h}		F_{5r}		F_6	
t-statistics	2	3	2	3	2	3	2	3	2	3
RUNC	0	0	4	0	2	0	3	0	2	0
FUNC	0	0	0	0	0	0	1	0	0	0
IP	5	0	2	0	1	0	1	0	2	0
OIL	7	0	4	1	4	2	3	0	3	0
URATE	0	0	0	0	1	0	0	0	0	0
NDCON	1	0	2	0	1	0	2	0	2	0
SCON	3	0	3	0	3	1	2	1	1	1
Behavioral										
BWSENT	1	0	2	0	2	0	2	0	2	0
HJTZSENT	0	0	0	0	0	0	0	0	1	0
PEAD	6	1	4	2	2	1	4	1	1	0
FIN	23	20	21	18	21	17	18	17	16	9
Microstructure										
TLIQ	4	1	11	3	10	3	9	3	11	3
NTLIQ	17	6	5	0	3	0	6	0	7	0
Other										
UMD	3	1	10	4	8	1	9	3	7	1
STREV	21	13	8	3	7	1	10	5	8	2
LTREV	20	19	7	6	9	7	9	5	9	7
MGMT	17	15	20	15	15	11	19	15	11	8
PERF	17	12	13	11	15	6	16	13	5	0

Note: The table presents the number of statistically significant portfolios for the 25 *Size-B/M*, 25 *Size-OP*, and 25 *Size-Inv* portfolios from the time-heterogeneous Student's *t* DLR model; October 1973 to December 2016, 519 months. The right-hand side of the regressions includes the Gram-Schmidt orthonormal trend polynomials, the monthly dummy variables for the months of February through December, *MKT*, the additional factor, f^*, the formation factors, *HML*, *RMW*, and the lagged variables of the respective excess portfolio return and *MKT*. Panel A reports the number of statistically significant portfolios for the 25 *Size-B/M* portfolios. Panel B reports the number of statistically significant portfolios for the 25 *Size-OP* portfolios. Panel C reports the number of statistically significant portfolios for the 25 *Size-Inv* portfolios. Each panel reports the number of statistically significant portfolios for the respective additional factor, f^*, in each set of 25 regressions. The number of portfolios with *t*-statistics greater than 2 and 3 is reported in separate columns. The estimated models are: two-factor models (F_2) that combine *MKT* and f^*; four-factor models (F_4) that combine *MKT*, f^*, and the formation factors; five-factor models (F_5) that combine *MKT*, f^*, the formation factors, and either *HML* or *RMW*, depending on which of the two is missing from the four-factor models; and six-factor models (F_6) that combine *MKT*, f^*, the formation factors, and both *HML* and *RMW*.

and selecting the most relevant factors among a set of possible ones. Additionally, it can help identify whether any factors are missing from a time-series asset pricing model.

Through empirical applications, the article also contributes to the expanding body of literature on the factor zoo. First, the empirical results offer further insights into the redundancy of *HML* by indicating that it is not a redundant factor, while also suggesting that *CMA* is a redundant factor for describing returns. This finding

TABLE 8 Jensen's alphas from the time-heterogeneous Student's *t* DLR for the 25 Size-B/M, 25 Size-OP, and 25 Size-Inv portfolios.

Panel A: 25 *Size-B/M*

Size\B/M	α_i^J Low	2	3	4	High	α_i^J Low	2	3	4	High
	1. *SISSUE*					2. *QMJ*				
Small	−0.49	0.08	0.01	0.15	0.09	−0.34	0.10	−0.04	0.11	0.09
2	−0.09	0.02	0.01	0.06	−0.07	−0.09	−0.03	−0.05	−0.02	−0.09
3	0.07	0.07	−0.07	0.02	0.06	0.02	0.04	−0.06	0.00	0.11
4	0.10	−0.03	−0.00	0.06	0.05	0.06	−0.02	0.03	0.02	0.02
Big	0.03	−0.04	0.04	−0.18	−0.14	−0.08	0.03	0.10	−0.14	−0.15
	3. *OSCORE*					4. *SSKEW*				
Small	−0.44	0.10	0.02	0.18	0.13	−0.46	0.12	0.07	0.23	0.15
2	−0.11	0.01	0.02	0.06	−0.08	−0.11	0.05	0.04	0.09	−0.06
3	0.03	0.04	−0.07	0.01	0.09	−0.04	0.01	−0.12	0.01	0.07
4	0.05	−0.03	−0.05	0.07	0.00	−0.05	−0.08	−0.10	0.01	−0.05
Big	0.02	−0.01	0.08	−0.18	−0.21	0.07	−0.01	0.10	−0.15	−0.23
	5. *MGMT*					6. *SISSUE,QMJ,OSCORE,SSKEW,MGMT*				
Small	−0.44	0.11	−0.05	0.12	0.09	−0.35	0.15	−0.06	0.10	0.07
2	−0.01	0.05	−0.00	0.00	−0.09	0.05	0.04	−0.03	−0.02	−0.10
3	0.14	0.08	−0.08	−0.02	0.07	0.06	0.05	−0.07	0.05	0.10
4	0.16	−0.02	−0.07	0.05	0.01	0.12	0.02	0.03	0.04	−0.00
Big	0.02	−0.03	0.01	−0.15	−0.05	−0.05	0.06	0.12	−0.08	−0.06

