Dissecting Anomalies in Conditional Asset Pricing

VALENTINA RAPONI   PAOLO ZAFFARONI

IESE Business School   Imperial College London

This Draft

January 2, 2024

Abstract

We develop a methodology for estimating and testing the effect of anomalies in conditional asset pricing models when premia are time-varying. The methodology also provides an inferential procedure for characteristic-based portfolios. Our method, which builds on the two-pass methodology, is developed for ordinary and weighted least-squares estimation, considering both cases of correct specification and global misspecification of the candidate asset pricing model. A cross-sectional R-squared test to dissect anomalies is proposed, establishing its limiting properties under the null hypothesis of no effect of anomalies and its alternative. Using a dataset of $20,000$ individual US stock returns, we find that although anomalies are statistically significant in about half the cases (out of 170 anomalies), they explain a small fraction (less than 10%) of the cross-sectional variation of expected returns. Anomalies tend to be more important during economic and financial crises.

Keywords: Anomalies, characteristic-based portfolios, time-variation, two-pass methodology, OLS, WLS, global misspecification, cross-sectional R-square, large-N asymptotics.

This project has received funding from Grant PID2020-115069GB-I00, funded by AEI/ MCIN/ 10.13039/ 501100011033 and from the postdoctoral fellowships programme Beatriu de Pinos, funded by the Secretary of Universities and Research (Government of Catalonia) and by the Horizon 2020 programme of research and innovation of the European Union under the Marie Sktodowska-Curie grant agreement No 801370. We are grateful to Svetlana Bryzgalova, Victor DeMiguel, Patrick Gagliardini (SOFIE discussant), Daniele Massacci (York discussant), Georgios Skoulakis (WFA discussant), and to seminar partecipants at IESE, LBS, Imperial College London, and Rotterdam for comments.

1 Introduction

CAN WE CITE MCLEAN AND PONITIEFF?

This paper provides a general methodology to formally estimate and test for the economic significance of asset pricing anomalies, within conditional asset pricing models, when both risk premia and risk exposures are allowed to be time-varying.

Asset pricing theory implies that the cross-sectional variation in expected returns should be explained by the exposures to systematic risk factors (Sharpe (1964) and Lintner (1965)). However, over decades, researchers have identified many "anomalies", where some firm- or asset-specific characteristics can predict the cross-section of expected returns, even after controlling for risk factors and their risk exposures.[1]

Despite the very extensive literature on asset pricing anomalies, identifying and understanding such seemingly anomalous predictability represents one of the biggest challenges in empirical asset pricing, as it is essential for commanding what an investor considers to be risk. The availability of rigorous methodological approaches to estimate and test for anomalies is therefore of paramount importance, as highlighted by Fama and French (2008) and Hou, Chen, and Zhang (2020). Broadly speaking, there are two main standard approaches to test for anomalies: (i) sorting average returns on anomaly variables, and (ii) using anomaly variables as additional regressors in the Fama and MacBeth (1973) two-pass regression. In this latter case, the conventional approach involves T cross-sectional ordinary least squares regressions (CSR OLS hereafter) of asset returns on the anomaly variables, one for each period, and then interpreting the average of the T slopes' estimates as the anomaly's premia.

While sorting offers an immediate picture of how returns vary across the spectrum of the anomaly variable(s), it becomes unreliable when the sort is made on more than three variables at the time, as it can result in many empty sets, and, even more importantly, it does not allow to make inference on the significance of the anomalies, although recent advances addressed this issue (see Cattaneo, Crump, Farrell, and Schaumburg (2020)). On the other hand, the two-pass

[1] Examples of anomalies include momentum (Jegadeesh and Titman (1993)), the NASDAQ anomaly (Brennan, Chordia, and Subrahmanyam (1998)), firm size and the book-to-market ratio (Fama and French (1993)), liquidity (Acharya and Pedersen (2005) and Brennan, Chordia, Subrahmanyam, and Tong (2012)), carry (Koijen, Moskowitz, Pedersen, and Vrugt (2018)), and idiosyncratic volatility (Ang, Hodrick, Xing, and Zhang (2006)), among many others. Hou, Chen, and Zhang (2020) document 452 anomalies.

regression provides direct estimates of the marginal effects of each anomaly (together with their standard errors), offering a formal way to make inference on the potential existence of anomalies. However, in this paper, we show that the conditions required for its validity are often hard to justify in practice, hence invalidating many of the inferential results on the anomalies' premia.

In particular, we show that the conventional approach - based on the large-T sampling scheme - provides an accurate estimation of the average anomaly premium and, moreover, only if one is willing to assume orthogonality between the factor betas and the anomalies. This condition appears to be at odds with the empirical evidence here provided, where we very often find statistically significant (non-zero) correlation between the estimated betas and the anomalies. More importantly, we show that the conventional approach is ill-suited to estimate time-varying premia because time-variation in the premia would be completely obscured by averaging the T estimates. Averaging the premia estimates over very short rolling time windows would partially resolve this problem but at the cost of invalidating the large-T asymptotic theory, which underlies all the inferential results of the conventional approach. In particular, when T is asymptotically large, we show that the classical t-ratio of the average anomaly premium is downward biased whenever one assumes that the (true) anomaly premium varies over time. In other words, we could reject the null hypothesis of a zero average premium more often than we should, unless the premium turns out to be constant across time (and orthogonality between the factor betas and the anomalies holds).

Introducing a new methodology that resolves all these challenges is the objective of this paper. First, given the overwhelming evidence of time-varying risk premia in the empirical asset pricing literature, we design our methodology to capture time-variation in the anomalies' premia, leaving their dynamics unspecified (i.e., nonparametrically). This is accomplished by exploiting large cross-sections of size N of asset returns while keeping their time series dimension (T) fixed and possibly very small. Given the large availability of individual securities, such a setting has gained significant attention in recent years, thanks also to its flexibility to handle time variation of any (non-parametric) form, hence mitigating the risk of model misspecification and potential structural breaks in the data.[2] Moreover, by allowing the use of short (i.e., small T) unbalanced panels, our small-T approach allows one to mitigate the issue of missing data, which is a frequent, yet often

[2]Large-T asymptotic results require fully-specified parametric assumptions to capture time-variation of loadings and risk premia, such as eg., by assuming them to be linear functions of some observed state variables (see, e.g., Gagliardini, Ossola, and Scaillet (2016)).

overlooked, feature of company fundamentals. This problem affects the time-series availability of almost any characteristic and becomes extremely severe when one needs to analyze multiple characteristics over the same time span (see, for example, the recent contributions of Bryzgalova, Lerner, Lettau, and Pelger (2022) and Freyberger, Höppner, Neuhierl, and Weber (2021) for methodologies that tackle missing financial data).

Our methodology builds on the classical Fama and MacBeth (1973) two-pass procedure and maintains its computational ease and interpretability, despite not relying on its strict assumptions. In particular, we derive novel OLS-type estimators of both risk and anomaly time-varying premia and establish their asymptotic properties, showing how to derive closed-form standard errors to conduct correct inference on model's premia. The large-N and fixed-T setting allows us to work under very mild assumptions, which can now accommodate the more realistic case of both cross- and time-correlation between returns and anomalies, in contrast to existing methodologies.

We also extend our analysis in four main directions. The first extension introduces a new weighted least square (WLS) version of the estimator. This idea is strongly motivated by the recent literature that shows that microcaps can adversely affect the significance of anomalies. Indeed, as reported by Hou, Chen, and Zhang (2020), microcaps represent only 3.2% of the aggregate NYSE-Amex-NASDAQ market capitalization, but they account for more than 60% of the traded stocks in the market.[3] In this case, performing a simple CSR of returns on anomaly variables would make the estimates very sensitive to microcap outliers (see Hou, Chen, and Zhang (2020), Green, Hand, and Zhang (2017), and Fama and French (2008)). This impact could be mitigated by a WLS estimation, which minimizes a weighted sum of squared errors. The derivation of a WLS estimator is technically challenging, due to the potential presence of both time- and cross-sectional-correlation between weights and asset returns. This is very likely especially if the weights are defined to be equal (or proportional) to assets' market capitalization. We address these challenges and establish the limiting properties of our novel WLS estimator, providing its standard errors in closed-form.

The second extension of our analysis is about robustifying our inferential results to the case of global model misspecification.[4] Indeed, the significance of premia estimates can be dramatically

[3]In Hou, Chen, and Zhang (2020), microcaps identify all stocks with a market capitalization smaller than the 20th percentile in the distribution of all the NYSE stock market equity.

[4]By global misspecification in the context of beta-pricing models, we refer to deviations, of unspecified form, from exact pricing.

affected by the degree of model misspecification, which could arise due to the omission of potentially relevant risk factors from the model, or because one selects the wrong (or incomplete) set of anomalies (see Jagannathan and Wang (1998)).[5] To mitigate this risk, we provide asymptotically-valid standard errors of the anomalies' premia estimates, which are robust to global model misspecification.

As a third extension, we propose a cross-sectional R-squared measure, that can be used to quantify the joint effect of anomalies on the cross-section of expected returns. Indeed, although our t-ratios can be correctly used to assess the significance of a premium estimate corresponding to the single anomaly, a cross-sectional R-squared test permits quantifying the portion of the total asset variability *jointly* explained by the anomalies. For example, one might be interested in the joint effect of anomalies belonging to the category (say, e.g., all the momentum anomalies). Specifically, we establish the limiting distribution of the proposed R-squared measure under both the null hypothesis of zero anomalies' premia and the alternative hypothesis of priced anomalies.

The fourth extension appears to be the most relevant one. The ability of our methodology to estimate time-varying the slope coefficients to the anomalies, granted by by our large-N with fixed-T sampling scheme, implies that the same methodology provides an important advancement in the estimation of characteristic-based portfolios, given Fama (1976)'s insight that Fama-MacBeth regression coefficients (to characteristics) are returns to tradeable portfolios. To understand this by-product of our methodology, it sufficies to recognize that a conditional asset pricing anomalies with constant risk exposures and time-varying anomalies spannng the pricing errors, i.e. the alphas, can be (under suitable restrictions here formalized) equivalent to a model with constant pricing errors and time-varying risk exposures in the same anomalies. Hence, the asymptotic analysis of the OLS- and WLS-type estimators of the anomalies's slopes readily provides a novel inferential procedure for the cross-sectional characteristic-based portfolios, asymptotically-correct when N is large, once one narrows, as anomalies, on the special case of firms' characteristics. Moreover, our time-varying estimators, and their associated limiting distribution theory, can be deployed for accurate estimation of the risk premia associated with the characteristic-based portfolios. This is relevant as our analysis of the conventional large-T approach shows that the latter leads to invalid estimation of the risk premia to characteristics-based portolios when the former are time-varying.

[5] Jagannathan and Wang (1998) analyze the implications of model misspecification using the two-pass methodology, valid under the large-T set up.

We construct locally-averaged estimator of these risk premia and show their validity, together with a infential procedure.

We present an extensive empirical application using data provided by Chen and Zimmermann (2019), from which we extract 170 anomalies at the monthly frequency (January 1986 - December 2020). We find patterns of time-variation according to which the importance of anomalies emerge often during financial crises (about 70% of the times). Although statistically the contribution of anomalies appear significant (at 5% level) for about half of cases, anomalies explain a very small fraction of the cross-sectional variation of expected returns, with only 4% of them explaining above 20%, and more than half contributing to less than 1%. In contrast, the estimated betas do not show the same pattern across time, although explain a similar fraction to anomalies of the cross-section of asset returns. The large majority of the variation in the cross-section of asset returns remains unexplained.

The paper is structured as follows. Section 2 describes the main literature, while Section 3 introduces our conditional asset pricing framework. In Section 4 we provide both analytical and numerical evidence that highlights the pitfalls of the conventional large-T method used to detect anomalies. Our methodology is formalized in Sections 5 and 7, where we present our OLS-type and WLS-type estimators, respectively, with their corresponding statistical analysis. The implication of our methodology for inference on characteristic-based portfolios is elaborated in Section 6. Section 8 shows how to robustify our methodology to global misspecification, while Section 9 describes our cross-sectional R-square test. The empirical application is contained in Section 10. Section 11 concludes. The proofs of the theorems, together with some preliminary lemmas, are relegated to the Online Appendix (Raponi and Zaffaroni (2023)), referred to as OA throughout the manuscript.

2 Literature Review

The literature on asset pricing anomalies is very extensive, with a list of more than 400 papers proposing (or *dissecting*) anomalies thought to be relevant in explaining the cross-sectional variation of stock returns (see Hou, Chen, and Zhang (2020) for a detailed list). These empirical findings have spurred a growing literature that tries to summarize (or *digest*) this cross-sectional variation with new risk factors.[6] However, it seems that there are still many asset-specific characteristics that cannot be explained by any common risk factors, and that still represent the major determinants of average equity returns (see Daniel and Titman (1997), Lewellen (2015), and Dong, Yan, Rapach, and Zhou (2021), among others).

The apparent significance of such a wide range of anomalies can be in part attributed to a lack of proper statistical methodologies. A recent example is Hou, Chen, and Zhang (2020), which cast doubts on the empirical validity of 452 anomalies proposed in asset pricing and accounting literature, showing that 65% of them fail to explain the cross-section of average stock returns, with the biggest failure (96%) being observed in the trading frictions literature. This empirical finding is even more severe if one allows for multiple testing approaches (see Harvey, Liu, and Zhu (2016)).

The two-pass methodology augmented with anomalies has been studied and extended by the literature in many directions. Important results, valid under large-T and the assumption of constant premia, have been provided by Jagannathan and Wang (1998), who derived the limiting distribution of the CSR OLS estimator under the null hypothesis of no effect of anomalies. Brennan, Chordia, and Subrahmanyam (1998) propose to first net out average returns from the risk exposure to common risk factors, and then to regress these risk-adjusted average returns on observed firms' characteristics, to test for the potential effect of anomalies. Their approach has been further extended by Avramov and Chordia (2006), allowing for time variation in the factors' loadings through observed state variables. Chordia, Goyal, and Shanken (2015) examine the two-pass estimator in situations when N is much larger than T, and where the anomaly variables are also allowed to vary over time. However, in their work, a bootstrap procedure is proposed to derive the standard errors of the premia estimator.

Going beyond the two-pass methodology, alternative approaches have been also proposed to quantify the economic relevance of anomalies. Important examples are the semi-parametric estimation of Connor and Linton (2007), the Projected Principal Component Analysis of Fan, Liao, and Wang (2016), the Instrumented Principal Component analysis of Kelly, Pruitt, and Su (2019), and the Bayesian approach of Kozak, Nagel, and Santosh (2020). Other studies have also quantified the

[6] Prominent examples are the Fama and French (1993) and Fama and French (2015) factors, Carhart (1997) and Jegadeesh and Titman (1993) momentum factors, the liquidity factors of Pástor and Stambaugh (2003) and Acharya and Pedersen (2005), the Ang, Hodrick, Xing, and Zhang (2006) idiosyncratic risk factor, the Hou, Chen, and Zhang (2015) four q-factors, and the Stambaugh and Yuan (2017) lucky factors, among many others.

impact of firms' characteristics on investors' portfolio choices (see, e.g., DeMiguel, Martin-Utrera, Nogales, and Uppal (2020) and Kim, Korajczyk, and Neuhierl (2021)). Using non-parametric methods, Freyberger, Neuhierl, and Weber (2020) show that characteristics play a crucial role in terms of model selection and return predictability.

Moreover, most of the empirical asset pricing literature that deals with anomalies uses portfolio data constructed from a relatively small subset of asset-specific predictors. However, although the use of portfolios reduces the sampling variability of estimated loadings, it sensibly reduces returns' heterogeneity (see Ang, Liu, and Schwarz (2020)) and makes it impossible to investigate the joint effect of a high-dimensional set of anomalies. In addition, portfolio data could be highly sensitive to data-snooping biases, especially when the same data set is repetitively examined (see Lo and MacKinlay (1990), Brennan, Chordia, and Subrahmanyam (1998), Conrad, Cooper, and Kaul (2003), Barras, Scaillet, and Wermers (2010), McLean and Pontiff (2016), and Chen (2021)). These issues can be sensibly mitigated (if not entirely avoided) using our approach, which applies to large cross-sections of individual assets, much less scrutinized than portfolio data sets.

Our methodology contributes also to the literature on the estimation of cross-sectional characteristic-based portfolios, introduced by Fama (1976). Various methods have been employed to construct characteristic-based factors, ranging from sorting approaches (see Fama and French (1993)), to cross-sectional OLS regressions (see Back, Kapadia, and Ostdiek (2015) and Fama and French (2020)), linear combinations in rank-transformed centred characteristics (see Kozak, Nagel, and Santosh (2020)), and PCA-type methods (see Kelly, Pruitt, and Su (2019) and Kim, Korajczyk, and Neuhierl (2021)).[7] Daniel, Mota, Rottke, and Santos (2020) study how to clean up the characteristic-based portfolios from unpriced sources of risk. Our methodology permits accurate estimation of the OLS and WLS CSR characteristic-based portfolios with an asymptotically valid inferential procedure (in N).

[7] Kelly, Pruitt, and Su (2019) and Kim, Korajczyk, and Neuhierl (2021) can be interepreted as methodogies for either testing for anomalies (when these are posed to span the pricing errors) or estimating characteristic-based portfolios (when the risk exposures to the latent risk factors are posed as functions of firms' characteristics), or both.

3 Conditional Asset Pricing with Anomalies

Given our objective of estimating and testing for anomalies in a time-varying setting, the first step of our analysis requires the introduction of a *conditional* asset pricing factor model that admits the presence of anomalies. We assume that asset returns are governed by the following conditional asset pricing factor model:

$$R_{it} \quad = \quad \alpha_{i,t-1} + \boldsymbol{\beta}'_{i,t-1}\mathbf{f}_t + \epsilon_{it}, \quad \text{for } i = 1, \cdots, N, \;\; t = 1, \ldots, T \tag{1}$$

where R_{it} represents the gross return of stock i at time t, $\alpha_{i,t-1}$ is a potentially time-varying and asset specific intercept, $\boldsymbol{\beta}_{i,t-1} = (\beta_{i1,t-1}, \ldots, \beta_{iK_{\mathrm{f}},t-1})'$ is the vector of time-varying loadings on K_{f} *observed* risk factors $\mathbf{f}_t = (f_{1,t}, \ldots, f_{K_{\mathrm{f}},t})'$, and ϵ_{it} is the asset-specific error component. Using matrix notation, the asset pricing model in (1) can be re-written as

$$\mathbf{R}_t = \boldsymbol{\alpha}_{t-1} + \mathbf{B}_{t-1}\mathbf{f}_t + \boldsymbol{\epsilon}_t, \tag{2}$$

where $\mathbf{R}_t$ denotes the $N \times 1$ vector of asset returns at time t, $\boldsymbol{\alpha}_{t-1} \equiv [\alpha_{1,t-1}, \ldots, \alpha_{N,t-1}]'$, $\mathbf{B}_{t-1} \equiv (\boldsymbol{\beta}_{1,t-1}, \ldots, \boldsymbol{\beta}_{N,t-1})'$, and $\boldsymbol{\epsilon}_t \equiv (\epsilon_{1,t}, \ldots, \epsilon_{N,t})'$.

When conditional no-arbitrage and full diversification of the mean-variance frontier hold (see Chamberlain and Rothschild (1983, Corollary 1) and Hansen and Richard (1987) for an extension to a conditional asset pricing setup), exact pricing follows. That is:

$$\mathrm{E}\left[R_{it} | I_{t-1}, \boldsymbol{\Pi}\right] = \gamma_{0,t-1} + \boldsymbol{\gamma}'_{\mathrm{f,t-1}}\boldsymbol{\beta}_{i,t-1}, \tag{3}$$

where $\mathrm{E}[\cdot]$ denotes the expectation operator, I_t represents the information set available up to time t, and $\boldsymbol{\Pi}$ defines the complete set of parameters, known to the agent when evaluating expected returns, with $\{\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_{\mathrm{f}}, \mathbf{B}\} \subset \boldsymbol{\Pi}$, where $\boldsymbol{\gamma}_0 = (\gamma_{0,1}, \cdots, \gamma_{0,T-1})'$ denotes the zero-beta rate vector, and $\boldsymbol{\gamma}_{\mathrm{f}} = (\boldsymbol{\gamma}_{\mathrm{f},1}, \cdots, \boldsymbol{\gamma}_{\mathrm{f,T-1}})'$ denotes the risk premia matrix associated with the observed risk factors $\mathbf{f}_t$, and $\mathbf{B} = (\mathbf{B}_1, \cdots, \mathbf{B}_{T-1})$ is the loadings matrix.

However, we are specifically interested in situations where (3) might not hold and, in fact, we replace it by

$$\mathrm{E}\left[R_{it} | I_{t-1}, \boldsymbol{\Pi}\right] = a_{i,t-1} + \gamma_{0,t-1} + \boldsymbol{\gamma}'_{\mathrm{f,t-1}}\boldsymbol{\beta}_{i,t-1}, \tag{4}$$

for some time-varying and asset-specific *pricing errors* $\mathbf{a} \subset \boldsymbol{\Pi}$, with $\mathbf{a} = (\mathbf{a}_1, \cdots, \mathbf{a}_{T-1})'$, and where $\mathbf{a}_{t-1} = (a_{1,t-1}, \cdots, a_{N,t-1})'$.

In this paper, we assume that the pricing errors $a_{i,t-1}$ are governed by some *observed* characteristics, represented by a $K_{\mathrm{z}} \times 1$ vector of *asset-specific* and possibly time-varying variables, $\mathbf{z}_{i,t-1}$, which we refer to as *anomalies.* Formally, we assume that

$$a_{i,t-1} = \boldsymbol{\gamma}_{\mathrm{z,t-1}}' \mathbf{z}_{i,t-1}, \quad \text{for } i = 1, \cdots, N, \;\; t = 1, \ldots, T \tag{5}$$

for some vector of coefficients $\boldsymbol{\gamma}_{\mathrm{z}} \subset \boldsymbol{\Pi}$, where $\boldsymbol{\gamma}_{\mathrm{z}} = (\boldsymbol{\gamma}_{\mathrm{z},1}, \cdots, \boldsymbol{\gamma}_{\mathrm{z,T-1}})'$ denotes the *anomalies' premia* matrix. Clearly, should all the elements of $\boldsymbol{\gamma}_{\mathrm{z}}$ be zero, then exact pricing (3) holds, and no anomaly affects the cross-section of expected returns.[8]

Using (5), the asset pricing relationship in (4) becomes

$$\mathrm{E}\left[R_{it} | I_{t-1}, \boldsymbol{\Pi}\right] = \gamma_{0,t-1} + \boldsymbol{\gamma}_{\mathrm{f,t-1}}' \boldsymbol{\beta}_{i,t-1} + \boldsymbol{\gamma}_{\mathrm{z,t-1}}' \mathbf{z}_{i,t-1}. \tag{6}$$

The expression in (6) represents our new asset pricing restriction.[9] It is worth noticing that, while allowing for anomalies, condition (6) does not necessarily represent a deviation from no-arbitrage, but only from exact pricing (see Proposition OA.5 in Section OA.9 of the Online Appendix for more details).