Panel B: 25 *Size-OP*

Size\OP	α_i^J Low	2	3	4	High	α_i^J Low	2	3	4	High
	1. *SISSUE*					2. *QMJ*				
Small	−0.28	0.06	0.05	0.05	−0.10	−0.16	−0.01	−0.02	−0.03	−0.07
2	−0.11	−0.06	0.09	−0.04	0.06	−0.01	−0.13	−0.02	−0.12	0.02
3	0.05	0.03	0.00	0.01	0.05	0.10	−0.04	−0.03	−0.02	0.05
4	0.14	0.11	−0.00	0.01	−0.03	0.20	0.05	−0.04	−0.02	0.01
Big	0.14	0.03	0.04	−0.06	−0.00	0.10	−0.02	0.06	−0.03	0.02
	3. *OSCORE*					4. *SSKEW*				
Small	−0.23	0.06	0.06	0.04	−0.08	−0.24	0.14	0.11	0.13	−0.09
2	−0.05	−0.05	0.04	−0.08	0.01	−0.07	0.03	0.06	−0.06	−0.01
3	0.06	−0.02	−0.01	−0.01	0.03	0.01	−0.02	−0.04	−0.07	−0.07
4	0.14	0.07	−0.04	−0.02	−0.05	0.02	0.02	−0.06	−0.05	−0.15
Big	0.07	0.02	0.07	0.00	0.01	0.04	0.04	0.12	0.03	0.03

TABLE 8 (Continued)

Panel B: 25 *Size-OP*										
	α_i^J					α_i^J				
Size\OP	Low	2	3	4	High	Low	2	3	4	High
	5. *MGMT*					6. *SISSUE,QMJ,OSCORE,SSKEW,MGMT*				
Small	−0.27	0.04	0.01	0.02	−0.06	−0.22	0.04	0.01	0.06	−0.03
2	−0.08	−0.07	0.07	−0.03	0.12	0.01	−0.06	0.04	−0.04	0.09
3	0.09	0.02	0.01	0.02	0.11	0.09	0.04	0.02	0.00	0.04
4	0.14	0.10	−0.03	0.02	0.00	0.16	0.06	0.00	0.06	0.01
Big	0.18	−0.00	0.05	−0.07	0.01	0.15	−0.00	0.12	−0.05	−0.01

Panel C: 25 *Size-Inv*										
	α_i^J					α_i^J				
Size\Inv	Low	2	3	4	High	Low	2	3	4	High
	1. *SISSUE*					2. *QMJ*				
Small	−0.02	0.17	0.14	0.05	−0.37	0.07	0.13	0.07	−0.02	−0.30
2	−0.03	0.05	0.10	0.06	−0.14	0.03	−0.04	0.02	0.00	−0.12
3	0.04	0.11	0.10	0.09	0.03	0.10	0.08	0.07	0.05	0.04
4	−0.05	−0.01	0.02	0.08	0.10	−0.01	−0.02	0.00	0.07	0.16
Big	0.00	−0.12	−0.07	0.05	0.15	−0.05	−0.15	−0.06	0.05	0.13
	3. *OSCORE*					4. *SSKEW*				
Small	−0.03	0.16	0.16	0.05	−0.32	−0.04	0.21	0.20	0.10	−0.35
2	−0.04	0.04	0.07	0.04	−0.10	−0.06	0.11	0.11	0.08	−0.12
3	0.05	0.09	0.08	0.08	0.03	−0.01	0.06	0.06	0.06	−0.05
4	−0.07	−0.00	0.03	0.06	0.10	−0.13	−0.05	−0.01	−0.01	−0.02
Big	−0.01	−0.10	−0.03	0.05	0.10	0.03	−0.09	−0.01	0.12	0.10
	5. *MGMT*					6. *SISSUE,QMJ,OSCORE,SSKEW,MGMT*				
Small	−0.00	0.18	0.08	0.04	−0.35	0.04	0.20	0.07	0.00	−0.29
2	−0.01	0.04	0.06	0.03	−0.07	0.03	0.02	0.06	0.01	−0.06
3	0.04	0.10	0.06	0.09	0.09	0.08	0.09	0.05	0.06	0.05
4	−0.07	−0.03	−0.01	0.07	0.17	0.02	−0.03	0.03	0.06	0.17
Big	0.02	−0.12	−0.06	0.03	0.14	0.01	−0.14	−0.07	0.05	0.14