Now, under (6), the asset pricing model in (2) generalizes to

$$\mathbf{R}_t = \gamma_{0,t-1} \mathbf{1}_N + \mathbf{Z}_{t-1} \boldsymbol{\gamma}_{\mathrm{z,t-1}} + \mathbf{B}_{t-1} \boldsymbol{\delta}_{\mathrm{f,t-1}} + \boldsymbol{\epsilon}_t, \tag{7}$$

where $\mathbf{1}_N$ denotes a $N \times 1$ vector of ones, $\mathbf{Z}_{t-1} = (\mathbf{z}_{1,t-1}, \cdots, \mathbf{z}_{N,t-1})'$ represents the $N \times K_{\mathrm{z}}$ matrix of anomalies at time $t-1$, and where we set

$$\boldsymbol{\delta}_{\mathrm{f,t-1}} \equiv \boldsymbol{\gamma}_{\mathrm{f,t-1}} + \mathbf{f}_t - \mathrm{E}\left[\mathbf{f}_t | I_{t-1}, \boldsymbol{\Pi}\right], \tag{8}$$

which we denominate as the vector of *ex-post* risk premia.[10] An important special case of (7) arises

[8] When imposing no-arbitrage, the APT (see Ross (1976) and Chamberlain and Rothschild (1983), among others) imposes the following constraint on the pricing errors $\mathbf{a}_{t-1} = (a_{1,t-1}, \cdots, a_{N,t-1})'$,

$$\mathbf{a}_{t-1}' [\mathrm{Var}_{t-1}(\boldsymbol{\epsilon}_t)]^{-1} \mathbf{a}_{t-1} \leq \delta_{t-1} < \infty,$$

for some (unknown) finite quantity δ_{t-1}. Our main analysis will abstract from any consideration associated with no-arbitrage (see Section OA.9 of the Online Appendix for some details).

[9] Condition (6) implies that the agent has full information on the anomaly variables $\mathbf{z}_{i,t-1}$ for every stock. If one suspects that the agent's information is not complete (for example, because firm's balance sheet data is released less frequently or with delays), then the asset pricing restriction (6) generalizes to $\mathrm{E}\left[R_{i,t} | I_{t-1}, \boldsymbol{\Pi}\right] = \gamma_{0,t-1} + \boldsymbol{\gamma}_{\mathrm{f,t-1}}' \boldsymbol{\beta}_{i,t-1} + \boldsymbol{\gamma}_{\mathrm{z,t-1}}' \mathrm{E}\left[\mathbf{z}_{i,t-1} | I_{t-1}, \boldsymbol{\Pi}\right]$, and all our arguments continue to be valid.

[10] The notion of *ex-post* risk premia was originally coined by Shanken (1992) to denote a noisy version of the ex-ante risk premia $\boldsymbol{\gamma}_{\mathrm{f},t-1}$ due to the unexpected factor outcomes $\mathbf{f}_t - \mathrm{E}\left[\mathbf{f}_t | I_{t-1}, \boldsymbol{\Pi}\right]$, arising whenever one considers the fixed-T case. Shanken (1992) considers the time-average of $\boldsymbol{\delta}_{\mathrm{f,t-1}}$ with constant premia and expected value, yielding $\boldsymbol{\delta}_{\mathrm{f}} \equiv \boldsymbol{\gamma}_{\mathrm{f}} + \bar{\mathbf{f}} - \mathrm{E}[\mathbf{f}_t]$, whereas here we consider the time-varying ex-post risk premia (8), with a slight abuse of notation as $\boldsymbol{\delta}_{\mathrm{f,t-1}}$ contains $\mathbf{f}_t$, justified by latency of $\boldsymbol{\delta}_{\mathrm{f,t-1}}$.

when the risk factors represent returns of traded portfolios, in which case one simply replaces $\gamma_{0,t-1}$ with the gross risk-free rate ($R_{\mathrm{f},t-1}$), and sets $\boldsymbol{\gamma}_{\mathrm{f},t-1} = \mathrm{E}\left[\mathbf{f}_t | I_{t-1}, \boldsymbol{\Pi}\right] - \gamma_{0,t-1}\mathbf{1}_{K_\mathrm{f}}$.

Whenever the vector of anomalies' premia $\boldsymbol{\gamma}_{\mathrm{z,t-1}}$ in (7) is non-zero, we say that the anomalies *affect* (or, *are priced* in) the cross-section of expected returns through (6). The objective of this paper is to provide a formal methodology to estimate the anomalies' premia $\boldsymbol{\gamma}_{\mathrm{z,t-1}}$ and test for their statistical significance, using the model specification in (7).

3.1 Risk Exposures: Time-Variation and Identification

To estimate the asset pricing model (7), it is essential to specify the form of time-variation of the loadings $\boldsymbol{\beta}_{i,t-1}$. Two different approaches can be considered in this case.

First, one could postulate a parametric specification of the loadings, such as, for example:

$$\boldsymbol{\beta}_{i,t-1} = \boldsymbol{\beta}_{0i} + \mathbf{B}_{1i}\mathbf{g}_{t-1} + \mathbf{B}_2\mathbf{z}_{it-1}, \tag{9}$$

for some matrices of coefficients $\boldsymbol{\beta}_{0i}$ ($K_\mathrm{f} \times 1$), $\mathbf{B}_{1i}$ ($K_\mathrm{f} \times K_\mathrm{g}$), and $\mathbf{B}_2$ ($K_\mathrm{f} \times K_\mathrm{z}$).

When $\mathbf{B}_2 = \mathbf{0}_{K_\mathrm{z} \times K_\mathrm{f}}$, it follows that only the K_g-dimensional vector of asset-invariant variables $\mathbf{g}_{t-1}$ is driving the loadings. In this case, it is well-known that one can re-express the conditional asset pricing model (7) as a model with constant loadings on ($K_\mathrm{f} + K_\mathrm{g}$) *rescaled* risk factors ($\mathbf{f}_t \otimes \mathbf{g}_{t-1}$), whit $\otimes$ denoting the Kronecker product (see Shanken (1990), Ferson and Harvey (1999), and Lettau and Ludvigson (2001), among others). When $\mathbf{B}_2 \neq \mathbf{0}_{K_\mathrm{z} \times K_\mathrm{f}}$, asset-specific characteristics also contribute to the time-variation of the loadings. This is discussed for example in Avramov and Chordia (2006) and Gagliardini, Ossola, and Scaillet (2016), where heterogeneity of the coefficients (with $\mathbf{B}_{2i}$ replacing $\mathbf{B}_2$ in (9)) and interactive effects of $\mathbf{g}_{t-1}$ and $\mathbf{Z}_{t-1}$ are also considered.

An alternative method consists in leaving the time-varying $\boldsymbol{\beta}_{i,t-1}$ model-free, with their practical estimation obtained through (short) rolling samples of size T (see Brennan, Chordia, and Subrahmanyam (1998), among others). Specifically, we formalize this approach by introducing the following smoothing assumption on the loadings.

Assumption 1 (*smoothness of the loadings*)**.** *For every $s = 1, \cdots, T-1$, there exists a locally-constant matrix $\mathbf{B}$, depending on the interval $\{1, \cdots, T\}$, such that*

$$\frac{(\mathbf{B}_s - \mathbf{B})'(\mathbf{B}_s - \mathbf{B})}{N} = o(N^{-\frac{1}{2}}). \tag{10}$$

When (10) holds, then model (7) can be replaced by a conditional asset pricing model with *locally-constant* risk exposures

$$\mathbf{R}_t = \gamma_{0,t-1}\mathbf{1}_N + \mathbf{Z}_{t-1}\boldsymbol{\gamma}_{\mathrm{z,t-1}} + \mathbf{B}\boldsymbol{\delta}_{\mathrm{f,t-1}} + \boldsymbol{\epsilon}_t, \tag{11}$$

without altering the statistical properties of both testing and model estimation.[11] Notice that the conditional asset pricing model (11) is still characterized by time-varying loadings, because in practice $\mathbf{B}$ would (non-parametrically) rely on the specific time period being considered.[12] We have chosen not to explicitly write this just to facilitate both the notation and exposition. Section OA.10 of the Online Appendix reports some numerical illustrations of relevant examples of time-varying betas that satisfy Assumption 1.

Since we consider T to be (almost) arbitrarily small in our analysis, Assumption 1 appears quite mild in practice and offers a significant advantage by remaining robust against any form of model misspecification. This stands in contrast to the parametric approach in (9), which could result in invalid inference whenever an incorrect specification of $\boldsymbol{\beta}_{i,t-1}$ is assumed, for example because $\mathbf{g}_{t-1}$ and/or $\mathbf{Z}_{t-1}$ are either mis-measured or incomplete.

For these reasons, in the following sections, we will operate under the assumption that (11) holds (by means of Assumption 1), thereby examining conditional asset pricing models with locally-constant loadings. However, it is worth noting that our analysis can easily extend to the scenario of model (7) with parametric time-varying loadings as in (9) if deemed more appropriate.

3.1.1 Identification Condition

When taking into account the locally-constant asset pricing model (11), an identification issue may arise. Indeed, the specification in (11) can be *observationally equivalent* to a model in which $\boldsymbol{\gamma}_{\mathrm{z,t-1}} = \mathbf{0}_{K_z}$ (that is, exact pricing holds), but having time-varying risk exposures, with their time-variation driven by the anomalies $\mathbf{Z}_{t-1}$. To see this, consider the following exact-pricing model with time-varying betas $\mathbf{B}_{t-1} = \mathbf{B} + \mathbf{Z}_{t-1}\mathbf{B}_2'$:

[11] Formally, one obtains that (7) can be rewritten as $\mathbf{R}_t = \gamma_{0,t-1}\mathbf{1}_N + \mathbf{Z}_{t-1}\boldsymbol{\gamma}_{\mathrm{z,t-1}} + \mathbf{B}\boldsymbol{\delta}_{\mathrm{f,t-1}} + \boldsymbol{\epsilon}_t + (\mathbf{B}_{t-1} - \mathbf{B})\boldsymbol{\delta}_{\mathrm{f,t-1}}$. Then, under Assumption 1, this model would have the same statistical properties (in terms of both testing and estimation) of model (11). See Section OA.3 of the Online Appendix for details (Lemma OA.10).

[12] Formally, one can define a generic time window of length T as $\mathcal{T}_t = t - T + 1, \ldots, t - 1$. Then, given t, $\mathbf{B}$ implicitly refers to $\mathbf{B}_t = \mathbf{B}(\mathcal{T}_t)$, which varies across different data windows but remains constant within the same window.

$$\mathbf{R}_t = \gamma_{0,t-1}\mathbf{1}_N + \mathbf{B}_{t-1}\boldsymbol{\delta}_{\mathrm{f,t-1}} + \boldsymbol{\epsilon}_t$$
$$= \gamma_{0,t-1}\mathbf{1}_N + (\mathbf{B} + \mathbf{Z}_{t-1}\mathbf{B}_2')\boldsymbol{\delta}_{\mathrm{f,t-1}} + \boldsymbol{\epsilon}_t. \quad (12)$$

It is easy to see that (12) yields exactly (11), whenever $\boldsymbol{\gamma}_{\mathrm{z,t-1}}$ in (11) satisfies:

$$\boldsymbol{\gamma}_{\mathrm{z},t-1} = \mathbf{B}_2'\boldsymbol{\delta}_{\mathrm{f},t-1}. \quad (13)$$

The validity of the restriction in (13) can be empirically verified. Once model (7) is estimated, the time series of the estimated $\boldsymbol{\gamma}_{\mathrm{z},t-1}$ and $\boldsymbol{\delta}_{\mathrm{f},t-1}$ become available; one can then analyze the R-squared of the regression of $\boldsymbol{\gamma}_{\mathrm{z},t-1}$ on $\boldsymbol{\delta}_{\mathrm{f},t-1}$ or the canonical correlations between $\boldsymbol{\gamma}_{\mathrm{z},t-1}$ and $\mathbf{B}_2'\boldsymbol{\delta}_{\mathrm{f},t-1}$.[13] This evaluation will be incorporated into our empirical analysis in Section 10, to ensure the correct identification of anomalies' impact on pricing errors.

4 Two-Pass Methodology for Anomalies: Conventional Approach

The most common and intuitive approach to test for the presence of anomalies is based on the estimation of model (11) by means of the two-pass Fama and MacBeth (1973) regression. It first entails obtaining the estimated matrix of loadings $\hat{\mathbf{B}}$ from (2) through time-series OLS regressions (one for each asset) of asset returns on observed risk factors $\boldsymbol{f}_t$, and then estimating the premia parameters ($\gamma_{0,t-1}, \boldsymbol{\delta}_{\mathrm{f,t-1}}$, and $\boldsymbol{\gamma}_{\mathrm{z,t-1}}$) through CSR OLS (one for each period of time), using $\hat{\mathbf{B}}$ in (11).

However, recognizing that inference would necessarily be affected by the error-in-variable (EIV) problem due to the use of $\hat{\mathbf{B}}$ in (11) (see Shanken (1992)), Fama and French (2008) advocate estimation of the anomalies' premia by simple OLS cross-sectional regressions (one for each period of time) of $\mathbf{R}_t$ on $\mathbf{Z}_{t-1}$ and an intercept, hence excluding the estimated $\mathbf{B}$ from (11), yielding the *time-varying* anomaly premium estimator

$$\tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1} \equiv (\mathbf{Z}_{t-1}'\mathbb{M}_{1_N}\mathbf{Z}_{t-1})^{-1}\mathbf{Z}_{t-1}'\mathbb{M}_{1_N}\mathbf{R}_t, \quad (14)$$

[13]When (13) holds, the canonical correlations between $\boldsymbol{\gamma}_{\mathrm{z},t-1}$ and $\mathbf{B}_2'\boldsymbol{\delta}_{\mathrm{f},t-1}$ are all equal to one in population. Note that when $K_{\mathrm{z}} > K_{\mathrm{f}}$ (with full row-rank $\mathbf{B}_2$), one obtains that $\boldsymbol{\delta}_{\mathrm{f},t-1} = \mathbf{B}_2^*\boldsymbol{\gamma}_{\mathrm{z},t-1}$ for $\mathbf{B}_2^* = (\mathbf{B}_2\mathbf{B}_2')^{-1}\mathbf{B}_2$. Therefore, one could also assess the validity of (13) using the R-squared of the multivariate regression of $\boldsymbol{\delta}_{\mathrm{f},t-1}$ on the vector $\boldsymbol{\gamma}_{\mathrm{z},t-1}$. Formal tests for canonical correlations exist (see Knapp (1978)), although their significance should be considered only approximately valid as the premia parameters are estimated.

where $\mathbb{M}_{1_N} \equiv \mathbf{I}_N - \mathbf{1}_N\mathbf{1}_N'/N$ is used to de-mean the data, with $\mathbf{I}_N$ denoting an identity matrix of dimension N. This implies that $\mathbb{M}_{1_N}\mathbf{R}_t = \mathbf{R}_t - \mathbf{1}_N\bar{R}_t$, with $\bar{R}_t \equiv \sum_{i=1}^N R_{it}/N$ denoting the cross-sectional sample average of returns. Similarly, $\mathbb{M}_{1_N}\mathbf{Z}_{t-1} = \mathbf{Z}_{t-1} - \mathbf{1}_N\bar{\mathbf{Z}}_{t-1}'$, setting $\bar{\mathbf{Z}}_{t-1} \equiv \sum_{i=1}^N \mathbf{z}_{i,t-1}/N$. Fama and French (2008) justify the approach in (14) by recognizing that $\tilde{\gamma}_{\mathrm{z},t-1}$ is equivalent to the two-pass estimator applied to (11), whenever the loadings $\mathbf{B}$ and the anomalies $\mathbf{Z}_{t-1}$ in (11) are orthogonal to each other, an assumption claimed to hold empirically. This orthogonality condition is implied when the loadings are cross-sectionally invariant.

Inference is typically carried out in terms of the *average* premium, taking the time-series average of the premia estimates $\tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1}$ in (14). This yields the conventional *average premium* estimator

$$\bar{\tilde{\boldsymbol{\gamma}}}_{\mathrm{z}} \equiv \frac{1}{T-1}\sum_{t=2}^{T}\tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1} \tag{15}$$

for which the corresponding t-ratios is evaluated. To illustrate, consider the case of univariate regressions (i.e. $K_{\mathrm{z}} = 1$). In this case, the t-ratio of the average premium associated to the z-th anomaly is simply

$$t_{\mathrm{z}} \equiv \frac{\bar{\tilde{\gamma}}_{\mathrm{z}}}{\sqrt{\tilde{\Sigma}_{\gamma_{\mathrm{z}}}/(T-1)}}, \tag{16}$$

where $\tilde{\Sigma}_{\gamma_{\mathrm{z}}}$ is the sample variance of the CSR OLS estimates $\tilde{\gamma}_{\mathrm{z},t-1}$, namely:

$$\tilde{\Sigma}_{\gamma_{\mathrm{z}}} = \frac{1}{T-1}\sum_{t=2}^{T}(\tilde{\gamma}_{\mathrm{z},t-1} - \bar{\tilde{\gamma}}_{\mathrm{z}})^2. \tag{17}$$

The t-ratio in (16) is then compared with the critical values of the standard Normal distribution, conjecturing that the inference on $\bar{\tilde{\boldsymbol{\gamma}}}_{\mathrm{z}}$ is valid as $T \to \infty$. We denote this approach as the *conventional approach*.

Given the extensive use of the conventional approach in empirical studies (see Fama and French (2008) and Hou, Chen, and Zhang (2020), among others), it seems essential to understand the inferential properties of both the time-varying estimator in (14) and the average estimator in (15), as well as their ability to capture time-variation in the (true) premia, and the potential consequences of omitting factors' loadings from the estimation of model (11).

We now show that the statistical validity of the conventional approach is not always warranted, unless extremely strict conditions are applied. In particular, we show below that the asymptotic properties of the conventional approach crucially depend on the sampling scheme under consideration, namely the relative magnitude of N and T. Moreover, regardless of the adopted sampling scheme, the conventional t-ratios are never appropriate whenever one faces a model with time-varying premia parameters, making standard inference seriously problematic. To show our results, throughout this section for simplicity we assume that $K_{\mathrm{z}} = 1$ and consider three different sampling schemes: (i) the large-T–fixed-N case, (ii) the large-N–fixed-T case, and (iii) the large-T–large-N case. Formal derivations of the following results, including the generalization to the case of $K_{\mathrm{z}} > 1$, are reported in the Online Appendix OA.6.

Let us consider first the case (i) of $T \to \infty$ with fixed N. This situation applies, for example, when one uses a panel consisting of a small number of portfolios, for which a long time-series of data is available. As N is kept fixed in this sampling scheme, it follows that no asymptotic properties can be established for the time-varying estimator $\tilde{\gamma}_{t-1,z}$ in (14). One can only assert the unbiasedness of the estimator (14), which can be established only under some regularity conditions that include, among others, the *finite-N orthogonality* condition:

$$\mathbf{Z}'_{t-1}\mathbb{M}_{1_N}\mathbf{B} = \mathbf{0}_{N\times K_{\mathrm{f}}}, \tag{18}$$

namely the (in sample) cross-sectional orthogonality between factor betas and the anomaly variable $\mathbf{Z}_{t-1}$, with $\mathbf{0}_{N\times K_{\mathrm{f}}}$ representing the zero matrix of dimension $N \times K_{\mathrm{f}}$.

Under the same sampling scheme, instead, the average premium estimator (15) satisfies:

$$\bar{\tilde{\gamma}}_{\mathrm{z}} \to_p \bar{\gamma}^0_{\mathrm{z}} \equiv \lim_{T\to\infty} \bar{\gamma}_{\mathrm{z}}, \text{ with } \bar{\gamma}_{\mathrm{z}} \equiv \frac{1}{T-1}\sum_{t=2}^{T} \gamma_{\mathrm{z},t-1} \tag{19}$$

It follows that $\bar{\tilde{\gamma}}_{\mathrm{z}}$ converges to a constant quantity, $\bar{\gamma}^0_{\mathrm{z}}$, which we refer to as the *long-run* anomaly premium. Alternatively, (19) tells us that $\bar{\tilde{\gamma}}_{\mathrm{z}}$ consistently estimates the *constant* premium γ_{z}, whenever $\gamma_{\mathrm{z},t-1} = \gamma_{\mathrm{z}}$, for every $t = 1, .., T-1$. It is important to note that the results in (19) are valid under some regularity conditions, including again the orthogonality condition in (18). Moreover, under some further regularity conditions (See Theorem OA.4 of the Online Appendix OA.6.1), as $T \to \infty$ and N is fixed, $\bar{\tilde{\gamma}}_{\mathrm{z}}$ is also asympotically normally distributed, such that

$$\sqrt{T}\,(\bar{\tilde{\gamma}}_{\mathrm{z}} - \bar{\gamma}_{\mathrm{z}}) \to_d \mathcal{N}\,(0, V_N)\,,$$

with V_N denoting the large-T asymptotic variance of the estimator, and where the subscript N is used to remark its dependency on the N-dimension as well. To conduct inference, one needs to consistently estimate V_N, which is typically done in the literature by using the the sample variance $\tilde{\Sigma}_{\gamma_z}$ of the CSR OLS estimates, as defined in (17). However, we show that $\tilde{\Sigma}_{\gamma_z}$ can only work in the case where the true anomaly premium is assumed to be time-invariant, i.e., when one assumes that $\gamma_{z,t} = \gamma_z$ for every t in (11). More formally, we show that

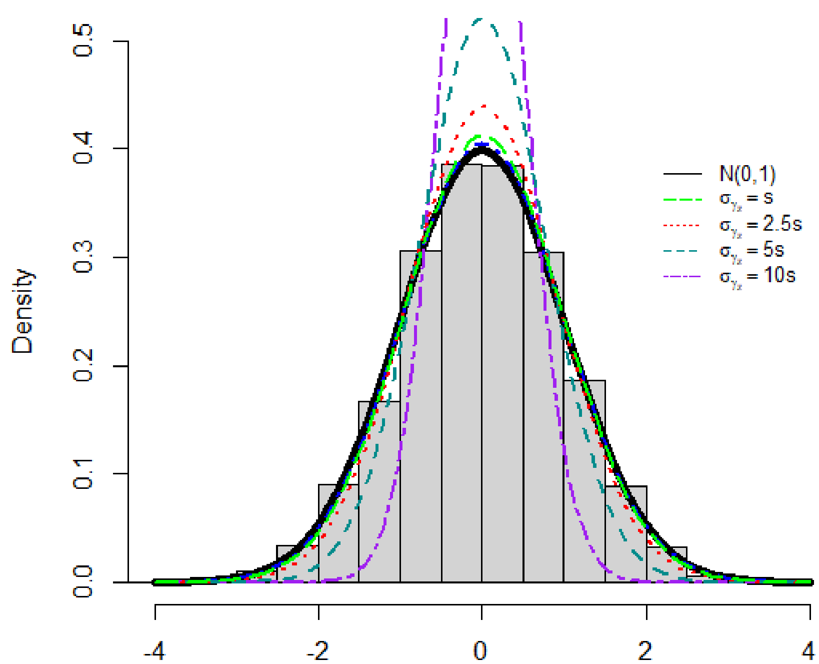
$$\tilde{\Sigma}_{\gamma_z} \to_p \sigma^2_{\gamma_z} + V_N, \qquad \text{with} \quad \sigma^2_{\gamma_z} \equiv \lim_{T\to\infty} \frac{1}{(T-1)} \sum_{t=2}^{T} (\gamma_{z,t-1} - \bar{\gamma}_z)^2. \tag{20}$$

From (20), it is immediate to see that $\tilde{\Sigma}_{\gamma_z}$ will consistently estimate V_N only when $\sigma^2_{\gamma_z} \equiv 0$, which happens if, and only if, $\gamma_{z,t-1} = \gamma_z$ for every $t = 2, ..., T$. Whenever this condition is violated, then $\sigma^2_{\gamma_z}$ will be a positive quantity, implying that the t-ratio in (16) is downward biased. In other words, whenever one assumes that the true premia in (11) are time-varying and uses the conventional t-ratio in (16) to make inference on the average anomaly premium, then one tends to under-reject the null hypothesis of zero (long-run) premium than prescribed by the chosen nominal size. Therefore, a statistically significant t-ratio could provide a strong indication of a non-zero average premium, even though it leaves inference undetermined when it is found to be not significant. This is an important and crucial result, which could invalidate or raise doubts on many of the findings established in the empirical literature on anomalies.

To demonstrate the potential implications of this result, we consider a simple simulation exercise, where the true anomaly premium has been generated using a time-varying scheme. Specifically, using $N = 25$ and $T = 360$, we simulate B=2,000 samples of asset returns, using the data generating process $\mathbf{R}_t = \gamma_{0,t-1}\mathbf{1}_N + \mathbf{Z}_{t-1}\gamma_{\mathrm{z,t-1}}\boldsymbol{\epsilon}_t$, where $K_z = 1$ and $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}_N, \sigma^2_\epsilon \mathbf{I}_N)$, with $\sigma^2_\epsilon = 0.1$. For simplicity, we set $\gamma_{0,t-1} = \gamma_0 = 0$, while $\gamma_{\mathrm{z,t-1}}$ has been generated using an AR(1) model $\gamma_{\mathrm{z},t-1} = \mu_z(1-\phi_z) + \phi_z\gamma_{\mathrm{z},t-1} + u_z$, with $u_z \sim \mathcal{N}(0, \sigma^2_u)$. This implies that the variance $\sigma^2_{\gamma_z}$ in (20) is equivalent to σ^2_u. The parameters μ_z and ϕ_z have been calibrated by fitting an AR(1) model on the estimated time series of $\tilde{\gamma}_{z,t}$, obtained by regressing observed monthly returns $\mathbf{R}_t$ on the book leverage anomaly variable $\mathbf{Z}_{t-1}$, while for σ^2_u we consider different increasing values, from $\sigma^2_u = s$, up to $\sigma^2_u = 10s$, with $s = 0.001$. Then, for each simulated sample, and for each different value of σ^2_u, we estimate the anomaly average premium with (15) and construct the corresponding t-ratio in (16), which we plot in Figure 1. The figure clearly shows the inferential consequences of time-varying premia. When there is very little time variation, the classical approach works quite well (see the light green dotted curve). As the time-variation (i.e., the variance) of $\gamma_{\mathrm{z},t-1}$

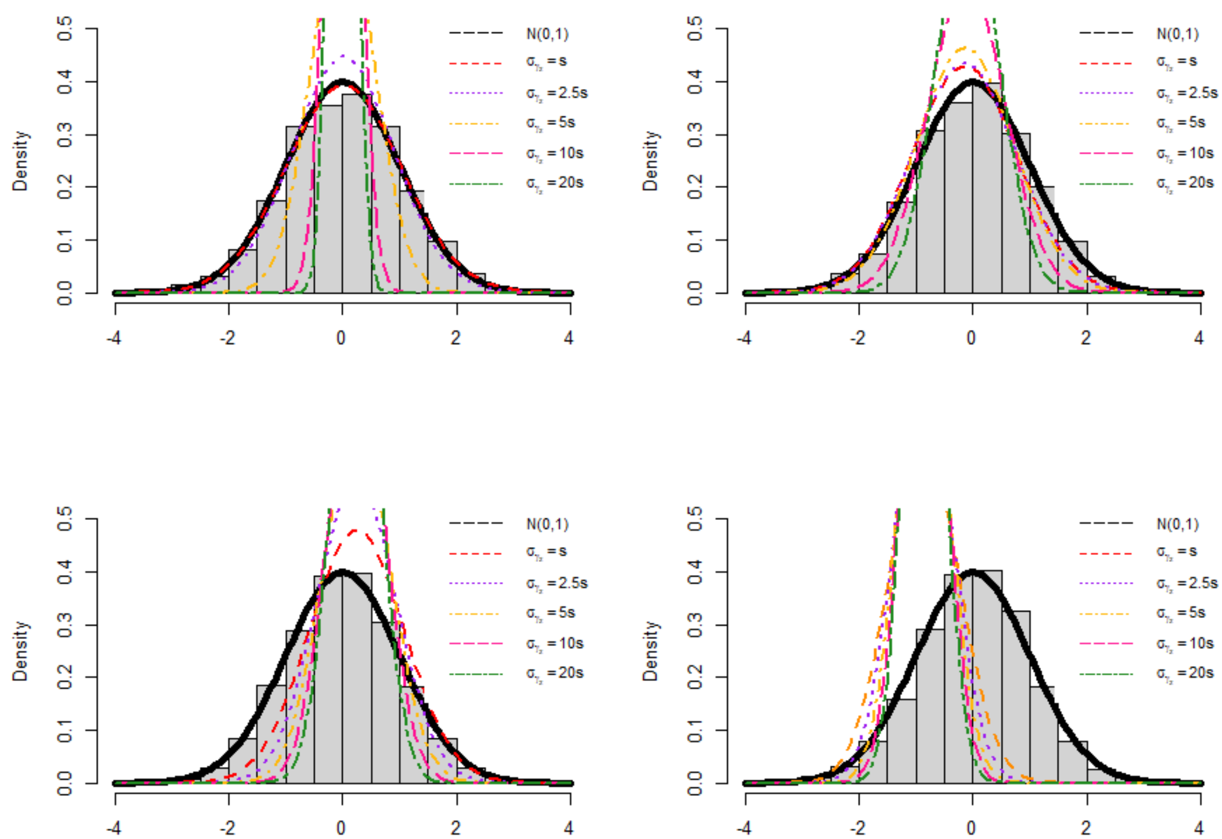
increases, the distribution of the corresponding t-ratio departs substantially from the standard normal distribution, pointing to a severe under-rejection.

Figure 1: **Conventional t-ratios under a time-varying setting.** The figure shows the distribution of the conventional t-ratios in (16), when the true anomaly premium $\gamma_{\mathrm{z,t-1}}$ follows a time-varying process. Specifically, using $N = 25$ and $T = 360$, we simulate B=2,000 samples of asset returns, using the data generating process $\mathbf{R}_t = \gamma_{0,t-1}\mathbf{1}_N + \mathbf{Z}_{t-1}\gamma_{\mathrm{z,t-1}}\boldsymbol{\epsilon}_t$, where $K_{\mathrm{z}} = 1$ and $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}_N, \sigma^2_\epsilon \mathbf{I}_N)$, with $\sigma^2_\epsilon = 0.1$. For simplicity, we set $\gamma_{0,t-1} = \gamma_0 = 0$, while $\gamma_{\mathrm{z,t-1}}$ has been generated using an AR(1) process $\gamma_{\mathrm{z},t-1} = \mu_z(1-\phi_z) + \phi_z\gamma_{\mathrm{z},t-1} + u_z$, with $u_z \sim \mathcal{N}(0, \sigma^2_u)$. This implies that the variance $\sigma^2_{\gamma_{\mathrm{z}}}$ in (20) is equivalent to $\sigma^2_u/(1-\phi^2_z)$. The parameters σ^2_u, μ_z and ϕ_z have been calibrated by fitting an AR(1) process on the estimated time series of $\tilde{\gamma}_{z,t}$, obtained by regressing observed monthly returns $\mathbf{R}_t$ on the book leverage anomaly variable $\mathbf{Z}_{t-1}$. Then, for each simulated sample, we estimate the anomaly average premium using the conventional estimator in (15) and construct the corresponding t-ratio in (16). We then plot the distribution of the B=2,000 t-ratios and repeat the same exercise for increasing values of σ^2_u. Monthly returns are from the Center for Research in Security Prices (CRSP), while data on the anomaly variables are provided by Chen and Zimmermann (2019).



The results presented above have clearly strong inferential implications which, however, provide only a partial view of the overall picture. In fact, the previous exercise assumes that the true model contains only the anomaly variables, thus excluding the estimated $\mathbf{B}$ from the return generating process. This would coincide with the two-pass estimator applied to (11), whenever the loadings $\mathbf{B}$ and the anomalies $\mathbf{Z_{t-1}}$ in (11) are orthogonal to each other. Whenever this assumption is not satisfied, the accuracy of the results could be even more compromised. The inferential consequences of excluding $\mathbf{B}$ from the estimated model are presented in Figure 2. The figure depicts the outcomes of a simulation exercise where now the true return generating process follows the model in (11), but where $\gamma_{z,t-1}$ is still estimated using (14) - hence omitting the loadings $\mathbf{B}$. Specifically, using the same parameters of the above exercise with $K_{\mathrm{f}} = 1$, we generate asset returns using the process $\mathbf{R}_t = \gamma_{0,t-1}\mathbf{1}_N + \mathbf{Z}_{t-1}\boldsymbol{\gamma}_{z,t-1} + \mathbf{B}\boldsymbol{\delta}_{\mathrm{f},t-1} + \boldsymbol{\epsilon}_{t-1}$, where $\boldsymbol{\delta}_{\mathrm{f},t-1}$ and $\mathbf{B}$ have been calibrated using data on the market factor and its loadings on observed monthly returns $\mathbf{R}_t$, respectively. To account for different degrees of correlation between $\mathbf{B}$ and $\mathbf{Z}_{t-1}$, we define the anomaly variable $\mathbf{Z}_{t-1} = [\theta\mathbb{M}_{\mathrm{B}} + (1-\theta)\mathbb{P}_{\mathrm{B}}]\tilde{\mathbf{Z}}_{t-1}$, where $\tilde{\mathbf{Z}}_{t-1}$ has been calibrated using firms' book leverage data and where we set $\mathbb{P}_{\mathrm{B}} = \mathbf{B}(\mathbf{B}'\mathbf{B})^{-\mathbf{1}}\mathbf{B}$, and $\mathbb{M}_{\mathrm{B}} = \boldsymbol{I}_{\mathrm{N}} - \mathbb{P}_{\mathrm{B}}$. The parameter θ ranges between 0 and 1, where $\theta = 0$ represents the case of perfect correlation between $\mathbf{B}$ and $\mathbf{Z}_{t-1}$, while $\theta = 1$ indicates no correlation between the loadings and the anomaly variable. In our experiment, we consider different degrees of correlation by setting $\theta = \{1, 0.75, 0.25, 0\}$. As before, for each of the $B = 2,000$ simulated samples and for each different value of θ and σ_u^2, we estimate the average anomaly premium as in (15) and construct the corresponding t-ratios defined in (16), which we then plot in Figure 2. Each panel in Figure 2 corresponds to a different value of the parameter θ, namely $\theta = 1$ (top-left panel), $\theta = 0.75$ (top-right panel), $\theta = 0.25$ (bottom-left panel), and $\theta = 0$ (bottom-right panel). As expected, when $\theta = 1$, we re-obtain the same results of Figure 1, confirming the fact that the estimator in (14) coincides with the conventional two-pass estimator applied to (11), whenever $\mathbf{B}$ and $\mathbf{Z}_{t-1}$ are orthogonal to each other. However, as the correlation between the anomaly variable and the loadings increases, the estimation bias becomes more pronounced and combines with the downward bias arising from time variation in the anomaly premium process.

Figure 2: The figure shows the outcomes of a simulation exercise where the true return generating process follows the model in (11), but where $\gamma_{z,t-1}$ is estimated using (14) - hence omitting the loadings $\mathbf{B}$. Specifically, using the same parameters of the exercise described in Figure 1 with $K_{\mathrm{f}} = 1$, asset returns have generated using the process $\mathbf{R}_t = \gamma_{0,t-1}\mathbf{1}_{\mathrm{N}} + \mathbf{Z}_{t-1}\boldsymbol{\gamma}_{\mathrm{z,t-1}} + \mathbf{B}\boldsymbol{\delta}_{\mathrm{f},t-1} + \boldsymbol{\epsilon}_{t-1}$, where $\boldsymbol{\delta}_{\mathrm{f},t-1}$ and $\mathbf{B}$ have been calibrated using data on the market factor and its loadings on observed monthly returns $\mathbf{R}_t$, respectively. To account for different degrees of correlation between $\mathbf{B}$ and $\mathbf{Z}_{t-1}$, the anomaly variable has been generate as $\mathbf{Z}_{t-1} = [\theta\mathbb{M}_{\mathrm{B}} + (1-\theta)\mathbb{P}_{\mathrm{B}}]\tilde{\mathbf{Z}}_{\mathbf{t-1}}$, where $\tilde{\mathbf{Z}}_{\mathbf{t-1}}$ has been calibrated using firms' book leverage data and where we set $\mathbb{P}_{\mathrm{B}} = \mathbf{B}(\mathbf{B}'\mathbf{B})^{-\mathbf{1}}\mathbf{B}$, and $\mathbb{M}_{\mathrm{B}} = \mathbf{I}_{\mathrm{N}} - \mathbb{P}_{\mathrm{B}}$. The parameter θ ranges between 0 and 1, where $\theta = 0$ represents the case of perfect correlation between $\mathbf{B}$ and $\mathbf{Z}_{t-1}$, while $\theta = 1$ indicates no correlation between the loadings and the anomaly variable. The experiment considers different degrees of correlation, setting $\theta = \{1, 0.75, 0.25, 0\}$. Then, for each of the $B = 2,000$ simulated samples and for each different value of θ and σ_u^2, the average anomaly premium is estimated using (15) and the corresponding t-ratios defined in (16) have been plotted. Each panel of the figure corresponds to a different value of the parameter θ, namely $\theta = 1$ (top-left panel), $\theta = 0.75$ (top-right panel), $\theta = 0.25$ (bottom-left panel), and $\theta = 0$ (bottom-right panel). Monthly returns are from the Center for Research in Security Prices (CRSP), while data on the anomaly variables are provided by Chen and Zimmermann (2019).



Let us now consider the case of estimating (11) using the time-varying estimator in (14) and the average estimator in (15), when now $N \to \infty$ with fixed T. This situation commonly arises when one uses data on the thousands of individual stock returns - rather than portfolios - over short time windows. Under suitable regularity conditions, and assuming the *large-N orthogonality* condition

$$\frac{\mathbf{Z}'_{t-1}\mathbb{M}_{1_N}\mathbf{B}}{N} \to_p \mathbf{0}_{N\times K_{\mathrm{f}}}, \tag{21}$$

then, the time-varying estimator in (14) and the average estimator in (15) satisfy:

$$\tilde{\gamma}_{\mathrm{z},t-1} \to_p \gamma_{\mathrm{z},t-1} \quad \text{and} \quad \bar{\tilde{\gamma}}_{\mathrm{z}} \to_p \bar{\gamma}_{\mathrm{z}}. \tag{22}$$

The results in (22) imply that the time-varying estimator (14) is now able to capture the true time-varying anomaly premium, with the average estimator in (15) now converging to the *local* average premium, defined over a fixed (and possibly small) time window of length T. Moreover, when the condition in (21) is replaced by the stronger assumption in (18) - namely when the cross-sectional orthogonality condition between the factor betas and the anomaly holds in sample - we get

$$\begin{aligned} \sqrt{N}\left(\tilde{\gamma}_{\mathrm{z},t-1} - \gamma_{\mathrm{z},t-1}\right) &\to_d \mathcal{N}\left(0, V_{t-1}\right), \text{ and} \\ \sqrt{N}\left(\bar{\tilde{\gamma}}_{\mathrm{z}} - \bar{\gamma}_{\mathrm{z}}\right) &\to_d \mathcal{N}\left(0, \bar{V}\right), \quad \text{with} \quad \bar{V} = \frac{1}{(T-1)^2}\sum_{t=2}^{T} V_{t-1} \end{aligned}$$

where V_{t-1} denotes the large-N asymptotic variance of the time-varying estimator, and where we use the subscript $t-1$ to emphasize its time dependence. However, in this large-N–fixed-T setting, inference based on conventional t-ratios becomes even more problematic than the previous large-T case, for both the time-varying and the average estimators. Indeed, the finite-T sampling scheme implies that

$$\tilde{\Sigma}_{\gamma_{\mathrm{z}}} \to_p \frac{1}{(T-1)}\sum_{t=2}^{T}\left(\gamma_{\mathrm{z},t-1} - \bar{\gamma}_{\mathrm{z}}\right)^2, \tag{23}$$

which is now a positive constant that could be, in general, bigger or smaller than $\bar{V}$, making any conclusion on the over- or under-rejection of the t-ratio in (16) impossible. Moreover, notice that the conventional t-ratios would involve the incorrect $\sqrt{T}$-normalization, rather than $\sqrt{N}$, even though this would be easy to rectify. Therefore, under the large-N–fixed-T sampling scheme, except for the special circumstance when condition (21) holds, the two conventional estimators in (14) and (15) could not be used to estimate the time-varying premia in (11) and its time-average, respectively.

Moreover, a new inferential theory would be needed in this case, to equip the results with correct standard errors and t-ratios. Filling this gap is one of the objective of this paper.

Finally, let us consider the case where both N and T are allowed to diverge. Under this setting, it is easy to show that the time-varying estimator $\tilde{\gamma}_{z,t-1}$ in (14) maintains the same identical behavior of the large-N–fixed-T case discussed above, so we omit the discussion to avoid repetition. Instead, for the average estimator, we get

$$\sqrt{NT}(\bar{\tilde{\gamma}}_z - \bar{\gamma}_z) \to_d \mathcal{N}\left(0, \bar{\mathcal{V}}\right), \tag{24}$$

where $\bar{\mathcal{V}}$ denotes the large-(N,T) asymptotic variance of the average estimator, such that $(T-1)^{-1}\sum_{t=2}^{T} V_{t-1} \to_p \bar{\mathcal{V}}$. Notice that, in this case, the average estimator $\bar{\tilde{\gamma}}_z$ converges at the fast rate $O(\sqrt{NT})$ to the long-run risk premium. As for the previous case, inference remains still problematic if one uses conventional t-ratios based on $\tilde{\Sigma}_{\gamma_z}$.[14]

To summarize, our results show that the conventional approach is unable to capture and make inference on time-varying premia, whenever $T \to \infty$ and N is kept fixed. That is, in a model with time-varying anomaly premia as in (11), one can only hope to consistently estimate the (long-run) average anomaly premia $\bar{\gamma}_z$, but not the anomaly premia at each point in time $\gamma_{z,t}$. Inference is even more complicated in this setting, with the conventional t-ratio of the average premium being downward biased, hence making standard inferential results potentially invalid. Only in the special case of time-constant anomaly premia, then the conventional approach works, even though it would still require stringent assumptions.

Under the large-N–fixed-T setting, the conventional time-varying estimator in (14) could in principle be used to consistently estimate time-varying anomaly premia, even though the validity of this result requires that the stringent orthogonality condition (21) holds in the data. At any rate, conventional t-ratios (of both the time-varying and the average premium estimators) are not valid, rendering all the inferential results potentially highly misleading. The same conclusions hold if one considers the double-asymptotic setting, where both N and T jointly diverge. In this respect, our paper offers an important contribution to the literature to fill this gap.

Indeed, exploiting the large-N–fixed-T setting, we show below how it is possible to *adjust* the

[14]In this case, it is possible to show that inference could be carried out if one further assumes that $\boldsymbol{B} = \mathbf{0}_{N\times K}$, that is if none of the risk factors in the model is correlated with the test assets' returns. See Remark OA.24 in the Online Appendix OA.6.3 for formal derivations.

conventional time-varying estimators (14) and (15), and make them working under the presence of estimated betas in model (11) - hence resolving the EIV problem - and relaxing any orthogonality assumption between factor loadings and anomalies such as (21). Moreover, we provide the limiting distribution of a new time-varying estimator, showing how to derive closed-form standard errors to conduct valid inference when N becomes large. Essentially, our aim is to propose a time-varying methodology which is simple and easy to implement, and which is based on the Fama and MacBeth (1973) two-pass principle, uncovering the required adjustments to make it work.

To conclude this section, we would like to give a quick preview of some important implications of our new time-varying methodology, by analysing the performance of six categories of anomalies, namely, *Momentum*, *Value versus Growth*, *Investment*, *Profitability*, *Intangibles* and *Trading Frictions*, as in Hou, Chen, and Zhang (2020). We report the main results in Table I. Specifically, we use monthly firm-level characteristics data provided by Chen and Zimmermann (2019), from January 1986 to December 2020 and perform monthly cross-sectional regressions of each anomaly variable on monthly returns from the Center for Research in Security Prices (CRSP) using both the conventional approach and our proposed approach (which we define as "RZ Approach" in Table I), described in Section 5 below. In this latter case, and contrary to the conventional approach, cross-sectional regressions also consider the market factor in the model specification. We then group each anomaly in one of the above six categories using the classification adopted in Hou, Chen, and Zhang (2020) and report the main results, averaged across categories.[15] We repeat the same exercise for different time lengths, from $T = 12$ up to $T = 360$ months, using monthly rolling widows. Then, for each category, and for both the two approaches, in Table I we report: (i) the average percentage of times that the category has been found to be significant (Panel A), (ii) the average $|t|$-statistics to test the null hypothesis that the anomaly premium is equal to zero (Panel B), and (iii) the average anomaly premium (Panel C).

The downward bias of the conventional approach clearly emerges from Table I, especially when T is relatively small. Indeed, for all the categories, the percentage of significance obtained by using the conventional approach is always subtantially lower than the one we found with our approach. Noticeably, the result is stable across T for our RZ approach, suggesting its validity, whereas it changes sharply for the conventional approach. This is also confirmed in Panel B, where we find

[15] A complete list of the anomaly variables in each category is provided in Appendix OA.11.

that the RZ approach is almost always associated with a higher average $|t|$-ratio. Interestingly, for all the categories and regardless of the time-series length, the two approaches also show different average values of the anomaly premium (Panel C), suggesting that the correlation between the estimated betas and anomalies could be actually different from zero, rendering the conventional approach estimates biased.

Table I: Conventional Approach versus the RZ time-varying approach

	% of significance - Conventional Approach						% of significance - RZ Approach					
Panel A	$T=12$	$T=36$	$T=72$	$T=120$	$T=240$	$T=360$	$T=12$	$T=36$	$T=72$	$T=120$	$T=240$	$T=360$
Momentum	19.48	26.99	35.36	45.01	54.53	60.00	71.07	77.77	73.22	71.89	62.11	59.44
Value VS Growth	15.04	19.67	23.97	31.19	46.09	54.76	43.87	52.36	54.27	57.07	62.61	62.18
Investment	20.19	37.79	52.50	64.50	89.59	92.00	39.38	36.87	48.58	63.92	71.11	64.61
Profitability	12.98	15.11	19.81	25.92	33.07	37.00	36.21	44.16	43.76	42.20	38.44	45.58
Intangibles	11.63	17.33	25.06	31.71	40.32	56.00	28.08	33.72	37.26	41.18	37.36	33.54
Trade Frictions	11.31	15.91	19.88	25.45	36.55	51.70	37.76	35.09	39.08	44.10	47.80	46.48
	average $\lvert t\rvert$ - Conventional Approach						average $\lvert t\rvert$ - RZ Approach					
Panel B	$T=12$	$T=36$	$T=72$	$T=120$	$T=240$	$T=360$	$T=12$	$T=36$	$T=72$	$T=120$	$T=240$	$T=360$
Momentum	2.73	2.80	3.00	3.38	4.19	5.18	8.07	11.59	10.98	9.29	7.28	7.87
Value VS Growth	2.59	2.60	2.67	2.86	3.18	3.69	4.15	5.21	5.94	5.74	5.17	5.26
Investment	2.76	2.77	3.14	3.37	3.75	4.57	3.56	3.71	4.60	5.00	4.98	5.04
Profitability	2.60	2.48	2.45	2.39	2.94	3.32	4.24	4.70	4.97	4.83	3.76	3.19
Intangibles	2.69	2.64	2.79	2.82	2.98	3.38	3.67	4.40	4.90	4.89	4.61	5.76
Trade Frictions	2.93	3.03	3.37	3.19	3.05	3.43	4.64	4.52	4.53	4.10	3.81	3.82
	average premia - Conventional Approach						average premia- RZ Approach					
Panel C	$T=12$	$T=36$	$T=72$	$T=120$	$T=240$	$T=360$	$T=12$	$T=36$	$T=72$	$T=120$	$T=240$	$T=360$
Momentum	0.35	0.26	0.23	0.21	0.19	0.19	0.20	0.23	0.21	0.18	0.16	0.15
Value VS Growth	0.33	0.22	0.18	0.17	0.17	0.15	0.42	0.40	0.42	0.40	0.38	0.36
Investment	0.22	0.18	0.17	0.17	0.17	0.16	0.32	0.26	0.31	0.39	0.43	0.42
Profitability	0.30	0.19	0.15	0.13	0.12	0.12	0.58	0.52	0.50	0.45	0.44	0.38
Intangibles	0.31	0.21	0.18	0.17	0.15	0.14	0.45	0.53	0.56	0.61	0.66	0.73
Trade Frictions	0.36	0.24	0.21	0.19	0.17	0.16	0.72	0.47	0.38	0.35	0.34	0.24

5 Anomalies with Time-Varying Premia: OLS-Based Estimation

The results of the previous section show that the conventional approach is not valid whenever one postulates time variation in the (true) anomalies' premia and unless strict orthogonality conditions are satisfied. We now introduce our new results, valid when $N \to \infty$ and T remains fixed, and show how all these challenges related to the conventional approach can be resolved, by means of a new OLS *bias-adjusted* estimator of the time-varying premia $\boldsymbol{\delta}_{\mathrm{f},t-1}$ and $\boldsymbol{\gamma}_{\mathrm{z},t-1}$. All the results are established under several regularity conditions and mild assumptions that we report in Appendix A.1.