Note: The table presents Jensen's alphas for the 25 *Size-B/M*, 25 *Size-OP*, and 25 *Size-Inv* portfolios from the time-heterogeneous Student's *t* DLR model; October 1973 to December 2016, 519 months. The right-hand side of the regressions includes the Gram-Schmidt orthonormal trend polynomials, the monthly dummy variables for the months of February through December, *MKT*, the formation factors, *HML*, *RMW*, the identified potentially useful factors (*SISSUE*, *QMJ*, *OSCORE*, *SSKEW*, *MGMT*), and the lagged variables of the respective excess portfolio return and *MKT*. Panel A reports Jensen's alphas for the 25 *Size-B/M* portfolios. Panel B reports Jensen's alphas for the 25 *Size-OP* portfolios. Panel C reports Jensen's alphas for the 25 *Size-Inv* portfolios. Each panel reports Jensen's alphas for the time-series asset pricing models that combine *MKT*, the formation factors, *HML*, *RMW*, and either one of the identified potentially useful factors or all of them together. The reported Jensen's alphas are calculated by adding back the means of the associated deterministic and lagged variables to the estimated intercept of each regression (see Appendix B).

TABLE 9 Binary statistical significance of Gram-Schmidt orthonormal trend polynomials from the time-heterogeneous Student's t DLR for the 25 Size-B/M, 25 Size-OP, and 25 Size-Inv portfolios.

	Significance at the 5% level					Significance at the 1% level					Significance at the 0.1% level				
Panel A: 25 *Size-B/M*															
Size\B/M	**Low**	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**
Small	√	√	√	√	√	√	√		√	√	√	√		√	
2	√	√	√			√	√	√			√	√	√		
3	√	√	√	√		√	√	√	√		√	√	√	√	
4	√	√	√	√	√	√	√	√	√	√				√	
Big	√	√	√			√		√							
Panel B: 25 *Size-OP*															
Size\OP	**Low**	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**
Small	√	√	√	√	√	√	√	√	√		√	√		√	
2	√	√	√	√		√	√	√			√	√			
3	√	√	√	√	√	√	√	√	√	√	√	√	√		
4	√	√		√	√	√	√								
Big	√	√		√			√		√						
Panel C: 25 *Size-Inv*															
Size\Inv	**Low**	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**	**Low**	**2**	**3**	**4**	**High**
Small	√	√	√	√	√	√	√	√	√	√		√	√	√	√
2	√	√	√	√	√	√	√	√	√	√				√	√
3	√	√	√	√	√	√		√	√	√			√	√	√

TABLE 9 (Continued)

Panel C: 25 *Size-Inv*															
Size\Inv	Low	2	3	4	High	Low	2	3	4	High	Low	2	3	4	High
4		√	√	√	√			√					√		
Big			√	√	√			√	√	√					√

Note: The table presents the binary statistical significance of the Gram-Schmidt orthonormal trend polynomials for the 25 *Size-B/M*, 25 *Size-OP*, and 25 *Size-Inv* portfolios from the time-heterogeneous Student's *t* DLR model; October 1973 to December 2016, 519 months. The right-hand side of the regressions includes the Gram-Schmidt orthonormal trend polynomials, the monthly dummy variables for the months of February through December, *MKT*, the formation factors, *HML*, *RMW*, *QMJ*, and the lagged variables of the respective excess portfolio return and *MKT*. Panel A reports the binary statistical significance of the Gram-Schmidt orthonormal polynomials for the 25 *Size-B/M* portfolios. Panel B reports the binary statistical significance of the Gram-Schmidt orthonormal polynomials for the 25 *Size-OP* portfolios. Panel C reports the binary statistical significance of the Gram-Schmidt orthonormal polynomials for the 25 *Size-Inv* portfolios. Each panel reports the binary statistical significance of the Gram-Schmidt orthonormal polynomials at the significance levels of 5%, 1%, and 0.1% for the time-series asset pricing model that combines *MKT*, the formation factors, *HML*, *RMW*, and *QMJ*. Check marks indicate statistical significance at the respective significance level. Absence of check marks denotes statistical insignificance.