Consider again the *conditional* asset pricing model (with locally-constant loadings) in (11), and rewrite it as

$$\mathbf{R}_t = \mathbf{Z}_{t-1}\boldsymbol{\gamma}_{\mathrm{z},t-1} + \mathbf{X}\boldsymbol{\Gamma}_{\mathrm{f},t-1} + \boldsymbol{\epsilon}_t \tag{25}$$

where $\mathbf{X} = (\mathbf{1}_N, \mathbf{B})$ and $\boldsymbol{\Gamma}_{\mathrm{f},t-1} = (\gamma_{0,t-1}, \boldsymbol{\delta}'_{\mathrm{f},t-1})'$, with $\boldsymbol{\delta}_{\mathrm{f},t-1}$ defined in (8). Since the matrix $\mathbf{X}$ in (25) is unknown, one needs first to estimate the loadings $\mathbf{B}$ to make the estimation of (25) feasible. The conventional two-pass approach typically advocates a simple OLS regression of $\mathbf{R}_t$ on an intercept and the observed risk factors $\mathbf{f}_t$, that is:

$$\hat{\mathbf{B}} \equiv \mathbf{R}'\mathbb{M}_{\mathbf{1}_{T-1}}\mathbf{F}(\mathbf{F}'\mathbb{M}_{\mathbf{1}_{T-1}}\mathbf{F})^{-1} = \mathbf{R}'\mathbf{P}, \tag{26}$$

where $\hat{\mathbf{B}} = (\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_N)'$, $\mathbf{F} = (\mathbf{f}_2, \cdots \mathbf{f}_T)'$, $\mathbf{R} = (\mathbf{R}_2, \cdots, \mathbf{R}_T)'$, and $\mathbf{P} \equiv \mathbb{M}_{\mathbf{1}_{T-1}}\mathbf{F}(\mathbf{F}'\mathbb{M}_{\mathbf{1}_{T-1}}\mathbf{F})^{-1}$, where we assume that $\mathbf{P}'\mathbf{P} = (\mathbf{F}'\mathbb{M}_{\mathbf{1}_{T-1}}\mathbf{F})^{-1} > 0$ for every T (see Assumption 3 in the Appendix A.1). The matrix $\mathbb{M}_{1_{T-1}} \equiv \mathbf{I}_{T-1} - \mathbf{1}_{T-1}\mathbf{1}'_{T-1}/(T-1)$ de-means the data, that is $\mathbb{M}_{1_{T-1}}\mathbf{R} = \mathbf{R} - \mathbf{1}_{T-1}\bar{\mathbf{R}}'$ and $\mathbb{M}_{1_{T-1}}\mathbf{F} = \mathbf{F} - \mathbf{1}_{T-1}\bar{\mathbf{f}}'$, setting $\bar{\mathbf{R}} \equiv \sum_{t=2}^{T} \mathbf{R}_t/(T-1)$ and $\bar{\mathbf{f}} \equiv \sum_{t=2}^{T} \mathbf{f}_t/(T-1)$.

It is clear that the estimator in (26) *excludes* the potential effect of the anomalies $\mathbf{Z}_{t-1}$, as well as the time variation of their premia. This could induce sources of bias in the estimates, further exacerbated if $\mathbf{f}_t$ and $\mathbf{Z}_{t-1}$ were potentially correlated across time, making $\hat{\mathbf{B}}$ clearly invalid. The following smoothness Assumption 2 permits to overcome these challenges, by constraining the time variation of the premia parameters, implying their (temporal) orthogonality with the risk factors $\mathbf{f}_t$. As the time-series dimension T gets small (and as long as $T > K_{\mathrm{f}} + 1$), this assumption appears extremely mild, especially in terms of anomalies' premia, where the observed (time-varying) $\mathbf{z}_{i,t-1}$ could account for most of the time-variation of their overall contribution to expect returns.

Assumption 2 (*smoothness of the premia parameters*)**.** *The following hold:*

$$\mathbf{P}'\gamma_0 = \mathbf{0}_{K_{\mathrm{f}}}, \qquad \mathbf{P}'\breve{\boldsymbol{\delta}}_{\mathrm{f}} = \mathbf{0}_{K_{\mathrm{f}}\times K_{\mathrm{f}}}, \qquad \textit{and} \quad \mathbf{P}'\boldsymbol{\Delta}_z = \mathbf{0}_{K_{\mathrm{f}}\times N},$$

setting the $(T-1)\times K_{\mathrm{f}}$ *matrix* $\breve{\boldsymbol{\delta}}_{\mathrm{f}} = (\breve{\boldsymbol{\delta}}_{f,1}, \cdots, \breve{\boldsymbol{\delta}}_{\mathrm{f},t-1})'$, with $\breve{\boldsymbol{\delta}}_{\mathrm{f},t-1} \equiv \boldsymbol{\delta}_{\mathrm{f,t-1}} - \mathbf{f}_t = \boldsymbol{\gamma}_{\mathrm{f},t-1} - E(\mathbf{f}_t|I_{t-1}, \boldsymbol{\Pi})$, *and the* $(T-1)\times N$ *matrix*

$$\boldsymbol{\Delta}_z \equiv \begin{bmatrix} \boldsymbol{\gamma}'_{z,1} - \boldsymbol{\gamma}'_{\mathrm{z}} & \mathbf{0}'_{K_{\mathrm{z}}} & \cdots & \mathbf{0}'_{K_{\mathrm{z}}} \\ \mathbf{0}'_{K_{\mathrm{z}}} & \boldsymbol{\gamma}'_{z,2} - \boldsymbol{\gamma}'_{\mathrm{z}} & \cdots & \mathbf{0}'_{K_{\mathrm{z}}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}'_{K_{\mathrm{z}}} & \mathbf{0}'_{K_{\mathrm{z}}} & \cdots & \boldsymbol{\gamma}'_{\mathrm{z},t-1} - \boldsymbol{\gamma}'_{\mathrm{z}} \end{bmatrix} \begin{bmatrix} \mathbf{Z}'_1 \\ \mathbf{Z}'_2 \\ \vdots \\ \mathbf{Z}'_{T-1} \end{bmatrix},$$

for some constant $K_{\mathrm{z}} \times 1$ *vector* $\boldsymbol{\gamma}_{\mathrm{z}}$ *satisfying* $N^{-1}\sum_{i=1}^{N}(\mathbf{Z}'_i\mathbf{Z}_i)^{-1}\mathbf{Z}'_i\mathbf{R}_i \to_p \boldsymbol{\gamma}_{\mathrm{z}}$.

When the risk factors are traded, $\breve{\boldsymbol{\delta}}_{\mathrm{f},t-1} = -\gamma_0 \mathbf{1}'_{K_\mathrm{f}}$ for every t, and Assumption 2 only concerns the zero-beta rate. Assumption 2 is extremely mild in most cases of interest: when the premia parameters are *locally* constant, the test assets and the risk factors are expressed as excess returns, and assuming that a risk-free asset is also traded, then Assumption 2 is always satisfied.[16]

The additional source of bias, arising from the presence of $\mathbf{Z}_{t-1}$ in the first-pass, is instead dealt with by *orthogonalizing* the anomaly variables $\mathbf{Z}_{t-1}$ with respect to the observable factors $\mathbf{f}_t$, *before* running the two-step procedure. Therefore, $\mathbf{Z}_{t-1}$ can be interpreted as representing the *net* portion of the anomaly variables that affects expected returns, hence eliminating any *indirect* influence (or confounding effect) coming from the risk factors.[17]

To better understand the implications of the orthogonalization on the model's parameters and their corresponding interpretations, consider the case where the researcher postulates a model that involves a set of *initial* anomalies $\mathbf{Z}^{\dagger}_{t-1} = (\mathbf{z}^{\dagger}_{1,t-1}, \cdots, \mathbf{z}^{\dagger}_{N,t-1})'$, such that:

$$\mathbf{R}_t = \boldsymbol{\alpha}^{\dagger}_{t-1} + \mathbf{Z}^{\dagger}_{t-1}\gamma_{\mathrm{z},t-1} + \mathbf{B}^{\dagger}\mathbf{f}_{t-1} + \mathbf{e_t} \tag{27}$$

where $\mathbf{B}^{\dagger}$ will be, in general, different from $\mathbf{B}$. Then, starting from (27), one can construct the *orthogonal* anomalies $\mathbf{Z}_{t-1}$ as the residuals from projecting $\mathbf{Z}^{\dagger}_{t-1}$ onto the unit constant and $\mathbf{f}_t$, implying a zero sample covariance between $\mathbf{z}_{i,t-1}$ and $\mathbf{f}_t$, and where we re-centre each $\mathbf{z}_{i,t-1}$ so that their sample mean coincides with the sample mean of $\mathbf{z}^{\dagger}_{\mathbf{i},\mathbf{t-1}}$, for every $i = 1, ..., N$. This leads to:

$$\mathbf{z}_{i,t-1} \equiv \mathbf{z}^{\dagger}_{i,t-1} - \hat{\boldsymbol{\Sigma}}_{\mathbf{z}^{\dagger}_i \mathbf{f}} \hat{\boldsymbol{\Sigma}}^{-1}_{\mathbf{f}}(\mathbf{f}_t - \bar{\mathbf{f}}), \tag{28}$$

where $\hat{\boldsymbol{\Sigma}}_{\mathbf{z}^{\dagger}_i \mathbf{f}} = \widehat{\mathrm{Cov}}(\mathbf{z}^{\dagger}_{i,t-1}, \mathbf{f}') = \frac{1}{T-1}\mathbf{Z}^{\dagger\prime}_i \mathbf{F} - \bar{\mathbf{Z}}^{\dagger}_i \bar{\mathbf{f}}'$, and $\hat{\boldsymbol{\Sigma}}_{\mathbf{f}} = \widehat{\mathrm{Var}}(\mathbf{f}) = \frac{1}{T-1}\mathbf{F}'\mathbf{F} - \bar{\mathbf{f}}\bar{\mathbf{f}}'$, where $\mathbf{Z}^{\dagger}_i = (\mathbf{z}^{\dagger}_{i,1}, \cdots, \mathbf{z}^{\dagger}_{i,T-1})'$, $\bar{\mathbf{Z}}^{\dagger}_i = \mathbf{Z}^{\dagger\prime}_i \frac{\mathbf{1}_{T-1}}{T-1}$, and where we use $\widehat{\mathrm{Cov}}(\cdot)$ and $\widehat{\mathrm{Var}}(\cdot)$ to denote the sample covariance and sample variance estimators, respectively. Then, replacing (28) in (27), and

[16] One can avoid imposing the smoothness conditions of Assumption 2, and thus allowing for time-series dependence between the time-varying premia and the risk factors, but at the cost of more complicate expressions. In particular, (11) can be expressed as a panel data model with interactive-fixed effects:

$$\mathbf{R}_t = \boldsymbol{\alpha} + \mathbf{Z}_{t-1}\bar{\boldsymbol{\gamma}}_{\mathrm{z}} + \mathbf{B}\mathbf{f}_t + \mathbf{u}_t,$$

where the error term satisfies $\mathbf{u}_t = \boldsymbol{\xi}_t + \boldsymbol{\Delta}\mathbf{g}_t$ for an asset-specific error $\boldsymbol{\xi}_t$ and a vector of zero-mean latent factors $\mathbf{g}_t$ possibly correlated with the observed risk factors $\mathbf{f}_t$, with loadings $\boldsymbol{\Delta}$, and where $\bar{\boldsymbol{\gamma}}_{\mathrm{z}} = T^{-1}\sum_{t=1}^{T}\boldsymbol{\gamma}_{t-1,z}$. Assumption 2 implies orthogonality between $\mathbf{f}_t$ and $\mathbf{u}_t$, resurrecting the OLS estimator $\hat{\mathbf{B}}$. However, an alternative estimator for $\mathbf{B}$ exists that avoids Assumption 2 but leads to a more involved analysis of the CSR in the second pass. Details are available upon request.

[17] The orthogonalization between anomalies and risk factors implies that $\mathbf{Z}_{t-1}$ are no longer pre-determined. By standard arguments, this leads to a bias of order $O_p(T^{-1})$, which, however, turns out to be irrelevant in our large-N–fixed-T sampling scheme, given the fast rate at which the bias vanishes.

imposing the asset pricing restriction in (6), we get model (11), where, setting $\bar{\boldsymbol{\gamma}}_{\mathrm{z}} \equiv \boldsymbol{\Gamma}_z' \frac{\mathbf{1}_{T-1}}{T-1}$ with $\boldsymbol{\Gamma}_z = (\boldsymbol{\gamma}_{\mathrm{z},1}, \cdots, \boldsymbol{\gamma}_{\mathrm{z,T-1}})'$,

$$\boldsymbol{\beta}_i \equiv \boldsymbol{\beta}_i^{\dagger} + \hat{\boldsymbol{\Sigma}}_{\mathbf{f}}^{-1} \hat{\boldsymbol{\Sigma}}_{\mathbf{z}_i^{\dagger}\mathbf{f}}' \bar{\boldsymbol{\gamma}}_{\mathrm{z}}. \tag{29}$$

From (29), it is easy to see that, after the orthogonalization of the anomaly variables, the (transformed) $\mathbf{B}$ takes now into account not only the direct effect of the risk factors on the cross-section of expected returns, but also the indirect effect of $\mathbf{f}_t$, trough its possible dependence with $\mathbf{Z}_{t-1}^{\dagger}$.

This set-up is extremely convenient and allows us to estimate the matrix $\mathbf{B}$ by simply using (26), without now incurring in any source of bias coming from the exclusion of anomalies from the first-pass regression or due to the potential correlation between risk factors and anomalies. Therefore, the feasible version of (11) becomes

$$\mathbf{R}_t = \hat{\mathbf{X}} \boldsymbol{\Gamma}_{\mathrm{f},t-1} + \mathbf{Z}_{t-1} \boldsymbol{\gamma}_{\mathrm{z},t-1} + \boldsymbol{\eta}_t, \tag{30}$$

setting $\hat{\mathbf{X}} = (\mathbf{1}_N, \hat{\mathbf{B}})$, with $\hat{\mathbf{B}}$ defined in (26), $\boldsymbol{\eta}_t \equiv \boldsymbol{\epsilon}_t - (\hat{\mathbf{X}} - \mathbf{X}) \boldsymbol{\Gamma}_{\mathrm{f},t-1}$, and where $\mathbf{Z}_{t-1}$ satisfies (28), hence being uncorrelated with the risk factors. Running a single cross-sectional OLS regression on (30) yields the time-varying OLS estimator

$$\begin{bmatrix} \hat{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} \\ \hat{\gamma}_{\mathrm{z},t-1} \end{bmatrix} \equiv \begin{bmatrix} \hat{\mathbf{X}}'\hat{\mathbf{X}} & \hat{\mathbf{X}}'\mathbf{Z}_{t-1} \\ \mathbf{Z}_{t-1}'\hat{\mathbf{X}} & \mathbf{Z}_{t-1}'\mathbf{Z}_{t-1} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{X}}'\mathbf{R}_t \\ \mathbf{Z}_{t-1}'\mathbf{R}_t \end{bmatrix}, \tag{31}$$

where $\hat{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} \equiv (\hat{\gamma}_{0,t-1}, \hat{\boldsymbol{\delta}}_{\mathrm{f},t-1}')'$.[18] The estimator in (31) generalizes the conventional estimator $\tilde{\gamma}_{\mathrm{z},t-1}$ in (14) to the case of when both anomalies and (estimated) loadings are used as regressors in the feasible model. The two estimators coincide when $\hat{\mathbf{X}}'\mathbf{Z}_{t-1} = \mathbf{0}_{N \times K_{\mathrm{z}}}$, a condition which is, however, not warranted in general. When such orthogonality condition is violated, then $\hat{\boldsymbol{\gamma}}_{\mathrm{z},t-1}$ in (31) remains valid, but $\tilde{\gamma}_{\mathrm{z},t-1}$ in (14) becomes biased.[19]

Although (31) resolves the bias coming from the potential lack of orthogonality between the risk factors and the anomalies, unfortunately other sources of bias arise in our large-N–fixed-T set-up. The reason is that $\hat{\mathbf{B}}$ does not converge to $\mathbf{B}$ when T is fixed, making the OLS estimator in (31)

[18] To simplify the exposition we are slightly abusing the notation by indexing the OLS estimator (31) to time $t-1$ instead to time t.

[19] To clarify, notice that the orthogonality condition that we impose between $\mathbf{Z}_{t-1}$ and $\mathbf{f}_t$ represents a *time-series* restriction, which does not imply the *cross-sectional* restriction $\hat{\mathbf{X}}'\mathbf{Z}_{t-1} = \mathbf{0}_{N \times K_z}$.

biased due the EIV effect.[20]

However, we show that such biases can be consistently estimated, leading to our new *bias-adjusted* CSR OLS estimator:

$$\begin{bmatrix} \hat{\boldsymbol{\Gamma}}^*_{\mathrm{f},t-1} \\ \hat{\boldsymbol{\gamma}}^*_{\mathrm{z},t-1} \end{bmatrix} \equiv \begin{bmatrix} \hat{\mathbf{X}}'\hat{\mathbf{X}} - N\hat{\boldsymbol{\Lambda}}_1 & \hat{\mathbf{X}}'\mathbf{Z}_{t-1} \\ \mathbf{Z}'_{t-1}\hat{\mathbf{X}} & \mathbf{Z}'_{t-1}\mathbf{Z}_{t-1} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{X}}'\mathbf{R}_t - N\hat{\boldsymbol{\Lambda}}_{2,t-1} \\ \mathbf{Z}'_{t-1}\mathbf{R}_t \end{bmatrix}, \tag{32}$$

where $\hat{\boldsymbol{\Gamma}}^*_{\mathrm{f},t-1} \equiv (\hat{\gamma}^*_{0,t-1}, \hat{\boldsymbol{\delta}}^{*\prime}_{\mathrm{f},t-1})'$, and where we set

$$\hat{\boldsymbol{\Lambda}}_1 \equiv \begin{bmatrix} 0 & \mathbf{0}'_{K_\mathrm{f}} \\ \mathbf{0}_{K_\mathrm{f}} & \hat{\sigma}^2\mathbf{P}'\mathbf{P} \end{bmatrix}, \qquad \hat{\boldsymbol{\Lambda}}_{2,t-1} \equiv \hat{\sigma}^2 \begin{bmatrix} 0 \\ \mathbf{P}'\boldsymbol{\imath}_{t-1,T-1} \end{bmatrix}, \tag{33}$$

where $\boldsymbol{\imath}_{s,T-1}$ denotes the s-th row of the identity matrix $\mathbf{I}_{T-1}$, and where

$$\hat{\sigma}^2 \equiv \frac{\mathrm{tr}(\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}})}{N(T-K-2)}, \tag{34}$$

with $\mathrm{tr}(\cdot)$ denoting the trace operator, $K = K_\mathrm{f} + K_\mathrm{z}$, and where $\hat{\boldsymbol{\epsilon}}$ represents the OLS residuals, defined as $\hat{\boldsymbol{\epsilon}}_i \equiv \mathbf{M}_{\tilde{\mathbf{D}}_i}\mathbf{R}_i$, with $\mathbb{M}_{\tilde{\mathbf{D}}_i} = \mathbf{I}_{T-1} - \tilde{\mathbf{D}}_i(\tilde{\mathbf{D}}_i'\tilde{\mathbf{D}}_i)^{-1}\tilde{\mathbf{D}}_i'$, and $\tilde{\mathbf{D}}_i \equiv (\mathbf{D}, \tilde{\mathbf{Z}}_i)$, with $\mathbf{D} \equiv (\mathbf{1}_{T-1}, \mathbf{F})$, and $\tilde{\mathbf{Z}}_i \equiv \mathbb{M}_{\mathbf{1}_{T-1}}\mathbf{Z}_i$.[21]

The following theorem establishes the limiting properties of our novel bias-adjusted estimator. Let $\mathbf{Z} \equiv (\mathbf{z}_1, ..., \mathbf{z}_N)'$ define the overall $N \times K_\mathrm{z}(T-1)$ matrix of anomalies, with $\mathbf{z}_i$ being the $K_\mathrm{z}(T-1) \times 1$ vector $\mathbf{z}_i \equiv \left(\mathrm{z}^{(1)}_{i,1}, \cdots, \mathrm{z}^{(1)}_{i,T-1}, \cdots, \mathrm{z}^{(K_\mathrm{z})}_{i,1}, \cdots, \mathrm{z}^{(K_\mathrm{z})}_{i,T-1}\right)'$, with $\mathrm{z}^{(j)}_{i,T-1}$ denoting the value of the jth anomaly for stock i at time t. Let $\mathbf{0}_a$ and $\mathbf{1}_a$ denote an $a \times 1$ vector of zeros and ones, respectively. The following $K_\mathrm{z}(T-1) \times K_\mathrm{z}$ matrices of constants

$$\mathbb{J} = \frac{1}{T-1} \begin{bmatrix} \mathbf{1}_{T-1} & \mathbf{0}_{T-1} & \dots & \mathbf{0}_{T-1} \\ \mathbf{0}_{T-1} & \mathbf{1}_{T-1} & \dots & \mathbf{0}_{T-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{T-1} & \mathbf{0}_{T-1} & \dots & \mathbf{1}_{T-1} \end{bmatrix} = \left(\mathbf{I}_{K_\mathrm{z}} \otimes \frac{\mathbf{1}_{T-1}}{(T-1)}\right) = \frac{1}{T-1}\sum_{s=1}^{T-1} \mathbb{J}_s \tag{35}$$

[20]Moreover, as the estimator (31) is evaluated at each point in time, a second source of bias arises (besides the EIV) due to the fact that $\mathbf{P}'\boldsymbol{\imath}_{t-1,T-1}$ could be, in general, different from $\mathbf{0}_{K_\mathrm{f}}$. See Proposition OA.1 in the Online Appendix for a formal proof. In remark OA.19 we also show that the OLS estimator in (31) remains biased even when one assumes that T is large, but N is fixed. However, in this case, the bias would be a function of a random component, making the bias term impossible to be consistently estimated, unlike our large-N case.

[21]Note that, while computation of the OLS estimator $\hat{\boldsymbol{\beta}}_i$ only requires the regressors $\mathbf{D}$, the corresponding residuals must be evaluated with respect to both $\mathbf{D}_i$ and $\tilde{\mathbf{Z}}_i$, as it always happens in regressions with orthogonal independent variables.

with

$$\mathbb{J}_s = \begin{bmatrix} \boldsymbol{\iota}_{s,T-1} & \mathbf{0}_{T-1} & \dots & \mathbf{0}_{T-1} \\ \mathbf{0}_{T-1} & \boldsymbol{\iota}_{s,T-1} & \dots & \mathbf{0}_{T-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{T-1} & \mathbf{0}_{T-1} & \dots & \boldsymbol{\iota}_{s,T-1} \end{bmatrix} = \left(\mathbf{I}_{K_z} \otimes \boldsymbol{\iota}_{s,T-1}\right) \text{ for } 1 \leq s \leq T-1, \tag{36}$$

are needed to evaluate the sample means of the anomaly variables and to select the s-th observation from the $\mathbf{Z}$ matrix, yielding $\mathbf{Z}\mathbb{J}_s = \mathbf{Z}_s$ and $\mathbf{Z}\mathbb{J} = (T-1)\sum_{s=1}^{T-1} \mathbf{Z}_s$. Finally, let $\otimes$, vec$(\cdot)$ and $\odot$ denote the Kronecker product, the vec operator, and the Hadamard product, respectively, and let $\rightarrow_p, \rightarrow_d$ denote convergence in probability and distribution, respectively.