contradicts the U.S. study of Fama and French (2015) but aligns more closely with Fama and French's (2017) international tests, where *CMA* is found to be redundant for Europe and Japan, and its role is observed to be marginal for Asia Pacific. Moreover, this finding is in line with recent studies by Barillas and Shanken (2018), who find that *HML* is a robust explanator of returns, and Wahal (2019), who shows that *CMA* is redundant in the 1940–1963 period.

Second, the empirical results validate prior research and contribute supplementary evidence, suggesting that the vast majority of factors in the zoo, including macroeconomic (Kan & Robotti, 2008, 2009; Giglio & Xiu, 2021) and accounting (Linnainmaa & Roberts, 2018), are useless or redundant (Feng et al., 2020; Bryzgalova et al., 2023). Among the additional factors considered, the sole factor identified as useful is *QMJ* of Asness et al. (2019). Third, the results provide potential evidence of missing factors from common asset pricing models, such as the Fama-French five-factor model. This finding reinforces the proposition of He, Huang, Yuan, and Zhou (2021) that new factors are needed to better understand the cross-section of stock returns.

While the empirical applications in this article examine many of the prominent factors proposed in previous literature, the list is not exhaustive. Future work could consider other factors from the zoo, or possibly new factors, to determine whether a specific set of factors can fully explain the variation in returns captured by the deterministic orthogonal trend polynomials at the outset. Additionally, the starting point in the empirical applications is the Fama-French five-factor model, while results are obtained only for the 25 *Size-B/M*, 25 *Size-OP*, and 25 *Size-Inv* portfolios of Fama and French (2015). Future research could explore different starting points and portfolios, such as those in Hou, Xue, and Zhang (2015). Lastly, while the modeling approach of capturing variation in returns using deterministic variables can provide the basis for identifying whether relevant factors are missing from a time-series asset pricing model, it does not reveal specific information related to the exact number of missing factors. This limitation should be addressed in future research.

ACKNOWLEDGMENTS

I am very grateful to Erik Devos (Editor), the two anonymous referees, Wayne Ferson (discussant), Milan Ficura (discussant), Denada Ibrushi (discussant), Jae Hoon Kim, Georgios Koimisis, Jiadong Liu (discussant), Robert Martin, Tomas Nonnenmacher, Ryo Ogura (discussant), Yun Pu (discussant), Aris Spanos, Kwok Ping Tsang, Tony Via (discussant), and numerous conference participants for their useful comments and suggestions on earlier drafts of this paper. Any remaining errors are solely my responsibility.

ORCID

Michael Michaelides http://orcid.org/0009-0009-6727-5563

REFERENCES

Andreou, E., Pittis, N., & Spanos, A. (2001). On modelling speculative prices: the empirical literature. *Journal of Economic Surveys*, *15*(2), 187–220.

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

Asness, C. S., Frazzini, A., & Pedersen, L. H. (2019). Quality minus junk. *Review of Accounting Studies*, *24*(1), 34–112.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, *61*(4), 1645–1680.

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

Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *Journal of Finance*, *32*(3), 663–682.

Breeden, D. T., Gibbons, M. R., & Litzenberger, R. H. (1989). Empirical tests of the consumption-oriented CAPM. *Journal of Finance*, *44*(2), 231–262.

Bryzgalova, S., Huang, J., & Julliard, C. (2023). Bayesian solutions for the factor zoo: we just ran two quadrillion models. *Journal of Finance*, *78*(1), 487–557.

Burnside, C. (2016). Identification and inference in linear stochastic discount factor models with excess returns. *Journal of Financial Econometrics*, *14*(2), 295–330.

Campbell, J. Y. (1996). Understanding risk and return. *Journal of Political Economy*, *104*(2), 298–345.
Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In search of distress risk. *Journal of Finance*, *63*(6), 2899–2939.
Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of finance*, *52*(1), 57–82.
Chan, K. C., Chen, N. F., & Hsieh, D. A. (1985). An exploratory investigation of the firm size effect. *Journal of Financial Economics*, *14*(3), 451–471.
Chen, N. F., Roll, R., & Ross, S. A. (1986). Economic forces and the stock market. *Journal of Business*, *59*(3), 383–403.
Chen, L., Novy-Marx, R., & Zhang, L. (2011). An alternative three-factor model (Working Paper). Available at SSRN: https://ssrn.com/abstract=1418117
Chernov, M., Lochstoer, L. A., & Lundeby, S. R. (2018). Conditional dynamics and the multi-horizon risk-return trade-off (NBER Working Paper No. w25361). National Bureau of Economic Research. Available at NBER: https://www.nber.org/papers/w25361
Chib, S., & Zeng, X. (2020). Which factors are risk factors in asset pricing? A model scan framework. *Journal of Business & Economic Statistics*, *38*(4), 771–783.
Chib, S., Zeng, X., & Zhao, L. (2020). On comparing asset pricing models. *Journal of Finance*, *75*(1), 551–577.
Cochrane, J. H. (2011). Presidential address: discount rates. *Journal of Finance*, *66*(4), 1047–1108.
Cooper, M. J., Gulen, H., & Schill, M. J. (2008). Asset growth and the cross-section of stock returns. *Journal of Finance*, *63*(4), 1609–1651.
Daniel, K., Mota, L., Rottke, S., & Santos, T. (2020). The cross-section of risk and returns. *Review of Financial Studies*, *33*(5), 1927–1979.
Daniel, K., & Titman, S. (2006). Market reactions to tangible and intangible information. *Journal of Finance*, *61*(4), 1605–1643.
Daniel, K., Hirshleifer, D., & Sun, L. (2020). Short-and long-horizon behavioral factors. *Review of Financial Studies*, *33*(4), 1673–1736.
De Bondt, W. F., & Thaler, R. (1985). Does the stock market overreact? *Journal of Finance*, *40*(3), 793–805.
Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: empirical tests. *Journal of Political Economy*, *81*(3), 607–636.
Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*, *47*(2), 427–465.
Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.
Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22.
Fama, E. F., & French, K. R. (2017). International tests of a five-factor asset pricing model. *Journal of Financial Economics*, *123*(3), 441–463.
Fang, K.-T., Kotz, S., & Ng, K.-W. (1990). *Symmetric Multivariate and Related Distributions*, Chapman and Hall.
Feng, G., Giglio, S., & Xiu, D. (2020). Taming the factor zoo: A test of new factors. *Journal of Finance*, *75*(3), 1327–1370.
Ferson, W. E., & Harvey, C. R. (1991). The variation of economic risk premiums. *Journal of Political Economy*, *99*(2), 385–415.
Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, *111*(1), 1–25.
Freyberger, J., Neuhierl, A., & Weber, M. (2020). Dissecting characteristics nonparametrically. *Review of Financial Studies*, *33*(5), 2326–2377.
Gertler, M., & Grinols, E. L. (1982). Unemployment, inflation, and common stock returns. *Journal of Money, Credit and Banking*, *14*(2), 216–233.
Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, *57*(5), 1121–1152.
Giglio, S., & Xiu, D. (2021). Asset pricing with omitted factors. *Journal of Political Economy*, *129*(7), 1947–1990.
Golub, G. H., & Van Loan, C. F. (2013). *Matrix Computations* (4th ed.). Johns Hopkins.
Gospodinov, N., Kan, R., & Robotti, C. (2013). Chi-squared tests for evaluation and comparison of asset pricing models. *Journal of Econometrics*, *173*(1), 108–125.
Gospodinov, N., Kan, R., & Robotti, C. (2014). Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors. *Review of Financial Studies*, *27*(7), 2139–2170.
Gospodinov, N., Kan, R., & Robotti, C. (2017). Spurious inference in reduced-rank asset-pricing models. *Econometrica*, *85*(5), 1613–1628.
Gospodinov, N., Kan, R., & Robotti, C. (2019). Too good to be true? fallacies in evaluating risk factor models. *Journal of Financial Economics*, *132*(2), 451–471.
Haddad, V., Kozak, S., & Santosh, S. (2020). Factor timing. *Review of Financial Studies*, *33*(5), 1980–2018.
Hagerman, R. L. (1978). More evidence on the distribution of security returns. *Journal of Finance*, *33*(4), 1213–1221.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*(4), 1029–1054.
Harvey, C. R. (2017). Presidential address: the scientific outlook in financial economics. *Journal of Finance*, *72*(4), 1399–1440.
Harvey, C. R., & Liu, Y. (2019). A census of the factor zoo (Working Paper). Available at SSRN: https://ssrn.com/abstract=3341728
Harvey, C. R., Liu, Y., & Zhu, H. (2016). …and the cross-section of expected returns. *Review of Financial Studies*, *29*(1), 5–68.
He, A., Huang, D., Yuan, M., & Zhou, G. (2021). Tests of asset pricing models with a large number of assets (Working Paper). Available at SSRN: https://ssrn.com/abstract=3143752
He, Z., Kelly, B., & Manela, A. (2017). Intermediary asset pricing: new evidence from many asset classes. *Journal of Financial Economics*, *126*(1), 1–35.
Hendry, D. F., & Richard, J. F. (1983). The econometric analysis of economic time series. *International Statistical Review*, *51*(2), 111–148.
Hendry, D. F., & Richard, J. F. (1990). On the formulation of empirical models in dynamic econometrics. In C. W. J. Granger (Ed.), *Modelling economic series: Readings in econometric methodology* (pp. 304–334). Oxford University Press. (Original work published 1982 in *Journal of Econometrics*, 20(1), 3–33).
Heracleous, M. S., & Spanos, A. (2006). The Student's *t* dynamic linear regression: Re-examining volatility modeling. In D. Terrell & T. B Fomby (Eds.), *Econometric Analysis of Financial and Economic Time Series* (*Advances in Econometrics*, Vol. 20, Part A, pp. 289–319). Emerald Group Publishing Limited.
Hirshleifer, D., Hou, K., Teoh, S. H., & Zhang, Y. (2004). Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics*, *38*, 297–331.
Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: an investment approach. *Review of Financial Studies*, *28*(3), 650–705.
Huang, D., Jiang, F., Tu, J., & Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *Review of Financial Studies*, *28*(3), 791–837.
Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance*, *48*(1), 65–91.
Jensen, M. C. (1968). The performance of mutual funds in the period 1945-1964. *Journal of Finance*, *23*(2), 389–416.
Jurado, K., Ludvigson, S. C., & Ng, S. (2015). Measuring uncertainty. *American Economic Review*, *105*(3), 1177–1216.
Kan, R., & Robotti, C. (2008). Specification tests of asset pricing models using excess returns. *Journal of Empirical Finance*, *15*(5), 816–838.
Kan, R., & Robotti, C. (2009). Model comparison using the Hansen-Jagannathan distance. *Review of Financial Studies*, *22*(9), 3449–3490.
Kan, R., Robotti, C., & Shanken, J. (2013). Pricing model performance and the two-pass cross-sectional regression methodology. *Journal of Finance*, *68*(6), 2617–2649.
Kan, R., & Zhang, C. (1999a). GMM tests of stochastic discount factor models with useless factors. *Journal of Financial Economics*, *54*(1), 103–127.
Kan, R., & Zhang, C. (1999b). Two-pass tests of asset pricing models with useless factors. *Journal of Finance*, *54*(1), 203–235.
Kan, R., & Zhou, G. (1999). A critique of the stochastic discount factor methodology. *Journal of Finance*, *54*(4), 1221–1248.
Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā: The Indian Journal of Statistics, Series A*, *32*(4), 419–430.
Kelly, B. T., Pruitt, S., & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, *134*(3), 501–524.
Kim, J. H., & Ji, P. I. (2015). Significance testing in empirical finance: A critical review and assessment. *Journal of Empirical Finance*, *34*, 1–14.
Kleibergen, F. (2009). Tests of risk premia in linear factor models. *Journal of Econometrics*, *149*(2), 149–173.
Koundouri, P., Kourogenis, N., & Pittis, N. (2016). Statistical modeling of stock returns: explanatory or descriptive? A historical survey with some methodological reflections. *Journal of Economic Surveys*, *30*(1), 149–164.
Langlois, H. (2020). Measuring skewness premia. *Journal of Financial Economics*, *135*(2), 399–424.
Lettau, M., & Pelger, M. (2020). Factors that fit the time series and cross-section of stock returns. *Review of Financial Studies*, *33*(5), 2274–2325.
Linnainmaa, J. T., & Roberts, M. R. (2018). The history of the cross-section of stock returns. *Review of Financial Studies*, *31*(7), 2606–2649.