Theorem 1 (Large-N consistency and asymptotic normality of the time varying bias-adjusted CSR OLS estimator)**.** *As $N \rightarrow \infty$, under Assumptions 2–8 (listed in the Appendix A.1), then*

(i)

$$\hat{\boldsymbol{\Gamma}}^*_{\mathrm{f},t-1} - \boldsymbol{\Gamma}_{\mathrm{f},t-1} = O_p\left(\frac{1}{\sqrt{N}}\right) \quad \textit{and} \quad \hat{\boldsymbol{\gamma}}^*_{\mathrm{z},t-1} - \boldsymbol{\gamma}_{\mathrm{z},t-1} = O_p\left(\frac{1}{\sqrt{N}}\right), \tag{37}$$

(ii)

$$\sqrt{N}\begin{bmatrix} \hat{\boldsymbol{\Gamma}}^*_{\mathrm{f},t-1} - \boldsymbol{\Gamma}_{\mathrm{f},t-1} \\ \hat{\boldsymbol{\gamma}}^*_{\mathrm{z},t-1} - \boldsymbol{\gamma}_{\mathrm{z},t-1} \end{bmatrix} \rightarrow_d \mathcal{N}\left(\mathbf{0}_{K+1}, \mathbf{L}^{-1}_{t-1}\mathbf{O}_{t-1}\mathbf{L}^{-1\prime}_{t-1}\right), \tag{38}$$

for some $\mathbf{L}_{t-1} > 0$ and $\mathbf{O}_{t-1}$ defined in (OA.35).[22]

Proof. See Appendix OA.4.

To conduct statistical inference, we need a consistent estimator of the asymptotic covariance matrix in (38), which we present in the next theorem.

Theorem 2 (Standard errors of the time varying bias-adjusted CSR OLS estimator)**.** *As $N \rightarrow \infty$, under Assumptions 2–8, and the identification condition $\kappa_4 = 0$,*

$$\hat{\mathbf{L}}^{-1}_{t-1}\,\hat{\mathbf{O}}_{t-1}\,\hat{\mathbf{L}}^{-1\prime}_{t-1} \;\; \rightarrow_p \;\; \mathbf{L}^{-1}_{t-1}\,\mathbf{O}_{t-1}\,\mathbf{L}^{-1\prime}_{t-1} \tag{39}$$

[22]To ease the exposition, the definition of $\mathbf{L}_{t-1}0$ and $\mathbf{O}_{t-1}$ has been relegated to the proof of the theorem (see (OA.35)).

where

$$\hat{\mathbf{L}}_{t-1} \equiv \frac{1}{N} \begin{bmatrix} \hat{\mathbf{X}}'\hat{\mathbf{X}} - N\hat{\boldsymbol{\Lambda}}_1 & \hat{\mathbf{X}}'\mathbf{Z}_{t-1} \\ \mathbf{Z}'_{t-1}\hat{\mathbf{X}} & \mathbf{Z}'_{t-1}\mathbf{Z}_{t-1} \end{bmatrix}, \quad \textit{and} \quad \hat{\mathbf{O}}_{t-1} \equiv \begin{bmatrix} \hat{\mathbf{U}}_{t-1} & \hat{\sigma}^2 \hat{\mathbf{G}}_{t-1} \hat{\mathbf{H}}'_{t-1} \\ \hat{\sigma}^2 \hat{\mathbf{H}}_{t-1} \hat{\mathbf{G}}'_{t-1} & \hat{\mathbf{H}}_{t-1} \hat{\boldsymbol{\Sigma}}_{\mathrm{U}} \hat{\mathbf{H}}'_{t-1}, \end{bmatrix} \tag{40}$$

with $\hat{\mathbf{U}}_{t-1} \equiv \hat{\sigma}^2 \hat{\mathbf{Q}}'_{t-1} \hat{\mathbf{Q}}_{t-1} \left(\hat{\boldsymbol{\Sigma}}_{\mathrm{X}} - \hat{\boldsymbol{\Lambda}}_1 \right) + \begin{bmatrix} 0 & \mathbf{0}'_{K_{\mathrm{f}}} \\ \mathbf{0}_{K_{\mathrm{f}}} & \hat{\mathbf{V}}'_{t-1} \hat{\mathbf{U}}_{\epsilon} \hat{\mathbf{V}}_{t-1} \end{bmatrix}$, *setting* $\overline{\mathbb{M}}_{\tilde{\mathrm{D}}} \equiv N^{-1} \sum_{i=1}^{N} \mathbb{M}_{\tilde{\mathrm{D}}_i}$, $\hat{\boldsymbol{\Sigma}}_{\mathrm{X}} \equiv N^{-1} \hat{\mathbf{X}}'\hat{\mathbf{X}}$, $\hat{\boldsymbol{\Sigma}}_{\mathrm{ZB}} \equiv N^{-1} \mathbf{Z}'\hat{\mathbf{B}}$, $\hat{\boldsymbol{\mu}}_{\mathrm{z},T-1} \equiv N^{-1} \mathbf{Z}'\mathbf{1}_N$, *and* $\hat{\boldsymbol{\Sigma}}_{\mathrm{U}} \equiv (\hat{\sigma}^2 \boldsymbol{I}_{T-1} \otimes \mathbf{Z}'\mathbf{Z}/N)$, *with* $\hat{\boldsymbol{\Lambda}}_1$ *and* $\hat{\sigma}^2$ *defined in* (33) *and* (34)*, respectively, and we define the following matrices*

$$\begin{aligned}
\hat{\mathbf{Q}}_{t-1} &\equiv \boldsymbol{\imath}_{t-1,T-1} - \mathbf{P} \hat{\boldsymbol{\delta}}^*_{\mathrm{f},t-1}, \quad \hat{\mathbf{H}}_{t-1} \equiv \hat{\mathbf{Q}}'_{t-1} \otimes \mathbb{J}'_{t-1}, \\
\hat{\mathbf{G}}_{t-1} &\equiv \left[\hat{\mathbf{Q}}_{t-1} \otimes \hat{\boldsymbol{\mu}}_{\mathrm{z},T-1}, \hat{\mathbf{Q}}_{t-1} \otimes \hat{\boldsymbol{\Sigma}}_{\mathrm{ZB}} \right]', \ \textit{and} \\
\hat{\mathbf{V}}_{t-1} &\equiv (\hat{\mathbf{Q}}_{t-1} \otimes \mathbf{P}) - \left(\frac{\mathrm{vec}(\overline{\mathbb{M}}_{\tilde{\mathrm{D}}})}{(T-K-2)} \right) \hat{\mathbf{Q}}'_{t-1} \mathbf{P},
\end{aligned}$$

where $\hat{\mathbf{U}}_{\epsilon}$ *is obtained plugging* $\kappa_4 = 0$ *and* $\hat{\sigma}^4 = N^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \hat{\epsilon}^4_{it} / 3 \, \mathrm{tr} \left(\overline{\mathbb{M}}^{(2)}_{\tilde{\mathrm{D}}} \right)$, *with* $\overline{\mathbb{M}}^{(2)}_{\tilde{\mathrm{D}}} \equiv \frac{1}{N} \sum_{i=1}^{N} \left(\mathbb{M}_{\tilde{\mathrm{D}}_i} \odot \mathbb{M}_{\tilde{\mathrm{D}}_i} \right)$, *into* $\mathbf{U}_{\epsilon} = \mathbf{U}_{\epsilon}(\kappa_4, \sigma^4)$ *(see Remark 2 to Assumption 7).*

Proof. See Appendix OA.4.

The square root of the diagonal elements of $\hat{\mathbf{L}}^{-1}_{t-1} \hat{\mathbf{O}}_{t-1} \hat{\mathbf{L}}^{-1}_{t-1}$ in (40), divided by $\sqrt{N}$, represent the standard errors of the premia estimators $\hat{\boldsymbol{\Gamma}}^*_{\mathrm{f},t-1}$ and $\hat{\boldsymbol{\gamma}}^*_{\mathrm{z},t-1}$, which can be used to construct asymptotically valid confidence intervals.

6 Implications for Characteristic-Based Portfolios

An increasing strand of cross-sectional asset pricing literature, pioneered by Fama (1976), interprets (31) as characteristic-based portfolios. Therefore, a key and immediate implication of our novel asymptotic analysis relies in its ability to provide inference - specifically, standard errors and forecasting bands - for such portfolios, as we explain below.

Various methods have been employed to construct characteristic-based factors, ranging from sorting approaches (see Fama and French (1993)) to cross-sectional OLS regressions (see Back, Kapadia, and Ostdiek (2015) and Fama and French (2020)). Other techniques involve linear combinations derived from rank-transformed centered characteristics (as discussed in Kozak, Nagel, and Santosh (2020)), as well as PCA-type methods, as illustrated by Kelly, Pruitt, and Su (2019) and Kim, Korajczyk, and Neuhierl (2021). Kozak and Nagel (2023) spell out the necessary conditions for a generic (referred to as *heuristic*) estimator of characteristic-based portfolios to be (approximately) correct, thereby circumventing the challenging task of inverting a large covariance matrix central to the GLS approach, which they show is the only setting where characteristic-based portfolios can attain mean-variance efficiency.[23]

The key connection to understand how our anomaly-dissecting testing procedure can be used to accurately estimate characteristic-based portfolios hinges on the identification condition detailed in Section 3.1.1. Specifically, this condition asserts that a factor model characterized by time-varying risk exposures influenced by specific firms' characteristics $\mathbf{Z}_{t-1}$ can be observationally indistinguishable from a model featuring constant risk exposures and time-varying alphas driven by the same $\mathbf{Z}_{t-1}$.[24]

Indeed, consider again model (2) under exact pricing (3) and assume time varying risk exposures $\mathbf{B}_{t-1} = [\mathbf{Z}_{t-1}, \mathbf{B}]$. This implies that the K risk factors in $\mathbf{f}_t$ are made by K_c *latent* factors $\mathbf{f}_t^c$ (representing the characteristic-based portfolios) having observed time-varying risk exposures $\mathbf{Z}_{t-1}$, and by K_o *observed* risk factors $\mathbf{f}_t^o$ with locally constant risk exposures $\mathbf{B}$ (by Assumption 1).[25] That is, denoting $\mathbf{f}_t = [\mathbf{f}_t^{c\prime}, \mathbf{f}_t^{o\prime}]'$ and re-arranging, gives:

$$\mathbf{R}_t = \gamma_{0,t-1}\mathbf{1}_N + \mathbf{Z}_{t-1}\mathbf{F}_t^c + \mathbf{B}\mathbf{F}_t^o + \boldsymbol{\epsilon}_t, \tag{41}$$

where $\mathbf{F}_t^c$ and $\mathbf{F}_t^o$ represent re-centred versions of the risk factors around the corresponding risk premia, that is

$$\mathbf{F}_t^c \equiv \boldsymbol{\gamma}_{\mathrm{c,t-1}} + \mathbf{f}_t^c - \mathrm{E}\left[\mathbf{f}_t^c | I_{t-1}, \boldsymbol{\Pi}\right] \text{ and } \mathbf{F}_t^o \equiv \boldsymbol{\gamma}_{\mathrm{o,t-1}} + \mathbf{f}_t^o - \mathrm{E}\left[\mathbf{f}_t^o | I_{t-1}, \boldsymbol{\Pi}\right], \tag{42}$$

where $\boldsymbol{\gamma}_{\mathrm{c,t-1}}$ and $\boldsymbol{\gamma}_{\mathrm{o,t-1}}$ are the risk premia of the latent and observed risk factors $\mathbf{f}_t^c$ and $\mathbf{f}_t^o$, respectively.[26]

[23] Acknowledging that some of these heuristic methodologies can result in characteristics-based factors contaminated by unpriced risk, Daniel, Mota, Rottke, and Santos (2020) propose a corrective approach to eliminate the influence of these unpriced components rooted in the concept of hedging portfolios.

[24] A crucial difference, not explored here as we focus primarily on expected returns, arise between a model with characteristic-based alphas and characteristic-based betas in terms of the covariance matrix of the test assets' returns, which must be a function of characteristics driving the risk exposures in the time-varying betas specification (see Kozak and Nagel (2023)) but not necessarily in the time-varying alphas specification.

[25] Kelly, Pruitt, and Su (2019) and Kim, Korajczyk, and Neuhierl (2021) pioneered the interpretation of latent risk factors estimators as characteristics-based portfolios - see also the discussion in Kozak and Nagel (2023), Section III.

[26] To clarify, using our earlier notation in (8), it is evident that $\mathbf{F}_t^c = \boldsymbol{\delta}_{\mathrm{c,t-1}}$ and $\mathbf{F}_t^o = \boldsymbol{\delta}_{\mathrm{o,t-1}}$. However, we prefer to adopt this notation in this section to emphasize estimating cross-sectional portfolio returns instead of risk premia.

The cross-sectional estimation of $\mathbf{F}_t^c$, along with $\mathbf{F}_t^o$ and $\gamma_{0,t-1}$, can be therefore obtained using our two-step procedure. Firstly, $\mathbf{B}$ is estimated using (26), which entails regressing the returns on an intercept and the observed risk factors $\mathbf{f}_t^o$ only, a choice justified by the orthogonality assumption between observed and latent factors. Secondly, a cross-sectional regression is performed for each period t, wherein $\mathbf{R}_t$ is regressed onto an intercept, the observed characteristics $\mathbf{Z}_{t-1}$, and the estimated $\hat{\mathbf{B}}$ from the fist-pass. The outcome of this regression yields the estimated *cross-sectional OLS factors*, which can be interpreted as portfolio returns:

$$\begin{bmatrix} \hat{\gamma}_{0,t-1} \\ \hat{\mathbf{F}}_t^o \\ \hat{\mathbf{F}}_t^c \end{bmatrix} \equiv \mathbb{W}_{t-1}^{ols\prime} \mathbf{R}_t \text{ with weights } \mathbb{W}_{t-1}^{ols} \equiv \begin{bmatrix} \hat{\mathbf{X}} & \mathbf{Z}_{t-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{X}}'\hat{\mathbf{X}} & \hat{\mathbf{X}}'\mathbf{Z}_{t-1} \\ \mathbf{Z}_{t-1}'\hat{\mathbf{X}} & \mathbf{Z}_{t-1}'\mathbf{Z}_{t-1} \end{bmatrix}^{-1}, \tag{43}$$

where $\hat{\mathbf{X}} = [\mathbf{1}_N, \hat{\mathbf{B}}]$ and $\hat{\mathbf{B}}$ is defined in (26). Clearly, $\mathbb{W}_{t-1}^{ols\prime}\mathbf{R}_t = (\hat{\mathbf{\Gamma}}_{\mathrm{f},t-1}', \hat{\gamma}_{\mathrm{z},t-1}')'$, which coincides with our estimator in (31). However, delving into the interpretation of cross-sectional risk factors, one can also easily see that the portfolio weights satisfy:

$$\mathbb{W}_{t-1}^{ols\prime} \begin{bmatrix} \hat{\mathbf{X}} & \mathbf{Z}_{t-1} \end{bmatrix} = \begin{bmatrix} \mathbb{W}_{t-1}^{0\prime}\mathbf{1}_N & \mathbb{W}_{t-1}^{0\prime}\hat{\mathbf{B}} & \mathbb{W}_{t-1}^{0\prime}\mathbf{Z}_{t-1} \\ \mathbb{W}_{t-1}^{\beta\prime}\mathbf{1}_N & \mathbb{W}_{t-1}^{\beta\prime}\hat{\mathbf{B}} & \mathbb{W}_{t-1}^{\beta\prime}\mathbf{Z}_{t-1} \\ \mathbb{W}_{t-1}^{Z\prime}\mathbf{1}_N & \mathbb{W}_{t-1}^{Z\prime}\hat{\mathbf{B}} & \mathbb{W}_{t-1}^{Z\prime}\mathbf{Z}_{t-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{44}$$

where we denote $\mathbb{W}_{t-1}^{ols} = \begin{bmatrix} \mathbb{W}_{t-1}^{0} & \mathbb{W}_{t-1}^{\beta} & \mathbb{W}_{t-1}^{Z} \end{bmatrix}$, and where in (44) we set $K_{\mathrm{f}} = K_{\mathrm{z}} = 1$ for simplicity. Back, Kapadia, and Ostdiek (2015) refer to the cross-sectional estimator of $\mathbf{F}_t^c$ as the (characteristics) *pure-play* portfolios because their weights satisfy (44).[27] Therefore, the intercept portfolio $\hat{\gamma}_{0,t-1}$ with weights $\mathbb{W}_{t-1}^{0}$ represents a unit-long portfolio ($\mathbb{W}_{t-1}^{0\prime}\mathbf{1}_N = 1$) with a zero spread in both the characteristics and the factor's risk exposures (since $\mathbb{W}_{t-1}^{0\prime}\mathbf{Z}_{t-1} = \mathbb{W}_{t-1}^{0\prime}\hat{\mathbf{B}} = 0$). The characteristic-based portfolio $\hat{\mathbf{F}}_t^c$ (i.e, the OLS factor) represents a zero-investment long-short portfolio ($\mathbb{W}_{t-1}^{0\prime}\mathbf{1}_N = 0$), which is factor-neutral (i.e, $\mathbb{W}_{t-1}^{0\prime}\hat{\mathbf{B}} = 0$) but has a unit spread in the $\mathbf{Z}_{t-1}$ characteristic (i.e, $\mathbb{W}_{t-1}^{0\prime}\mathbf{Z}_{t-1} = 1$). Specular properties holds for the beta portfolio $\hat{\mathbf{F}}_t^o$, having only a unit spread in the factor's risk exposures.

Our asymptotic analysis derived in Theorems 1 and 2 serves as a rigorous foundation for estimating cross-sectional OLS portfolios accurately and drawing correct inference on them, such

[27] Specifically, every column of $\mathbb{W}_{t-1}^{ols}$, say the jth column with $j = 1, \cdots, K+1$, is the solution of the minimization problem $\min \mathbf{w}'\mathbf{w}$ such that $\begin{bmatrix} \hat{\mathbf{X}} & \mathbf{Z}_{t-1} \end{bmatrix}' \mathbf{w} = \iota_j$, where $\boldsymbol{\iota}_j$ denotes the jth column (or row) of the identity matrix of dimension $(K+1) \times (K+1)$, leading to the interpretation of the columns of $\mathbb{W}_{t-1}^{ols}$ as maximally diversified pure plays.

as, for instance, the construction of accurate forecasting intervals. The use and the validity of our methodology are further supported by the fact that our analysis is developed under the large-N setting, which represents an ideal sampling scheme in evaluating cross-sectional regressions. Moreover, keeping T fixed not only has enabled us to formulate our methodology within a context of conditional asset pricing with time-varying premia, but also to interpret these results as time-varying risk factors in characteristic-based portfolios construction. This is in contrast with existing PCA-related estimators, which typically hinge on a double asymptotic framework, where both N and T tend to infinity (see Kelly, Pruitt, and Su (2019) and Kim, Korajczyk, and Neuhierl (2021)). We also differ from these contributions by assuming that the risk exposures are parameter-free. This ensures that the estimates of latent risk factors, i.e., the cross-sectional OLS portfolios, remain uncontaminated by any unknown rotation matrix.

Notice that to achieve accurate estimation of such characteristic-based portfolios requires the bias adjustment in (32). This is because the risk exposures to observed factors are contaminated by estimation noise (due to the fixed nature of T), unless, of course, one opts to exclude $\mathbf{f}_t^o$ from the asset pricing model (41), as commonly practiced in this literature. As discussed in Section 4, however, this latter method may be affected by significant biases and corrupted inference whenever the observed factors are priced and cannot be effectively excluded. Alternatively, it becomes a concern if their risk exposures lack orthogonality to the characteristics $\mathbf{Z}_{t-1}$.

6.1 Risk Premia Estimation of Characteristic-Based Portfolios

Whereas our OLS CRS inferential procedure of Section 5 can be readily used for estimation and inference on the cross-sectional OLS factors, a suitable modification of our large-N methodology allows to estimate accurately their risk premia even when the latter are time-varying, in contrast to the conventional large-T approaches which give biased risk premia estimates, as explained in Section 4.

Considering the asset pricing model (41), one obtains $\boldsymbol{\gamma}_{\mathrm{c,t-1}} = \mathrm{E}\left[\mathbf{F}_t^c | I_{t-1}, \boldsymbol{\Pi}\right]$.[28] Therefore, a local average of the $\mathbf{F}_t^c$ provides a natural nonparametric estimator (specifically, with a rectangular kernel) for its risk premia vector $\boldsymbol{\gamma}_{\mathrm{c,t-1}}$ as the latter is a conditional mean.[29] This raises two

[28]NO EXTEND TO OBSERVED FACTORS IN TEXT The same arguments apply for $\boldsymbol{\gamma}_{\mathrm{o,t-1}} = \mathrm{E}\left[\mathbf{F}_t^o | I_{t-1}, \boldsymbol{\Pi}\right]$ but we focus here on estimation of the risk premia associated with the characteristic-based risk factor $\mathbf{F}_t^c$.

[29]By standard arguments, the sample mean represents a valid nonparametric estimator of the conditional expecta-

challenges, namely that $\mathbf{F}_t^c$ is not observed, and second that our asymptotic analysis works for fixed T, which we now tackle. First, given that T remains fixed, one can identify a unbiased proxi of the risk premia, that is one can estimate $\bar{\boldsymbol{\delta}}_{\rm c} \equiv T^{-1}\sum_{t=1}^{T}\bar{\mathbf{F}}_t^c = \bar{\boldsymbol{\gamma}}_{\rm c} + \bar{\mathbf{f}}_{\rm c} - \overline{\rm E}[\mathbf{F}^c] \approx \bar{\boldsymbol{\gamma}}_{\rm c}$, known as the ex-post risk premia (see Shanken (1992), here extended to the case of time-varying moments and risk premia).[30] The ex-post risk premia, amid the noisiness related to the fixed T, remains an extremely valuable quantity, both empirically and theoretically. For example, Zaffaroni (2022) shows that the SDF built on the ex-post risk premia induces pricing errors of the order of $O(T^{-1})$ despite $\bar{\boldsymbol{\delta}}_{\rm c}$ differs from $\bar{\boldsymbol{\gamma}}_{\rm c}$ by an order $O(T^{-\frac{1}{2}})$. Second, as (by Theorems 1 and 2) $\hat{\mathbf{F}}_t^c$ accurately captures the latent risk factor vector $\mathbf{F}_t^c$ for every given point in time, one should be able to estimate $\bar{\boldsymbol{\delta}}_{\rm c}$ accurately. We now formalize this conjecture, providing the corresponding inferential analysis.