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *Journal of Finance*, *20*(4), 587–615.
Linton, O. (2019). *Financial econometrics*. Cambridge University Press.
Lo, A. W., & MacKinlay, A. C. (1990). Data-snooping biases in tests of financial asset pricing models. *Review of Financial Studies*, *3*(3), 431–467.
Loughran, T., & Ritter, J. R. (1995). The new issues puzzle. *Journal of Finance*, *50*(1), 23–51.
Michaelides, M. (2021). Large sample size bias in empirical finance. *Finance Research letters*, *41*, 101835.
Michaelides, M., & Spanos, A. (2020). On modeling heterogeneity in linear models using trend polynomials. *Economic Modelling*, *85*, 74–86.
Novy-Marx, R. (2013). The other side of value: the gross profitability premium. *Journal of Financial Economics*, *108*(1), 1–28.
Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, *18*(1), 109–131.
Pástor, L., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, *111*(3), 642–685.
Shanken, J., & Zhou, G. (2007). Estimating and testing beta pricing models: alternative methods and their performance in simulations. *Journal of Financial Economics*, *84*(1), 40–86.
Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, *19*(3), 425–442.
Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *Accounting Review*, *71*(3), 289–315.
Spanos, A. (1990). Towards a unifying methodological framework for econometrics modelling. In C. W. J. Granger (Ed.), *Modelling economic series: Readings in econometric methodology* (pp. 335–364). Oxford University Press. (Original work published 1988 in *Economic Notes*, 20(1), 107–134).
Stambaugh, R. F., & Yuan, Y. (2017). Mispricing factors. *Review of Financial Studies*, *30*(4), 1270–1315.
Titman, S., Wei, K. J., & Xie, F. (2004). Capital investments and stock returns. *Journal of Financial and Quantitative Analysis*, *39*(4), 677–700.
Wahal, S. (2019). The profitability and investment premium: Pre-1963 evidence. *Journal of Financial Economics*, *131*(2), 362–377.

How to cite this article: Michaelides, M. (2025). Time-series factor modeling and selection. *Journal of Financial Research*, *48*, 839–875. https://doi.org/10.1111/jfir.12429

APPENDIX A: STATISTICAL MODEL

General version

The general version of the time-heterogeneous Student's *t* DLR model is as follows:

$$y_t = \alpha + \boldsymbol{\beta}^\top \mathbf{X}_t + \boldsymbol{\gamma}^\top \mathbf{D}_t + \sum_{j=1}^{p} \boldsymbol{\delta}_j^\top \mathbf{Z}_{t-j} + \epsilon_t, t \in \mathbb{N}, \tag{A1}$$

where y_t is the dependent variable, $\mathbf{X}_t := (x_{1t}, x_{2t}, ..., x_{lt})^\top$ is a $(l \times 1)$ vector of independent variables, $\mathbf{D}_t := (d_{1t}(t), d_{2t}(t), ..., d_{qt}(t))^\top$ is a $(q \times 1)$ vector of deterministic components, $\mathbf{Z}_t := (y_t, \mathbf{X}_t^\top)^\top, \mathbf{Z}_{t-j}$, for $j = 1, 2, ..., p$, is a $((l+1) \times 1)$ vector of lagged variables, and $\boldsymbol{\vartheta} := (\alpha, \boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\delta}^\top)^\top$, with $\boldsymbol{\delta}^\top := \left(\delta_1^\top, \delta_2^\top, ..., \delta_p^\top\right)$, is a vector of parameters. The error term, ϵ_t, is distributed $\text{St}(0, h^2(t); v + l^*)$.

Degrees of freedom

The error term, ϵ_t, is distributed $\text{St}(0, h^2(t); v + l^*)$. v is the degrees of freedom of the joint Student's *t* distribution. $l^* = l + ((l+1) \times p)$ denotes the number of the conditioning non-deterministic variables in the model. Therefore, the degrees of freedom of the conditional Student's *t* distribution is equal to $v + l^*$.

Conditional moments

The first two conditional moments of the model, stemming from the Student's t distribution, take the following form:

$$E\left(y_t \left| \sigma\left(\mathbf{X}_t, \mathbf{Z}_{t-1}^0\right)\right.\right) = \alpha + \boldsymbol{\beta}^\top \mathbf{X}_t + \boldsymbol{\gamma}^\top \mathbf{D}_t + \sum_{j=1}^{p} \boldsymbol{\delta}_j^\top \mathbf{Z}_{t-j}, t \in \mathbb{N}, \tag{A2}$$

$$Var\left(y_t \left| \sigma\left(\mathbf{X}_t, \mathbf{Z}_{t-1}^0\right)\right.\right) = \left(\frac{v\sigma^2}{v + l^* - 2}\right)\left(1 + \frac{2\varrho^2}{v}\right), t \in \mathbb{N}, \tag{A3}$$