Specifically, let $\bar{\mathbf{Z}} = \frac{1}{(T-1)}\sum_{t=1}^{T-1}\mathbf{Z}_t$ be the $N \times K_{\rm z}$ matrix of characteristics' time-series averages. Then, by averaging the second-pass relationship in (41) across time, and noticing that $(T-1)^{-1}\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}\mathbf{F}_t^c = \bar{\mathbf{Z}}\bar{\boldsymbol{\delta}}_{\rm c} + \widehat{\rm Cov}(\mathbf{Z}_{t-1}, \mathbf{F}_t^c)$, with $\widehat{\rm Cov}(\mathbf{Z}_{t-1}, \mathbf{F}_t^c) \equiv (T-1)^{-1}\sum_{t=1}^{T-1}(\mathbf{Z}_{t-1} - \bar{\mathbf{Z}})(\mathbf{F}_t^c - \bar{\boldsymbol{\delta}}_{\rm c})$, one obtains:

$$\bar{\mathbf{R}} = \bar{\gamma}_0\mathbf{1}_N + \bar{\mathbf{Z}}\bar{\boldsymbol{\delta}}_{\rm c} + \mathbf{B}\bar{\boldsymbol{\delta}}_{\rm o} + \bar{\boldsymbol{\epsilon}}^* = \bar{\mathbf{Z}}\bar{\boldsymbol{\delta}}_{\rm c} + \mathbf{X}\bar{\boldsymbol{\Gamma}}_{\rm o} + \bar{\boldsymbol{\epsilon}}^*, \tag{45}$$

where $\bar{\boldsymbol{\Gamma}}_{\rm o} \equiv \left(\bar{\gamma}_0,\, \bar{\boldsymbol{\delta}}_{\rm o}'\right)'$, and $\bar{\boldsymbol{\epsilon}}^* \equiv \bar{\boldsymbol{\epsilon}} + \widehat{\rm Cov}(\mathbf{Z}_{t-1}, \mathbf{F}_t^c)$. Therefore, following the same steps adopted for the time-varying estimator in (32), we can derive the OLS bias-adjusted estimator of the *locally-averaged* risk premia as:

$$\begin{bmatrix} \hat{\bar{\boldsymbol{\Gamma}}}_{\rm o}^* \\ \hat{\bar{\boldsymbol{\delta}}}_{\rm c}^* \end{bmatrix} \equiv \begin{bmatrix} \hat{\mathbf{X}}'\hat{\mathbf{X}} - N\hat{\boldsymbol{\Lambda}}_1 & \hat{\mathbf{X}}'\bar{\mathbf{Z}} \\ \bar{\mathbf{Z}}'\hat{\mathbf{X}} & \bar{\mathbf{Z}}'\bar{\mathbf{Z}} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{X}}'\bar{\mathbf{R}} \\ \bar{\mathbf{Z}}'\bar{\mathbf{R}} \end{bmatrix}, \tag{46}$$

where $\hat{\boldsymbol{\Lambda}}_1$ is defined in (33), and where $\hat{\bar{\boldsymbol{\Gamma}}}_{\rm o}^* \equiv \left(\hat{\bar{\gamma}}_{\mathbf{0}}^*,\, \hat{\bar{\boldsymbol{\delta}}}_{\rm o}^{*\prime}\right)'$. Notice that now, compared with the time-varying estimator in (32), the estimator in (46) is immune of the bias term related to $\boldsymbol{\Lambda}_2$. The reason is that this bias vanishes when one constructs the *locally-averaged* estimators (46), because $(T-1)^{-1}\sum_{t=2}^{T}\mathbf{P}\boldsymbol{\imath}_{t-1,T-1} = \mathbf{P}'\mathbf{1}_{T-1} = \mathbf{0}_{K_{\rm o}}$ by construction (see Theorem 3 in the Online Appendix OA.5.2). The asymptotic properties of our risk premia estimators are given as follows, where we estimate jointly the (average) zero-beta rate as well as the risk premia to the characteristic-based risk factors and the observed risk factors.

tion when the true risk premia's time-variation is sufficiently smooth and not too abrupt, where smoothness can be reasonably assumed in our setting, as T can be chosen to be arbitrarily small, with the only requirement being that $T > (K_{\rm f} + 2)$.

[30]ALREADY DEFINED?By $\bar{A}$ we generically denote the sample mean of the quantities $A_1, \cdots, A_T$.

Theorem 3 (large-N–fixed-T - consistency and asymptotic normality of the risk premia CSR OLS estimator)**.** *Under Assumptions 2—8 and* $\widehat{\mathrm{Cov}}(\mathbf{Z}_{t-1}, \mathbf{F}_{\mathrm{c},t}) = o_p(N^{-1/2})$*, as* $N \to \infty$*,*

(i)

$$\hat{\bar{\mathbf{\Gamma}}}_{\mathrm{o}}^{*} - \bar{\mathbf{\Gamma}}_{\mathrm{o}} = \mathbf{O_p}\left(\frac{\mathbf{1}}{\sqrt{\mathbf{N}}}\right) \quad \textit{and} \quad \hat{\bar{\boldsymbol{\delta}}}_{\mathrm{c}}^{*} - \bar{\boldsymbol{\delta}}_{\mathrm{c}} = \mathbf{O_p}\left(\frac{\mathbf{1}}{\sqrt{\mathbf{N}}}\right), \tag{47}$$

(ii)

$$\sqrt{N}\begin{bmatrix} \hat{\bar{\mathbf{\Gamma}}}_{\mathrm{o}}^{*} - \bar{\mathbf{\Gamma}}_{\mathrm{o}} \\ \hat{\bar{\boldsymbol{\delta}}}_{\mathrm{c}}^{*} - \bar{\boldsymbol{\delta}}_{\mathrm{c}} \end{bmatrix} \to_d \mathcal{N}\left(\mathbf{0}_{K+1}, \mathbf{L}^{-1}\mathbf{O}\mathbf{L}^{-1'}\right), \tag{48}$$

where

$$\mathbf{L} \equiv \begin{bmatrix} \mathbf{\Sigma}_{\mathrm{X}} & \mathbf{\Sigma}_{\mathrm{ZX}}' \\ \mathbf{\Sigma}_{\mathrm{ZX}} & \mathbb{J}'\mathbf{\Sigma}_{\mathrm{Z}}\mathbb{J} \end{bmatrix} > 0, \quad \textit{and} \quad \mathbf{O} \equiv \begin{bmatrix} \mathbf{U} & \sigma^2\mathbf{G}\mathbf{H}' \\ \sigma^2\mathbf{H}\mathbf{G}' & \mathbf{H}\mathbf{\Sigma}_{\mathrm{U}}\mathbf{H}', \end{bmatrix} \tag{49}$$

with $\mathbf{U} \equiv \frac{\sigma^2}{T-1}\left[1 + (T-1)\bar{\boldsymbol{\delta}}_{\mathrm{o}}'\mathbf{P}'\mathbf{P}\bar{\boldsymbol{\delta}}_{\mathrm{o}}\right]\mathbf{\Sigma}_{\mathrm{X}} + \begin{bmatrix} 0 & \mathbf{0}_{K_{\mathrm{o}}}' \\ \mathbf{0}_{K_{\mathrm{o}}} & \mathbf{V}'\mathbf{U}_{\epsilon}\mathbf{V} \end{bmatrix}$, $\mathbf{U}_{\epsilon}$, $\mathbf{\Sigma}_{\mathrm{ZB}}$, $\mathbf{\Sigma}_{\mathrm{ZX}}$, $\mathbf{\Sigma}_{\mathrm{U}}$ *and* $\boldsymbol{\mu}_{\mathrm{z},T-1}$ *defined in Assumptions 4 and 8, and in Lemma 2, and where*

$$\begin{aligned} \mathbf{Q} &\equiv \frac{\mathbf{1}_{T-1}}{(T-1)} - \mathbf{P}\bar{\boldsymbol{\delta}}_{\mathrm{o}}, \\ \mathbf{V} &\equiv (\mathbf{Q} \otimes \mathbf{P}) - \frac{\mathrm{vec}(\mathbb{M}_{\tilde{D}})}{T-K-2}\mathbf{Q}'\mathbf{P}, \\ \mathbf{G} &\equiv \left[\mathbf{Q} \otimes \boldsymbol{\mu}_{\mathrm{z},T-1}, \quad \mathbf{Q} \otimes \mathbf{\Sigma}_{\mathrm{ZB}}\right]', \\ \mathbf{H} &\equiv \mathbf{Q}' \otimes \mathbb{J}'. \end{aligned}$$

Proof. See the Online Appendix OA.5.2

The following theorem shows how to construct asymptotically valid standard errors.

Theorem 4 (standard errors of the locally-averaged bias-adjusted CSR OLS estimator)**.** *Under Assumptions 2—8,* $\widehat{\mathrm{Cov}}(\mathbf{Z}_{t-1}, \mathbf{F}_{\mathrm{c},t-1}) = o_p(N^{-1/2})$*, and the identification condition* $\kappa_4 = 0$*, as* $N \to \infty$*,*

$$\hat{\mathbf{L}}^{-1}\,\hat{\mathbf{O}}\,\hat{\mathbf{L}}^{-1'} \quad \to_p \quad \mathbf{L}^{-1}\,\mathbf{O}\,\mathbf{L}^{-1'} \tag{50}$$

where

$$\hat{\mathbf{L}} \equiv \frac{1}{N}\begin{bmatrix} \hat{\mathbf{X}}'\hat{\mathbf{X}} - N\hat{\mathbf{\Lambda}}_1 & \hat{\mathbf{X}}'\mathbf{Z} \\ \mathbf{Z}'\hat{\mathbf{X}} & \mathbf{Z}'\mathbf{Z} \end{bmatrix}, \quad \textit{and} \quad \hat{\mathbf{O}} \equiv \begin{bmatrix} \hat{\mathbf{U}} & \hat{\sigma}^2\hat{\mathbf{G}}\hat{\mathbf{H}}' \\ \hat{\sigma}^2\hat{\mathbf{H}}\hat{\mathbf{G}}' & \hat{\mathbf{H}}\hat{\mathbf{\Sigma}}_{\mathrm{U}}\hat{\mathbf{H}}' \end{bmatrix}, \tag{51}$$

with $\hat{\mathbf{U}} \equiv \frac{\hat{\sigma}^2}{T-1}\left[1+(T-1)\hat{\hat{\boldsymbol{\delta}}}_{\mathrm{o}}^{*\prime}\mathbf{P}'\mathbf{P}\hat{\hat{\boldsymbol{\delta}}}_{\mathrm{o}}^{*}\right](\hat{\boldsymbol{\Sigma}}_{\mathrm{X}}-\hat{\boldsymbol{\Lambda}}_1)+\begin{bmatrix} 0 & \mathbf{0}'_{K_{\mathrm{f}}} \\ \mathbf{0}_{K_{\mathrm{o}}} & \hat{\mathbf{V}}'\hat{\mathbf{U}}_{\epsilon}\hat{\mathbf{V}} \end{bmatrix}$ *and where* $\hat{\mathbf{U}}_{\epsilon} = \mathbf{U}_{\epsilon}(\kappa_4 = 0, \hat{\sigma}^4)$ *is a consistent plug-in estimator of* $\mathbf{U}_{\epsilon} = \mathbf{U}_{\epsilon}(\kappa_4, \sigma^4)$ *obtained by replacing* σ^4 *with*

$$\hat{\sigma}^4 = \frac{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\hat{\epsilon}_{it}^4}{3\,\mathrm{tr}\left(\overline{\mathbb{M}}_{\tilde{\mathrm{D}}}^{(2)}\right)}, \qquad \textit{with} \quad \overline{\mathbb{M}}_{\tilde{\mathrm{D}}}^{(2)} \equiv \frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{M}_{\tilde{\mathrm{D}}_i}\odot\mathbb{M}_{\tilde{\mathrm{D}}_i}\right), \tag{52}$$

recalling $\overline{\mathbb{M}}_{\tilde{\mathrm{D}}} = N^{-1}\sum_{i=1}^{N}\mathbb{M}_{\tilde{\mathrm{D}}_i}$, *with* $\mathbb{M}_{\tilde{\mathbf{D}}_i} = \mathbf{I}_{T-1} - \tilde{\mathbf{D}}_i(\tilde{\mathbf{D}}_i'\tilde{\mathbf{D}}_i)^{-1}\tilde{\mathbf{D}}_i'$, $\tilde{\mathbf{D}}_i = (\mathbf{D}, \tilde{\mathbf{Z}}_i)$, *with* $\mathbf{D} = (\mathbf{1}_{T-1}, \mathbf{F})$, $\hat{\boldsymbol{\Sigma}}_{\mathrm{X}} = N^{-1}\hat{\mathbf{X}}'\hat{\mathbf{X}}$, $\hat{\boldsymbol{\Sigma}}_{\mathrm{ZB}} = N^{-1}\mathbf{Z}'\hat{\mathbf{B}}$, $\hat{\boldsymbol{\mu}}_{\mathrm{z},T-1} = N^{-1}\mathbf{Z}'\mathbf{1}_N$, *and* $\hat{\boldsymbol{\Sigma}}_{\mathrm{U}} \equiv \hat{\sigma}^2\boldsymbol{I}_{T-1}\otimes\mathbf{Z}'\mathbf{Z}/N$, *with* $\hat{\sigma}^2$ *defined in* (34), *and defining*

$$\begin{aligned}
\hat{\mathbf{H}} &\equiv \hat{\mathbf{Q}}'\otimes\mathbb{J}', \hat{\mathbf{Q}} \equiv \frac{\mathbf{1}_{T-1}}{(T-1)} - \mathbf{P}\hat{\hat{\boldsymbol{\delta}}}_{\mathrm{o}}^{*}, \\
\hat{\mathbf{V}} &\equiv (\hat{\mathbf{Q}}\otimes\mathbf{P}) - \frac{\mathrm{vec}(\overline{\mathbb{M}}_{\tilde{\mathrm{D}}})}{T-K-2}\hat{\mathbf{Q}}'\mathbf{P}, \\
\hat{\mathbf{G}} &\equiv \left[\hat{\mathbf{Q}}\otimes\hat{\boldsymbol{\mu}}_{\mathrm{z},T-1}, \quad \hat{\mathbf{Q}}\otimes\hat{\boldsymbol{\Sigma}}_{\mathrm{ZB}}\right]'.
\end{aligned}$$

Proof. See the Online Appendix OA.5.2

Notice that the precision of our risk premia estimators improves when T increases, although the larger is T the more pronounced the smoothness of the risk premia must be for accurate estimation. On the other hand, our average premia estimator remains meaningful even under large-T asymptotics, because it generalizes the conventional two-pass estimator in (14), without requiring any stringent orthogonality assumption. In this case, however, $(\hat{\hat{\boldsymbol{\Gamma}}}_{\mathrm{o}}^{*\prime}, \hat{\hat{\boldsymbol{\delta}}}_{\mathrm{c}}^{*\prime})'$ will accurately estimate the *long-run* average of the time-varying premia, which clearly would not unveil any variation over time in the premia coefficients.[31]

7 Anomalies with Time-Varying Premia: WLS-Based Estimation

Fama and French (2008) and Hou, Chen, and Zhang (2020), among others, recognize that most of the empirical results on asset pricing anomalies can be seriously affected by the presence of micro-cap stocks. Small-cap equities typically show higher returns than large-cap stocks, but they also

[31] For example, following Ang and Kristensen (2012), one could assume that $\left(\boldsymbol{\Gamma}'_{\mathrm{o},t}, \boldsymbol{\delta}'_{\mathrm{c},t}\right)' = \left(\boldsymbol{\Gamma}'_{\mathrm{o}}(\frac{t}{T}), \boldsymbol{\delta}'_{\mathrm{c}}(\frac{t}{T})\right)'$, for some smooth functions $\boldsymbol{\Gamma}_{\mathrm{o}}(\cdot)$ and $\boldsymbol{\delta}_{\mathrm{c}}(\cdot)$. Then, as T goes to infinity, $(\hat{\hat{\boldsymbol{\Gamma}}}_{\mathrm{o}}^{*\prime}, \hat{\hat{\boldsymbol{\delta}}}_{\mathrm{c}}^{*\prime})'$ accurately estimate the *long-run* risk premia $\int_0^1 \begin{bmatrix}\boldsymbol{\Gamma}_{\mathrm{o}}(s) \\ \boldsymbol{\delta}_{\mathrm{c}}(s)\end{bmatrix} ds$, which, although of interest (and assuming that such quantity exists finite), would completely mask any form of time-variation in the risk premia parameters.

tend to have the largest cross-sectional dispersions both in terms of returns and anomaly variables. To mitigate this effect, Hou, Chen, and Zhang (2020) consider a Weighted Least Square (WLS) estimator of the premia parameters, with the weights being proportional to the corresponding stock's market capitalization.

Formally, let $\left(\hat{\mathbf{\Gamma}}^{(\mathrm{w})\prime}_{\mathrm{f},t-1}, \hat{\boldsymbol{\gamma}}^{(\mathrm{w})\prime}_{\mathrm{z},t-1}\right)^{\prime}$ denote the WLS estimator of the $(K+1)$-vector of premia coefficients, where $\hat{\mathbf{\Gamma}}^{(\mathrm{w})}_{\mathrm{f},t-1} \equiv \left(\hat{\gamma}^{(\mathrm{w})\prime}_{0,t-1}, \hat{\boldsymbol{\delta}}^{(\mathrm{w})\prime}_{\mathrm{f},t-1}\right)^{\prime}$ denotes the premia estimator of the zero-beta rate and the K_{f} risk factors, while $\hat{\boldsymbol{\gamma}}^{(\mathrm{w})}_{\mathrm{z},t-1}$ refers to the premia of the K_{z} anomaly variables. Let $\mathbf{W}_{t-1}$ be an $N \times N$ diagonal matrix containing the asset-specific weights at time $t-1$, i.e. $\mathbf{W}_{t-1} \equiv \mathrm{diag}\left(\mathrm{w}_{1,t-1}, ..., \mathrm{w}_{N,t-1}\right)$, where we assume $\mathrm{w}_{i,t} > 0$ for every asset i and period t without great loss of generality. Then, following Hou, Chen, and Zhang (2020), we have:

$$\begin{bmatrix} \hat{\mathbf{\Gamma}}^{(\mathrm{w})}_{\mathrm{f},t-1} \\ \hat{\boldsymbol{\gamma}}^{(\mathrm{w})}_{\mathrm{z},t-1} \end{bmatrix} \equiv \begin{bmatrix} \hat{\mathbf{X}}'\mathbf{W}_{t-1}\hat{\mathbf{X}} & \hat{\mathbf{X}}'\mathbf{W}_{t-1}\mathbf{Z}_{t-1} \\ \mathbf{Z}'_{t-1}\mathbf{W}_{t-1}\hat{\mathbf{X}} & \mathbf{Z}'_{t-1}\mathbf{W}_{t-1}\mathbf{Z}_{t-1} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{X}}'\mathbf{W}_{t-1}\mathbf{R}_t \\ \mathbf{Z}'_{t-1}\mathbf{W}_{t-1}\mathbf{R}_t \end{bmatrix} \tag{53}$$

where, for each stock i, the weight $\mathrm{w}_{i,t-1}$ in $\mathbf{W}_{t-1}$ is given by the corresponding stock market capitalization at time $t-1$. [32]

Similarly to the conventional time-varying OLS estimator defined in (31), we can show that analogous conclusions apply to the WLS estimator in (53). Indeed, whenever one wants to estimate time-varying premia under the traditional large-T–fixed-N setting, we show that the estimator in (53) would be invalid, because it is affected by a random (hence, unpredictable) bias.[33] In the large-N–fixed-T set-up, instead, the WLS estimator in (53) is still contaminated by several sources of bias which, however, can be consistently estimated, yielding our novel bias-adjusted CSR WLS estimator:[34]

$$\begin{bmatrix} \hat{\mathbf{\Gamma}}^{*(\mathrm{w})}_{\mathrm{f},t-1} \\ \hat{\boldsymbol{\gamma}}^{*(\mathrm{w})}_{\mathrm{z},t-1} \end{bmatrix} \equiv \begin{bmatrix} \hat{\mathbf{X}}'\mathbf{W}_{t-1}\hat{\mathbf{X}} - N\hat{\mathbf{\Lambda}}^{(\mathrm{w})}_{1,t-1} & \hat{\mathbf{X}}'\mathbf{W}_{t-1}\mathbf{Z}_{t-1} \\ \mathbf{Z}'_{t-1}\mathbf{W}_{t-1}\hat{\mathbf{X}} & \mathbf{Z}'_{t-1}\mathbf{W}_{t-1}\mathbf{Z}_{t-1} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{X}}'\mathbf{W}_{t-1}\mathbf{R}_t - N\hat{\mathbf{\Lambda}}^{(\mathrm{w})}_{2,t-1} \\ \mathbf{Z}'_{t-1}\mathbf{W}_{t-1}\mathbf{R}_t \end{bmatrix}, \tag{54}$$

[32]Our WLS estimator of the anomalies premia can also be used as an estimator of cross-sectional WLS portfolios, along the lines outlined in Section 6.

[33]We show this result in the Online Appendix OA.5.1 - see Remark OA.19.

[34]See the Online Appendix OA.5.1 - Proposition OA.2.

where $\hat{\boldsymbol{\Gamma}}^{*(\mathrm{w})}_{t-1} \equiv \left(\hat{\gamma}^{*(\mathrm{w})}_{0,t-1}, \hat{\boldsymbol{\delta}}^{*(\mathrm{w})\prime}_{\mathrm{f},t-1}\right)'$, and where we set

$$\hat{\boldsymbol{\Lambda}}^{(\mathrm{w})}_{1,t-1} \equiv \begin{bmatrix} 0 & \mathbf{0}'_{K_f} \\ \mathbf{0}_{K_f} & \hat{\sigma}^{2(\mathrm{w})}_{t-1}\mathbf{P}'\mathbf{P} \end{bmatrix}, \qquad \hat{\boldsymbol{\Lambda}}^{(\mathrm{w})}_{2,t-1} \equiv \hat{\sigma}^{2(\mathrm{w})}_{t-1} \begin{bmatrix} 0 \\ \mathbf{P}'\boldsymbol{\imath}_{t-1,T-1} \end{bmatrix} \tag{55}$$

with

$$\hat{\sigma}^{2(\mathrm{w})}_{t-1} \equiv \frac{\mathrm{tr}(\hat{\boldsymbol{\epsilon}}\mathbf{W}_{t-1}\hat{\boldsymbol{\epsilon}}')}{N(T-K-2)} \tag{56}$$

Before establishing the asymptotic properties of the WLS estimator in (54), it is important to highlight some necessary remarks. The choice of using stock's market capitalization in the $\mathbf{W}_{t-1}$ matrix makes the weighting scheme parameter-free. On one hand, this simplifies the WLS analysis, where the weights are instead typically defined as functions of unknown parameters (to be estimated) or set to be inversely proportional to the regression-error variance. On the other hand, however, market capitalization could be very likely correlated - both cross-sectionally and over time - with returns and other anomalies, making the asymptotic analysis of the estimator non-trivial. For this reason, we need to impose some conditions on the sample moments of anomalies, weights and asset-specific errors. Specifically, we assume that each asset-specific error is uncorrelated with past values of both anomaly variables and weights, but could be potentially correlated with their contemporary and future values (see Assumption 12 in Appendix A.1.1).

Moreover, the behavior of the weights plays a crucial role in determining the statistical properties of the WLS estimator, especially when $N \to \infty$. In particular, a condition that the weights should satisfy is the so-called *granuarity* assumption, which guarantees that the weights dissipate to zero sufficiently fast for every asset, as $N \to \infty$. When the granularity assumption fails, then the WLS estimator exhibits a random limit, making both estimation and inference invalid.Therefore, in the following theorems, we establish the limiting properties of the WLS estimator in (54) under the assumption that granularity holds (see Assumption 9 in Appendix A.1.1).

Theorem 5. *As $N \to \infty$, and under Assumptions 2–12,*

(i)
$$\hat{\boldsymbol{\Gamma}}^{*(\mathrm{w})}_{\mathrm{f},t-1} - \boldsymbol{\Gamma}_{\mathrm{f},t-1} = O_p\left(\frac{1}{\sqrt{N}}\right), \qquad \hat{\boldsymbol{\gamma}}^{*(\mathrm{w})}_{\mathrm{z},t-1} - \boldsymbol{\gamma}_{\mathrm{z},t-1} = O_p\left(\frac{1}{\sqrt{N}}\right). \tag{57}$$

(ii)
$$\sqrt{N}\begin{bmatrix}\hat{\boldsymbol{\Gamma}}^{*(\mathrm{w})}_{\mathrm{f},t-1}-\boldsymbol{\Gamma}_{\mathrm{f},t-1}\\ \hat{\boldsymbol{\gamma}}^{*(\mathrm{w})}_{\mathrm{z},t-1}-\boldsymbol{\gamma}_{\mathrm{z},t-1}\end{bmatrix}\to_d \mathcal{N}\left(\mathbf{0}_{\mathrm{K}+1},\mathbf{L}^{-1}_{t-1}\mathbf{O}^{(\mathrm{w})}_{t-1}\mathbf{L}^{-1}_{t-1}\right), \tag{58}$$
where $\mathbf{L}_{t-1}$ is the same as in Theorem 1, and for for some $\mathbf{O}^{(\mathrm{w})}_{t-1}$ defined in (OA.45).[35]

Proof. See Appendix OA.4.