for

$$\varrho^2 = \frac{1}{2}\left[(\mathbf{X}_t - \boldsymbol{\mu}_\mathbf{x}(t))^\top \mathbf{Q}_\mathbf{x}^{-1}(\mathbf{X}_t - \boldsymbol{\mu}_\mathbf{x}(t)) + \sum_{j=1}^{p}\left((\mathbf{Z}_{t-j} - \boldsymbol{\mu}_\mathbf{z}(t))^\top \mathbf{Q}_\mathbf{z}^{-1}(\mathbf{Z}_{t-j} - \boldsymbol{\mu}_\mathbf{z}(t))\right)\right],$$

where $\sigma(.)$ denotes the conditioning information set (σ-field), $\mathbf{Z}_{t-1}^0 := (\mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, ..., \mathbf{Z}_1)$, σ^2 is the unconditional variance, $\boldsymbol{\mu}_\mathbf{x}(t)$ and $\boldsymbol{\mu}_\mathbf{z}(t)$ are the time-dependent unconditional mean vectors of $\mathbf{X}_t$ and $\mathbf{Z}_{t-j}$, respectively, and $\mathbf{Q}_\mathbf{x}$ and $\mathbf{Q}_\mathbf{z}$ are the variance-covariance matrices of $\mathbf{X}_t$ and $\mathbf{Z}_{t-j}$, respectively.

For additional insights into the basic structure of the model, refer to Heracleous and Spanos (2006).

APPENDIX B: INTERCEPT TERM

The intercept of the time-heterogeneous Student's t DLR model in (1), which is calculated as:

$$\alpha_i = E(R_{it}) - \boldsymbol{\beta}_i^\top E(\mathbf{f}_t) - [\cdots], \tag{B1}$$

where $[\cdots]$ is used to denote the means of the deterministic and lagged variables included into the model in (1), is not the traditional Jensen's alpha (Jensen, 1968), thus not interpretable in the traditional way (see, for example, Fama & French, 1993, 2015).

Given that the model in (1) incorporates deterministic and lagged variables, the appropriate Jensen's alpha is calculated by adding back their means, as follows:

$$\alpha_i^J = \alpha_i + [\cdots], \tag{B2}$$

where α_i is the intercept term from (B1).

The Jensen's alpha in (B2) may be used in the traditional way to compare the relative performance among competing time-series asset pricing models. It is notable that the Jensen's alpha would not be equal to the estimated intercept coefficient of a model that does not include deterministic and lagged variables because under i.i.d. the estimates of factor loadings are likely to be biased.

APPENDIX C: ORTHOGONAL TREND POLYNOMIALS

In general, the most widely used orthogonal polynomials are the classical orthogonal polynomials, consisting of the Hermite, Laguerre, and Jacobi polynomials (including the Jacobi special cases of Gegenbauer, Chebyshev, and Legendre polynomials). However, when these classical polynomials are used as trend polynomials in a regression model to capture the mean and variance nonstationarity in returns, they are likely to give rise to collinearity problems. Conversely, the Gram-Schmidt orthonormal polynomials offer the flexibility to extend them up to higher degrees without giving rise to collinearity; for additional information, see Michaelides and Spanos (2020).

Gram-Schmidt orthonormal trend polynomials

The Gram-Schmidt orthonormal trend polynomials are computed as:

$$v_{jt} = \frac{\tilde{v}_{jt}}{\|\tilde{v}_{jt}\|}, j = 0, 1, 2, \ldots, m^*, \tag{C1}$$

where $\{\tilde{v}_{0t}, \tilde{v}_{1t}, \tilde{v}_{2t}, \ldots, \tilde{v}_{m^*t}\}$ is the orthogonal basis of $\{1, t, t^2, \ldots, t^{m^*}\}$, for $t = 1, 2, \ldots, n$, and n being the sample size.

Regarding the Gram-Schmidt orthonormal trend polynomials, it is important to note the following: (a) when they are to be included in a regression model with an intercept term, the polynomial of degree 0 (i.e., $\tilde{v}_{0t}$) – which is a vector of ones – must be dropped; and (b) in practice, it is advisable to use the modified Gram-Schmidt algorithm which is less sensitive to rounding errors and more numerically stable compared to the classical Gram-Schmidt algorithm; for additional information, see Golub and Van Loan (2013 (pp. 254–55)).