The next theorem shows how to construct asymptotically-valid standard errors for the WLS estimator.[36]

Theorem 6. *As $N\to\infty$, under Assumptions 2–12, and the identification condition $\kappa_4=0$,*
$$\hat{\mathbf{L}}^{(\mathrm{w})-1}_{t-1}\,\hat{\mathbf{O}}^{(\mathrm{w})}_{t-1}\,\hat{\mathbf{L}}^{(\mathrm{w})-1\prime}_{t-1}\quad\to_p\quad \mathbf{L}^{-1}_{t-1}\,\mathbf{O}^{(\mathrm{w})}_{t-1}\,\mathbf{L}^{-1\prime}_{t-1} \tag{59}$$
where
$$\hat{\mathbf{L}}^{(\mathrm{w})}_{t-1}\equiv\frac{1}{N}\begin{bmatrix}\hat{\mathbf{X}}'\hat{\mathbf{X}}-N\hat{\boldsymbol{\Lambda}}^{(\mathrm{w})}_{1,t-1} & \hat{\mathbf{X}}'\mathbf{W}_{t-1}\mathbf{Z}_{t-1}\\ \mathbf{Z}'_{t-1}\mathbf{W}_{t-1}\hat{\mathbf{X}} & \mathbf{Z}'_{t-1}\mathbf{W}_{t-1}\mathbf{Z}_{t-1}\end{bmatrix},\ \textit{and}\quad \hat{\mathbf{O}}^{(\mathrm{w})}_{t-1}\equiv\hat{\lambda}_{t-1}\begin{bmatrix}\hat{\boldsymbol{\mu}}_{\mathrm{x}}\hat{\boldsymbol{\mu}}'_{\mathrm{x}} & \hat{\boldsymbol{\mu}}_{\mathrm{x}}\hat{\boldsymbol{\mu}}'_{\mathrm{z}}\\ \hat{\boldsymbol{\mu}}_{\mathrm{z}}\hat{\boldsymbol{\mu}}'_{\mathrm{x}} & \hat{\boldsymbol{\mu}}_{\mathrm{z}}\hat{\boldsymbol{\mu}}'_{\mathrm{z}}\end{bmatrix}+\hat{\mathbf{M}}^{(\mathrm{w})}_{t-1} \tag{60}$$
with
$$\hat{\mathbf{M}}^{(\mathrm{w})}_{t-1}\equiv\begin{bmatrix}0 & \mathbf{0}'_{K_{\mathrm{f}}} & \mathbf{0}'_{K_{\mathrm{z}}}\\ \mathbf{0}_{K_{\mathrm{f}}} & \hat{\mu}^2_{\mathrm{w},t-1}\,\hat{\mathbf{V}}^{(\mathrm{w})\prime}_{t-1}\hat{\mathbf{U}}_{\epsilon}\hat{\mathbf{V}}^{(\mathrm{w})}_{t-1} & \mathbf{0}_{K_{\mathrm{f}}\times K_{\mathrm{z}}}\\ \mathbf{0}_{K_{\mathrm{z}}} & \mathbf{0}_{K_{\mathrm{z}}\times K_{\mathrm{f}}} & \hat{\mathbf{H}}^{(\mathrm{w})}_{t-1}\hat{\boldsymbol{\Sigma}}^{(\mathrm{w})}_{\mathrm{U}}\hat{\mathbf{H}}^{(\mathrm{w})\prime}_{t-1}+\hat{\mathbf{S}}^{(\mathrm{w})}_{t-1}+\hat{\mathbf{S}}^{(\mathrm{w})}_{t-1}\end{bmatrix} \tag{61}$$
setting $\hat{\boldsymbol{\mu}}_{\mathrm{x}}\equiv(1,\hat{\boldsymbol{\mu}}'_{\beta})'$, $\hat{\boldsymbol{\mu}}_{\beta}\equiv N^{-1}\hat{\mathbf{B}}\mathbf{1}_N$, $\hat{\boldsymbol{\mu}}_{\mathrm{z}}\equiv N^{-1}\mathbb{J}'\mathbf{Z}'\mathbf{1}_N$, $\hat{\mu}^2_{\mathrm{w},t-1}\equiv N^{-1}\mathbf{1}'_N\mathbf{W}^2_{t-1}\mathbf{1}_N$, $\hat{\boldsymbol{\Sigma}}^{(\mathrm{w})}_{\mathrm{U}}\equiv(\hat{\sigma}^{2(\mathrm{w})}_{t-1}\mathbf{I}_{T-1}\otimes N^{-1}\mathbf{Z}'\mathbf{Z})$, $\hat{\boldsymbol{\Sigma}}_{\mathrm{ZW}}\equiv N^{-1}\mathbf{Z}'\mathbf{W}$, and $\hat{\boldsymbol{\Sigma}}_{\mathrm{V}}\equiv\left(\hat{\sigma}^{2(\mathrm{w})}_{t-1}\mathbf{I}_{T-1}\otimes N^{-1}\sum_{i=1}^N\mathbf{w}_i\mathbf{w}'_i\right)$, where $\mathbf{w}_i\equiv(\mathrm{w}_{i,1},...,\mathrm{w}_{i,T-1})'$, with $\hat{\boldsymbol{\Lambda}}^{(\mathrm{w})}_{1,t-1}$ and $\hat{\sigma}^{2(\mathrm{w})}_{t-1}$ defined in (55) *and* (56)*, respectively, and we define the following matrices*
$$\begin{aligned}\hat{\mathbf{Q}}^{(\mathrm{w})}_{t-1}&\equiv\boldsymbol{\imath}_{t-1,T-1}-\mathbf{P}\hat{\boldsymbol{\delta}}^{*(\mathrm{w})}_{\mathrm{f},t-1},\quad \hat{\mathbf{H}}^{(\mathrm{w})}_{t-1}\equiv\hat{\mathbf{Q}}^{(\mathrm{w})\prime}_{t-1}\otimes\mathbb{J}'_{t-1},\\ \hat{\mathbf{Y}}_{t-1}&\equiv\hat{\mathbf{Q}}^{(\mathrm{w})}_{t-1}\otimes\boldsymbol{\imath}_{t-1,T-1},\quad \hat{\lambda}_{t-1}\equiv\hat{\mathbf{Y}}'_{t-1}\hat{\boldsymbol{\Sigma}}_{\mathrm{V}}\hat{\mathbf{Y}}_{t-1},\\ \hat{\mathbf{S}}_{t-1}&\equiv\hat{\boldsymbol{\mu}}_{\mathrm{z}}\hat{\mathbf{Y}}'_{t-1}(\hat{\sigma}^{2(\mathrm{w})}_{t-1}\boldsymbol{I}_{T-1}\otimes\hat{\boldsymbol{\Sigma}}'_{\mathrm{ZW}})\hat{\mathbf{H}}^{(\mathrm{w})\prime}_{t-1},\\ \hat{\mathbf{V}}^{(\mathrm{w})}_{t-1}&\equiv(\hat{\mathbf{Q}}^{(\mathrm{w})}_{t-1}\otimes\mathbf{P})-\left(\frac{\mathrm{vec}(\overline{\mathbf{M}}_{\tilde{\mathrm{D}}})}{T-K-2}\right)\hat{\mathbf{Q}}^{(\mathrm{w})\prime}_{t-1}\mathbf{P},\end{aligned}$$
where $\hat{\mathbf{U}}_{\epsilon}$ is defined in Theorem 2.

[35]To ease the exposition, the precise definition of $\mathbf{O}^{(\mathrm{w})}_{t-1}$ has been relegated to the proof of the theorem (see (OA.45)).
[36]We report the results without the proof, as if follows closely the proof of Theorem 2.

8 Anomalies with Time-Varying Premia: Global Misspecification

All the results so far established assume that the asset pricing model (11) is *correctly specified*, meaning that the true model does not omit any relevant variable (either a risk factor or an anomaly variable) or, alternatively, it does not include any irrelevant one.[37] When this assumption is violated - a very likely scenario - then the issue of *global misspecification* arises which, if ignored, could seriously compromise our inferential results.[38] Indeed, misspecification affects the standard errors obtained in the previous sections, with the risk of making an anomaly appear significant when instead its premium is null or, alternatively, making it statistically irrelevant when instead its effect is non-zero. Therefore, the objective of this section is to extend our methodology and robustify our inferential results to the case of a generic deviation from exact pricing of unknown form, i.e., global misspecification.

Consider the asset pricing restriction in (6) and assume now that, beyond the presence of anomalies, there is a further deviation from exact pricing, due to potential global misspecification. That is, assume that:

$$\mathrm{E}[R_{i,t}|I_{t-1}, \boldsymbol{\Pi}] = \tilde{\gamma}_{0,t-1} + \tilde{\boldsymbol{\delta}}'_{\mathrm{f},t-1}\boldsymbol{\beta}_i + \tilde{\boldsymbol{\gamma}}'_{z,t-1}\mathbf{z}_{i,t-1} + \mathrm{m}_{i,t-1}, \tag{62}$$

where $\mathrm{m}_{i,t-1}$ represents an additional pricing error, accounting for the fact the postulated model could potentially specify the wrong set of variables. In other words, one could think that the overall deviation from exact pricing in (3) has now a semiparametric structure, with the parametric part being linear in the $\mathbf{z}_{i,t-1}$, and a non-parametric component coming from the misspecification error $\mathrm{m}_{i,t-1}$, which is completely unspecified. Our objective is to test for the statistical relevance of the anomalies $\mathbf{z}_{i,t-1}$, *regardless* of whether they represent or not the full set of variables describing the

[37] A different form of misspecification, not explored in this paper, occurs when one (or more) vector of betas is a linear combination of the other ones, implying that $\mathbf{X}$ is not full-column rank. This happens, for example, when one or more of the candidate risk factors has zero (or almost zero) betas, a situation which is often referred to as the issue of *spurious* or *useless* factors. See, e.g., Jagannathan and Wang (1998), Kan and Zhang (1999b,a), Kleibergen (2009), Gospodinov, Kan, and Robotti (2014), Bryzgalova (2014), Burnside (2016), Ahn, Horenstein, and Wang (2018), Kleibergen and Zhan (2014, 2020), and Anatolyev and Mikusheva (2020), among others. The less restrictive cases of semi-strong, when $\mathbf{B}'\mathbf{B}/N = o(1)$ (see Connor and Korajczyk (2022)), and weak factors, when $\mathbf{B}'\mathbf{B} = O(1)$ (see Lettau and Pelger (2020) and Giglio, Xiu, and Zhang (2021)), are also ruled out by our assumptions. Kim, Raponi, and Zaffaroni (2020) develop an inferential procedure to test for spurious and weak factors, valid when N is large and T is fixed.

[38] Global misspecification has been studied widely in the large-T sampling scheme; see Jagannathan and Wang (1998), Shanken and Zhou (2007), Hou and Kimmel (2006), and Kan, Robotti, and Shanken (2013), among others. Gagliardini, Ossola, and Scaillet (2016) and Raponi, Robotti, and Zaffaroni (2020) show how to robustify their risk premia estimator to global misspecification in the large-N–large-T and in the large-N–fixed-T settings, respectively.

true asset pricing model, that is regardless of whether $\mathrm{m}_{i,t-1}$ is zero or not.

Although we do not impose any parameterization on $\mathrm{m}_{i,t-1}$, simple considerations suggest that $\mathrm{m}_{i,t-1}$ might be cross-sectionally correlated with $\epsilon_{i,s}$, for every $s \leq t$. As an illustrative example, consider the case where one omits some relevant risk factors and anomaly variables from the true model (11), and no other sources of misspecification are present. In this circumstance, the asset-pricing model can be written as

$$\begin{aligned} \mathbf{R}_t &= \mathbf{Z}_{t-1}\tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1} + \mathbf{X}\tilde{\boldsymbol{\Gamma}}_{t-1} + \boldsymbol{\epsilon}_t, \quad \text{with} \\ \boldsymbol{\epsilon}_t &= \breve{\boldsymbol{\epsilon}}_t + \breve{\mathbf{Z}}_{t-1}\breve{\boldsymbol{\gamma}}_{\mathrm{z},t-1} + \breve{\mathbf{B}}\breve{\boldsymbol{\delta}}_{\mathrm{f},t-1}, \end{aligned} \tag{63}$$

where $\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} = (\tilde{\gamma}_{0,t-1}, \tilde{\boldsymbol{\delta}}'_{\mathrm{f},t-1})'$, $\breve{\mathbf{Z}}_{t-1}$ represents the $N \times \breve{K}_{\mathrm{z}}$ set of *omitted* anomalies with corresponding premia $\breve{\boldsymbol{\gamma}}_{\mathrm{z},t-1}$, and where $\breve{\mathbf{B}}$ is the $N \times \breve{K}_{\mathrm{f}}$ matrix of loadings associated with the $\breve{K}_{\mathrm{f}}$ *omitted* risk factors $\breve{\mathbf{f}}_t$, having ex-post risk premia $\breve{\boldsymbol{\delta}}_{\mathrm{f},t-1}$. [39] Finally, $\breve{\boldsymbol{\epsilon}}_t$ represents the *genuine* asset-specific component of asset returns, coinciding with $\boldsymbol{\epsilon}_t$ in the case of correct model specification. Then, combining (62) with (63), we get

$$\mathbf{m}_{t-1} = (\mathrm{m}_{1,t-1}, \cdots, \mathrm{m}_{N,t-1})' = \breve{\mathbf{Z}}_{t-1}\breve{\boldsymbol{\gamma}}_{\mathrm{z},t-1} + \breve{\mathbf{B}}\breve{\boldsymbol{\delta}}_{\mathrm{f},t-1},$$

implying that $\mathbf{m}_t$ and $\boldsymbol{\epsilon}_s$ are *cross-sectionally* correlated, through either $\breve{\mathbf{Z}}_t$ or $\breve{\mathbf{B}}$, whenever $s \leq t$, unless 'of course the premia $\breve{\boldsymbol{\gamma}}_{\mathrm{z},t-1}$ and $\breve{\boldsymbol{\delta}}_{\mathrm{f},t-1}$ are null, that is when model (11) is correctly specified.

The relationship in (62) implies that the parameters $\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1}$, and $\tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1}$, represent the so-called *pseudo*-true values of the premia coefficients. Formally, let $\mathbf{c}_{\mathrm{z}}$ and $\mathbf{c}_{\mathrm{f}}$ denote two arbitrary vectors of dimension K_{z} and $K_{\mathrm{f}}+1$, respectively. Then, by generalizing Shanken and Zhou (2007) and Raponi, Robotti, and Zaffaroni (2020), we define the pseudo-true premia parameters

$$(\tilde{\boldsymbol{\Gamma}}'_{\mathrm{f},t-1}, \tilde{\boldsymbol{\gamma}}'_{\mathrm{z},t-1})' = \underset{\mathbf{c}_{\mathrm{z}},\mathbf{c}_{\mathrm{f}}}{\mathrm{argmin}} \frac{1}{N}\Big(\mathrm{E}[\mathbf{R}_t|\mathbf{I}_{t-1}, \boldsymbol{\Pi}] - \mathbf{Z}_{t-1}\mathbf{c}_{\mathrm{z}} - \mathbf{X}\mathbf{c}_{\mathrm{f}}\Big)'\Big(\mathrm{E}[\mathbf{R}_t|\mathbf{I}_{t-1}, \boldsymbol{\Pi}] - \mathbf{Z}_{t-1}\mathbf{c}_{\mathrm{z}} + \mathbf{X}\mathbf{c}_{\mathrm{f}}\Big), \tag{64}$$

When the model is correctly specified, then $\tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} = \boldsymbol{\Gamma}_{\mathrm{f},t-1}$ and $\tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1} = \boldsymbol{\gamma}_{t-1,z}$, that is we recover the vector of risk and anomalies' premia of Section 5.

[39] For convenience, assume that $(\mathbf{D}, \mathbf{Z}_i)'(\breve{\mathbf{F}}, \breve{\mathbf{Z}}_i) = \mathbf{0}_{K+1\times\breve{K}}$, with $\breve{K} = \breve{K}_{\mathrm{f}} + \breve{K}_{\mathrm{z}}$ and that $(\mathbf{X}, \mathbf{Z}_{t-1})'(\breve{\mathbf{B}}, \breve{\mathbf{Z}}_{t-1}) = \mathbf{0}_{K+1\times\breve{K}}$. This is with only a small loss of generality because, as discussed above, the estimated time-series regression of $\mathbf{R}_t$ on $\mathbf{f}_t$ and $\mathbf{Z}_{t-1}$ can be always re-arranged so that $(\mathbf{D}, \mathbf{Z}_i)$ and $(\breve{\mathbf{F}}, \breve{\mathbf{Z}}_i)$ are made orthogonal to each other for every i. The same applies for the estimated cross-sectional regression of $\mathbf{R}_i$ on $\boldsymbol{\beta}_i$ and $\mathbf{Z}_i$, leading to orthogonality between $\mathbf{X}, \mathbf{Z}_{t-1}$ and $\breve{\mathbf{B}}, \breve{\mathbf{Z}}_{t-1}$.

The cross-sectional correlation between $\mathbf{m}_t$ and $\boldsymbol{\epsilon}_s$, arising as a result of global misspecification, induces further biases to the CSR OLS estimator, which nevertheless can be consistently estimated, leading to our novel misspecification-robust premia estimators $\hat{\boldsymbol{\Gamma}}^{*(\mathrm{m})}_{\mathrm{f},t-1} \equiv (\hat{\gamma}^{*(\mathrm{m})}_{0,t-1}, \hat{\boldsymbol{\delta}}^{*(\mathrm{m})\prime}_{\mathrm{f},t-1})'$ and $\hat{\boldsymbol{\gamma}}^{*(\mathrm{m})}_{\mathrm{z},t-1}$, defined as follows.[40]

$$\begin{bmatrix} \hat{\boldsymbol{\Gamma}}^{*(\mathrm{m})}_{\mathrm{f},t-1} \\ \\ \hat{\boldsymbol{\gamma}}^{*(\mathrm{m})}_{\mathrm{z},t-1} \end{bmatrix} \equiv \begin{bmatrix} \hat{\mathbf{X}}'\hat{\mathbf{X}} - N(\hat{\boldsymbol{\Lambda}}_1 + \hat{\boldsymbol{\Lambda}}^{(\mathrm{m})}_{1,t-1}) & \hat{\mathbf{X}}'\mathbf{Z}_{t-1} - N\hat{\boldsymbol{\Lambda}}^{(\mathrm{m})}_{3,t-1} \\ \\ \mathbf{Z}'_{t-1}\hat{\mathbf{X}} & \mathbf{Z}'_{t-1}\mathbf{Z}_{t-1} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{X}}'\mathbf{R}_t - N(\hat{\boldsymbol{\Lambda}}_{2,t-1} + \hat{\boldsymbol{\Lambda}}^{(\mathrm{m})}_{2,t-1}) \\ \\ \mathbf{Z}'_{t-1}\mathbf{R}_t \end{bmatrix}, \tag{65}$$

setting $\hat{\boldsymbol{\Lambda}}_1$ and $\hat{\boldsymbol{\Lambda}}_{2,t-1}$ are defined in (33), and where we define the additional bias-correction terms

$$\hat{\boldsymbol{\Lambda}}^{(\mathrm{m})}_{1,t-1} \equiv \frac{1}{N}\begin{bmatrix} \mathbf{0}'_{K_{\mathrm{f}}+1} \\ \mathbf{P}'\hat{\boldsymbol{\Psi}}_{\mathrm{D}\hat{\mathrm{X}}} \end{bmatrix}, \quad \hat{\boldsymbol{\Lambda}}^{(\mathrm{m})}_{2,t-1} \equiv \frac{1}{N}\begin{bmatrix} 0 \\ \mathbf{P}'\hat{\boldsymbol{\Psi}}_{\mathrm{DR}} - \hat{\sigma}^2\mathbf{P}'\hat{\boldsymbol{\Psi}}_{\mathrm{D}\tilde{\mathrm{D}}} \end{bmatrix}, \text{ and } \hat{\boldsymbol{\Lambda}}^{(\mathrm{m})}_{3,t-1} \equiv \frac{1}{N}\begin{bmatrix} \mathbf{0}'_{K_z} \\ \mathbf{P}'\hat{\boldsymbol{\Psi}}_{\mathrm{DZ}} \end{bmatrix}, \tag{66}$$

with $\hat{\boldsymbol{\Psi}}_{\mathrm{D}\hat{\mathrm{X}}} \equiv \begin{bmatrix} \mathbf{M}^{(-1)}_{\mathrm{D},t-1}\hat{\boldsymbol{\epsilon}}\hat{\mathbf{X}} \\ \mathbf{0}_{(T-t+1)\times(K_{\mathrm{f}}+1)} \end{bmatrix}$, $\hat{\boldsymbol{\Psi}}_{\mathrm{DZ}} \equiv \begin{bmatrix} \mathbf{M}^{(-1)}_{\mathrm{D},t-1}\hat{\boldsymbol{\epsilon}}\mathbf{Z}_{t-1} \\ \mathbf{0}_{(T-t+1)\times K_{\mathrm{z}}} \end{bmatrix}$, $\hat{\boldsymbol{\Psi}}_{\mathrm{DR}} \equiv \begin{bmatrix} \mathbf{M}^{(-1)}_{\mathrm{D},t-1}\hat{\boldsymbol{\epsilon}}\mathbf{R}_t \\ \mathbf{0}_{T-t+1} \end{bmatrix}$, and $\hat{\boldsymbol{\Psi}}_{\mathrm{D}\tilde{\mathrm{D}}} \equiv \begin{bmatrix} \mathbf{M}^{(-1)}_{\mathrm{D},t-1}\mathbb{M}_{\mathrm{D}}\boldsymbol{\imath}_{t-1,T-1} \\ \mathbf{0}_{T-t+1} \end{bmatrix}$, setting the $(t-2)\times(T-1)$ matrix $\mathbf{M}^{(-1)}_{\mathrm{D},t-1} \equiv \mathbb{M}^{-1}_{11}[\mathbf{I}_{t-2}, \mathbf{0}_{(t-2)\times(T-t+1)}]$, where $\mathbb{M}_{11}$ denotes the $(t-2)\times(t-2)$ top-left block of $\mathbb{M}_D = \mathbf{I}_{T-1} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$.[41]

The following theorem derives the asymptotic properties of our robust estimator, extending the results of Theorem 1.

Theorem 7. *As $N \to \infty$, under Assumptions 2–8 and 13*

(i)

$$\hat{\boldsymbol{\Gamma}}^{*(\mathrm{m})}_{\mathrm{f},t-1} - \tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} = O_p\left(\frac{1}{\sqrt{N}}\right), \quad \hat{\boldsymbol{\gamma}}^{*(\mathrm{m})}_{\mathrm{z},t-1} - \tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1} = O_p\left(\frac{1}{\sqrt{N}}\right). \tag{67}$$

(ii)

$$\sqrt{N}\begin{bmatrix} \hat{\boldsymbol{\Gamma}}^{*(\mathrm{m})}_{\mathrm{f},t-1} - \tilde{\boldsymbol{\Gamma}}_{\mathrm{f},t-1} \\ \\ \hat{\boldsymbol{\gamma}}^{*(\mathrm{m})}_{\mathrm{z},t-1} - \tilde{\boldsymbol{\gamma}}_{\mathrm{z},t-1} \end{bmatrix} \to_d \mathcal{N}\left(\mathbf{0}_{\mathrm{K}+1}, \mathbf{L}^{-1}_{t-1}\,\mathbf{O}^{(\mathrm{m})}_{t-1}\,\mathbf{L}_{t-1}\right) \tag{68}$$

[40]See Section OA.5.3 for details of the derivation of (65).

[41]We use the partition $\mathbb{M}_D = \begin{bmatrix} \mathbb{M}_{11} & \mathbb{M}_{12} \\ \mathbb{M}_{21} & \mathbb{M}_{22} \end{bmatrix}$.

for some $\boldsymbol{\Omega}^{(\mathrm{m})}_{t-1}$ *defined in* (OA.47)*, and* $\mathbf{O}^{(\mathrm{m})}_{t-1} \equiv \mathbf{O}_{t-1} + \boldsymbol{\Omega}^{(\mathrm{m})}_{t-1}$ *with* $\mathbf{L}_{t-1} > 0$ *and* $\mathbf{O}_{t-1}$ *being the same as in Theorem 1.*[42]

Proof. See Appendix OA.4.

In Theorem OA.3 of the Online Appendix we establish $\hat{\mathbf{L}}^{(\mathrm{m})}_{t-1} \to_p \mathbf{L}_{t-1}$ and $\hat{\boldsymbol{\Omega}}^{(\mathrm{m})}_{t-1} \to_p \boldsymbol{\Omega}^{(\mathrm{m})}_{t-1}$, as $N \to \infty$, under the same assumptions of Theorem 7 and $\kappa_4 = 0$, for estimators $\hat{\mathbf{L}}^{(\mathrm{m})}_{t-1}$ and $\hat{\boldsymbol{\Omega}}^{(\mathrm{m})}_{t-1}$.

9 Measuring Anomalies' Contribution: Cross-Sectional $\mathbf{R^2}$ Test

Despite the considerable literature on asset pricing anomalies, how much of the cross-sectional variation in expected returns is accounted for by betas and how much by anomalies is still unclear and it still represents a challenging question.

Offering a simple criterion that can answer this question and allow to conduct formal inference on (joint) anomalies' contribution is the objective of this section. Following Chordia, Goyal, and Shanken (2015), one could consider the ratios of the (cross-sectional) variance of the beta component and of the characteristics component, with respect to the overall (cross-sectional) variance of average returns, to measure their relative contribution. Specifically, suppose one has estimated the model (30) using our bias-adjusted CSR OLS estimator, hence obtaining:

$$\mathbf{R}_t = \hat{\mathbf{X}}\hat{\boldsymbol{\Gamma}}^*_{\mathrm{f},t-1} + \mathbf{Z}_{t-1}\hat{\boldsymbol{\gamma}}^*_{\mathrm{z},t-1} + \hat{\boldsymbol{\eta}}_t, \tag{69}$$

where $\hat{\boldsymbol{\eta}}_t$ indicates the $N \times 1$ vector of residuals. Then, the fraction of the overall variance explained by the anomaly variables $\mathbf{Z}_{t-1}$ (at any point in time) would simply be

$$\hat{R}^{2(\mathrm{bench})}_{\mathrm{z},t-1} \equiv \frac{\hat{\boldsymbol{\gamma}}^{*\prime}_{\mathrm{z},t-1}\mathbf{Z}'_{t-1}\mathbb{M}_{1_N}\mathbf{Z}_{t-1}\hat{\boldsymbol{\gamma}}^*_{\mathrm{z},t-1}}{\mathbf{R}'_t\mathbb{M}_{1_N}\mathbf{R}_t}. \tag{70}$$

However, despite being a very simple and intuitive measure, the R-squared in (70) could lead to several problems. First, since beta and anomaly components are not necessarily orthogonal cross-sectionally, this can lead to a fraction of the cross-sectional variance explained by the betas and by the anomaly variables - expressed by the sum of the corresponding R^2 - that is jointly greater than 100%. In addition, while orthogonality between CSR residuals and the regressors (both $\hat{\mathbf{X}}$ and

[42]To ease the exposition, the definition of $\boldsymbol{\Omega}^{(\mathrm{m})}_{t-1}$ has been relegated to the proof of the theorem (see (OA.47)).

$\mathbf{Z}_{t-1}$) is, by construction, warranted by the conventional CSR OLS estimator in (31), this does not hold when considering our bias-adjusted estimator $\left(\hat{\boldsymbol{\Gamma}}^{*\prime}_{\mathrm{f},t-1}, \hat{\gamma}^{*\prime}_{\mathrm{z},t-1}\right)'$ of (32), implying that $\hat{R}^{2(\mathrm{bench})}_{t-1}$ is even wrongly centred.[43]

To overcome such (lack of) orthogonality issues, let us rearrange the estimated asset pricing model (69) as follows:

$$\begin{aligned}\mathbf{R}_t &= \hat{\mathbf{X}}\hat{\boldsymbol{\Gamma}}^*_{\mathrm{f},t-1} + \mathbf{Z}_{t-1}\hat{\gamma}^*_{\mathrm{z},t-1} + \mathbb{P}_{[\hat{\mathbf{X}},\mathbf{Z}_{t-1}]}\hat{\boldsymbol{\eta}}_t + \mathbb{M}_{[\hat{\mathbf{X}},\mathbf{Z}_{t-1}]}\hat{\boldsymbol{\eta}}_t \\ &= \mathbb{P}_{\hat{\mathbf{X}}}\left(\hat{\mathbf{X}}\hat{\boldsymbol{\Gamma}}^*_{\mathrm{f},t-1} + \mathbf{Z}_{t-1}\hat{\gamma}^*_{\mathrm{z},t-1} + \mathbb{P}_{[\hat{\mathbf{X}},\mathbf{Z}_{t-1}]}\hat{\boldsymbol{\eta}}_t\right) + \mathbb{M}_{\hat{\mathbf{X}}}\left(\mathbf{Z}_{t-1}\hat{\gamma}^*_{\mathrm{z},t-1} + \mathbb{P}_{[\hat{\mathbf{X}},\mathbf{Z}_{t-1}]}\hat{\boldsymbol{\eta}}_t\right) + \mathbb{M}_{[\hat{\mathbf{X}},\mathbf{Z}_{t-1}]}\hat{\boldsymbol{\eta}}_t.\end{aligned} \tag{71}$$

where we use the notation $\mathbb{M}_A \equiv \mathbf{I}_a - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' \equiv \mathbf{I}_a - \mathbb{P}_A$, with $\mathbb{P}_A \equiv \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$, for a generic matrix $\mathbf{A}$ of dimension $a \times b$ and rank $b < a$. Notice that, by construction, the three terms on the right-hand side of (71) are now mutually orthogonal, the second term reflecting the joint contribution of $\mathbf{Z}_{t-1}$. This yields our proposed R-squared test statistic:

$$\hat{R}^2_{\mathrm{z},t-1} \equiv \frac{\left(\hat{\gamma}^{*\prime}_{\mathrm{z},t-1}\mathbf{Z}'_{t-1} + \hat{\boldsymbol{\eta}}'_t\mathbb{P}_{[\hat{\mathbf{X}},\mathbf{Z}_{t-1}]}\right)\mathbb{M}_{\hat{X}}\left(\mathbf{Z}_{t-1}\hat{\gamma}^*_{\mathrm{z},t-1} + \mathbb{P}_{[\hat{\mathbf{X}},\mathbf{Z}_{t-1}]}\hat{\boldsymbol{\eta}}_t\right)}{\mathbf{R}'_t\mathbb{M}_{1_N}\mathbf{R}_t}, \tag{72}$$

which satisfies $0 \leq \hat{R}^2_{\mathrm{z},t-1} \leq 1$. Notice that (72) represents a meaningful quantity, which allows us to disentangle the contribution of that portion of anomalies that is *unexplained* by - i.e., orthogonal to - the loadings (through the term $\mathbb{M}_{\hat{\mathbf{X}}}\mathbf{Z}_{t-1}$), as well as the contribution that might arise from the term $\mathbb{P}_{[\hat{\mathbf{X}},\mathbf{Z}_{t-1}]}\hat{\boldsymbol{\eta}}_t$, which is now not guaranteed to be null in general.

Armed with $\hat{R}^2_{\mathrm{z},t-1}$, one could test for the null hypothesis of zero anomalies' contribution and, in case of rejection, construct an asymptotically valid confidence interval for it. This would require establishing the limiting statistical properties of $\hat{R}^2_{\mathrm{z},t-1}$, in particular its non-standard limiting distribution, occurring when $\boldsymbol{\gamma}_{\mathrm{z},t-1} = \mathbf{0}_{K_{\mathrm{z}}}$. Therefore, in the following, we derive the asymptotic distribution of $\hat{R}^2_{\mathrm{z},t-1}$ distinguishing between the two complementary cases of zero and non-zero anomalies' premia.

Theorem 8 (R^2 test of anomalies' contribution)**.** *Set the R^2 test statistic equal to*

$$\mathcal{T}^2_{\mathrm{z},t-1} \equiv N\left(\hat{R}^2_{\mathrm{z},t-1} - \frac{\hat{\boldsymbol{\eta}}'_t\mathbb{P}_{[\hat{\mathbf{X}},\mathbf{Z}_{t-1}]}\mathbb{M}_{\hat{\mathbf{X}}}\mathbb{P}_{[\hat{\mathbf{X}},\mathbf{Z}_{t-1}]}\hat{\boldsymbol{\eta}}_t + 2\,\hat{\boldsymbol{\eta}}'_t\mathbb{P}_{[\hat{\mathbf{X}},\mathbf{Z}_{t-1}]}\mathbb{M}_{\hat{\mathbf{X}}}\mathbf{Z}_{t-1}\hat{\gamma}^*_{\mathrm{z},t-1}}{\mathbf{R}'_t\mathbb{M}_{1_N}\mathbf{R}_t}\right). \tag{73}$$

[43] Finally, notice that when the (true) anomalies' premia $\boldsymbol{\gamma}_{\mathrm{z},t-1}$ are zero, then $\hat{R}^{2(\mathrm{bench})}_{\mathrm{z},t-1}$ will converge to zero in probability. This implies that we face a boundary problem - as necessarily $\hat{R}^{2(\mathrm{bench})}_{t-1} \geq 0$ - which could lead to a non-standard limiting distribution of the test statistic, under the null hypothesis of zero anomalies' premia.

Under Assumptions 2–8, as $N \to \infty$, then:

(i) *When* $\boldsymbol{\gamma}_{\mathrm{z},t-1} = \mathbf{0}_{K_{\mathrm{z}}}$,

$$\mathcal{T}_{\mathrm{z},t-1} \to_d \sum_{j=1}^{K_{\mathrm{z}}} d_{j,t-1}\chi^2_{1,j},$$

where $(\chi^2_{1,1}, \cdots, \chi^2_{1,K_{\mathrm{z}}})$ *are i.i.d* χ^2_1*-distributed random variables , and* $(d_{1,t-1}, \cdots, d_{K_{\mathrm{z}},t-1})$ *are the* K_{z} *eigenvalues of the matrix*

$$\left(\mathbf{L}^{-1}_{\mathrm{z},t-1}\mathbf{O}_{t-1}\mathbf{L}^{-1\prime}_{\mathrm{z},t-1}\right)^{\frac{1}{2}} \frac{\boldsymbol{\Sigma}_{\mathrm{Z}\hat{\mathrm{X}}\mathrm{Z},t-1}}{\sigma_{\tilde{\mathrm{R}},t}} \left(\mathbf{L}^{-1}_{\mathrm{z},t-1}\mathbf{O}_{t-1}\mathbf{L}^{-1\prime}_{\mathrm{z},t-1}\right)^{\frac{1}{2}},$$

where $\mathbf{L}_{\mathrm{z},t-1} \equiv \left[\mathbf{0}_{K_{\mathrm{z}}\times(K_{\mathrm{f}}+1)}, \mathbf{I}_{K_{\mathrm{z}}}\right]\mathbf{L}_{t-1}$, *with* $\mathbf{L}_{t-1}$ *and* $\mathbf{O}_{t-1}$ *defined in* (OA.35), *and where* $N^{-1}\mathbf{R}_t'\mathbb{M}_{1_N}\mathbf{R}_t \to_p \sigma_{\tilde{\mathrm{R}},t} > 0$, *while* $N^{-1}\mathbf{Z}'_{t-1}\mathbb{M}_{\hat{\mathbf{X}}}\mathbf{Z}_{t-1} \to_p \boldsymbol{\Sigma}_{\mathrm{Z}\hat{\mathrm{X}}\mathrm{Z},t-1}$.

(ii) *When* $\boldsymbol{\gamma}_{t-1,z} \neq \mathbf{0}_{K_z}$,

$$\mathcal{T}_{\mathrm{z},t-1} \to_p \infty.$$

Moreover, under the additional Assumption 14, together with $\kappa_4 = 0$, *for any* $0 < \alpha < 1$,

$$\Pr\left(\hat{R}^2_{\mathrm{z},t-1} - z_{\alpha/2}\left(\frac{\hat{\omega}_{\mathrm{z},t-1}}{N}\right)^{\frac{1}{2}} \leq R^2_{\mathrm{z},t-1} \leq \hat{R}^2_{\mathrm{z},t-1} + z_{\alpha/2}\left(\frac{\hat{\omega}_{\mathrm{z},t-1}}{N}\right)^{\frac{1}{2}}\right) \to (1-\alpha).$$

where $\hat{\omega}_{\mathrm{z},t-1}$ *is defined in (OA.53) and represents a consistent estimator of the asymptotic covariance matrix of* $\hat{R}^2_{\mathrm{z},t-1}$, $z_{\alpha/2}$ *denotes the* $\alpha/2$*-th quantile of the standard normal distribution, and* $R^2_{t-1,z}$ *denotes the limit (in probability) of* $\hat{R}^2_{t-1,z}$.

Proof. See Appendix OA.4.

The result of Theorem 8 resembles the limiting behavior of the Hansen and Jagannathan (2007) (HJ) distance, which is typically used to test the null hypothesis of a correctly specified stochastic discount factor (SDF), against the alternative of misspecified models. Indeed, under the null hypotheses of correct model specification and no anomalies, respectively, both the HJ and the $\hat{R}^2_{\mathrm{z},t-1}$ statistics show a non-standard limiting distribution, consisting of a linear combination of i.i.d chi-squares, each of them having one degree of freedom. In contrast, the conventional Normal distribution is restored for both test statistics when considering their alternative hypotheses of either model misspecification (for the HJ statistic) or priced anomalies (in the case of $\hat{R}^2_{\mathrm{z},t-1}$ statistics).

Practically, Theorem 8 suggests the following empirical testing procedure to assess and quantify the effect of anomalies. At first, one would test whether the contribution of the considered anomaly variables is null or not, using the limiting results of part (i), and eliminate such variables from the asset pricing model whenever they would not provide any statistically significant contribution to the cross-section of expected returns. Alternatively, if the test results to be statically significant - that is the candidate anomalies play a significant role in explaining the cross-section of expected returns - then one could construct a valid confidence interval for $\hat{R}^2_{\mathrm{z},t-1}$ using the results of part (ii).

Finally, we can show that analogous properties hold in the case of a cross-sectional R-squared test that uses the local average premia estimator (46), the WLS estimator in (54) and the misspecification-robust estimator defined in (65).[44]

10 Empirical Application

10.1 Data

For our empirical exercise, we use data provided by Chen and Zimmermann (2019)[45], which contains 202 predictive firm-level characteristics at the monthly frequency. The reference period is January 1986 - December 2020. For the anomalies which are not available for the entire time period, we consider the last available month. Since our theory is derived for large N, in our analysis we consider only predictors for which we have enough test assets (i.e. at least 20 observations) in any given time interval. This leaves us with 170 variables, which we group following the ex ante categorization of Hou et al. (2020) in six economic categories, namely *Momentum* (15 variables), *Value versus Growth* (29 variables), *Investment* (30 variables), *Profitability* (20 variables), *Intangibles* (49 variables), and *Trading Frictions* (27 variables). A detailed list of the variables is shown in Table A.1. Following Hou et al. (2020), when performing monthly cross-sectional regressions, we winsorize the regressors at the 1% - 99% levels each month to mitigate the impact of outliers. We then standardize each regressor by subtracting its cross-sectional mean and dividing by its cross-sectional standard deviation. Monthly returns are from the Center for Research in Security

[44]Details are available upon request.

[45]https://www.openassetpricing.com/data/. Details on the construction of return predictors can be found in their Online Appendix https://drive.google.com/file/d/1vXRzjxYucXZV-tgLxM26fvRZ5zKvlBXH/view

Prices (CRSP), while the monthly Fama-French factors are downloaded from the Kenneth French website.

10.2 Preliminary Empirical Analysis: Model Identification

Before presenting our main empirical results, we first empirically evaluate the restriction in (13) to ensure the accurate identification of the anomalies' effect. To this end, we initially estimate the parameters $\gamma_{\mathrm{z},t-1}$ and $\delta_{\mathrm{f},t-1}$ from the regression

$$\mathbf{R}_t = \hat{\mathbf{X}}\mathbf{\Gamma}_{\mathrm{f},t-1} + \mathbf{Z}_{t-1}\boldsymbol{\gamma}_{\mathrm{z},t-1} + \boldsymbol{\eta}_t, \quad t = 2, .., T-1$$

using our time-varying estimator in (32). In each regression, we consider one anomaly at the time (i.e., $K_{\mathrm{z}} = 1$), while the estimated betas in $\hat{\mathbf{X}}$ are obtained using (26), in which the matrix $\mathbf{F}$ contains the Fama-French five-factors (i.e., $K_{\mathrm{f}} = 5$). Then, for each anomaly, we regress the estimated time-series of $\hat{\boldsymbol{\gamma}}_{\mathrm{z}}$ against the series of $\hat{\boldsymbol{\delta}}_{\mathrm{f}}$, and report the corresponding R-squared, together with its confidence interval. The results are displayed in Figure 3. If the identification restriction is satisfied, we should expect a near-perfect correlation in the regression of $\hat{\boldsymbol{\gamma}}_{\mathrm{z}}$ on $\hat{\boldsymbol{\delta}}_{\mathrm{f}}$, resulting in an R-squared value close to 1. From Figure 3, the distribution of the R-squared across anomalies reveals an average value of 0.41 (purple dotted line), with the 5th and 95th percentiles ranging between 0.07 and 0.77, respectively (blue dotted lines). Moreover, the confidence intervals reliably exclude the value of 1, suggesting that our model specification should not suffer from any identification issue.

FIGURE 3 HERE

This establishes the foundation for our empirical analysis, which we will conduct in the following sections.

10.3 Preliminary Empirical Analysis: Time-Variation and Cross-Correlations

In this Section, we would like to provide some empirical insights on the shortcomings of the conventional approach in estimating time-varying anomaly premia. As explained in Section 4, the conventional methodology requires orthogonality between the factor betas and the anomalies, as well as a constant nature of premia over time. However, in the following we present some compelling evidence that suggests these two conditions frequently contradict the reality.

In the first exercise, we use our data to calculate and test for the linear correlation between each anomaly and the betas estimated from a given asset pricing factor model. Specifically, to get a (statistically) good estimate of the betas, we first estimate a first-pass regression with $T = 120$ months, using the CAPM specification. Then, at each point in time, we measure the correlation between the estimated matrix $\mathbf{B}$ and each $\mathbf{Z}_t$, using the R-squared of the regression of $\mathbf{Z}_t$ on $\mathbf{B}$. This would produce a time-series of correlation coefficients (one for each anomaly), for which we can compute the corresponding time-series of t-statistic, under the null hypothesis of zero relationship between the two variables. We then repeat the same exercise using the Fama-French 3-factor model (FF3) and the Fama-French 5-factor model (FF5) as alternative model specifications, and save the R-squared of the regression of each $\mathbf{Z}_t$ on the estimated matrix $\mathbf{B}$ at every point in time. We finally take the average of the time-series results and aggregate them at the category level. The main results are summarized in Table II below. The first column of the table shows the average R-squared (i.e., the square of the correlation coefficient), together with its minimum and maximum value (in parenthesis), for each of the six categories and for each of the three model specifications (CAPM in Panel A, FF3 in Panel B and FF5 in panel C). In the second column we report, instead, the average percentage of times in which the correlation coefficient has been found to be statistically different from zero.[46] From the table, a clear fact emerges: while the correlation between betas and anomalies may often be very small in magnitude, in most of the cases it appears to be statistically significant. Therefore, caution is needed before applying standard methodology, since ignoring this correlation could considerably distort the inferential results of the conventional two-pass estimator, as we have shown in Section4 (See Figure 2).

TABLE II HERE

In the next part of the section, our objective is to evaluate whether there is any supporting (or contradicting) evidence regarding the time-varying nature of premia parameters. To achieve this, we employ multiple exercises. In the first exercise, we use our data to derive the time series of the $\gamma_{\mathrm{z},t-1}$ estimates in (14), considering one anomaly at the time in the regression of $\mathbf{R}_t$ on $\mathbf{Z}_{t-1}$. Subsequently, on each time-series, we fit an ARMA(p,q) model, where p and q range from 0 to 6.

[46]To assess whether the correlation coefficient is statistically different from zero, we use the standard Pearson correlation test in the CAPM specification, while we use the p-value of the F-test for the FF3 and FF5 models. We report the results using a 5% confidence level.

The objective is to evaluate the performance of a purely random white noise process (i.e., when both p and q are set equal to zero) over different time-varying processes. We determine the best fitting model by evaluating the AIC (Akaike Information Criterion) value. The main results are summarized in Table III (see Column 1), where we present the percentage of anomalies (within each category) for which we find a statistically superior fitting ability of a time-varying model over a white noise process. In Colums 2–4, we repeat the same exercise, where now the time series of $\gamma_{z,t-1}$ is obtained by regressing the asset returns $\mathbf{R}_t$ on each of the anomaly $\mathbf{Z}_{t-1}$ at the time, but including also the betas estimated from the CAPM (Column 2), the FF3 (Column 3) and FF5 (Column 4) in the regression specification. The table clearly shows a significant time-variation in the dynamics of the premia parameters in each of the six categories and regardless of the chosen model, with percentages that are greater than 80% in all the cases.

TABLE III HERE

To further investigate our findings, we use the premia estimates obtained in the above exercise to test for the presence of potential structural changes in their time series. Specifically, for each anomaly, we define $\Delta\hat{\gamma}_{z,t} = \delta_1 + \delta_2\hat{\gamma}_{z,t} + u_t$ and test for structural changes by analyzing the cumulative sums of OLS-residuals (CUSUM) from the estimated linear model (see Ploberger, Krämer, and Kontrus (1989)).[47] Under the null hypothesis of no structural changes, we then report in Table IV the percentage of anomalies within each category for which the test rejects the null hypothesis. Overall, in each category, we find a non-ignorable set of variables that show a significant break in the parameter estimates, with the highest percentage in the Investment and Momentum categories.[48]

TABLE IV HERE

As a final exercise, we use again the time series of the premia estimates to test whether there is any statistically significant shift in the mean between any two sub-samples of periods. In the case of no time-variation, one should not find any significant mean difference when using the paired t-test on the differences between two sub-samples. Therefore, using our $\gamma_{z,t}$ estimates (one for each anomaly at the time), we consider all the possible pairs of sub-samples (not necessarily

[47] As a further robustness check, instead of using cumulative sums of the same residuals, we also consider the moving sums (MOSUM) of residuals (Chu, Hornik, and Kaun (1995)) and the moving estimates (ME) process Chu, Hornik, and Kuan (1995). The results lead to approximately the same conclusion as the COSUM case. Details are available upon request. See also Zeileis, Leisch, Hornik, and Kleiber (2002) for the implementation.

[48] A more detailed analysis specific for each anomaly is reported in the Online Appendix.

consecutive) of length equal to either three years or five years, and calculate the difference in the estimates between the corresponding values in each pair. We then calculate the mean difference between each pair of sub-samples and derive the corresponding t-statistic under the null hypothesis that there is no shift in the mean of the two sub-periods (i.e., the mean difference is equal to zero). Table V reports the average percentage of paired sub-periods (aggregated at the category level) for which we find a statistically significant (at 10% confidence level) mean difference. Panel A of the table refers to the case of sub-periods of length three years, while Panel B refers to sub-periods of five years. Column 1 refers to the time series of $\gamma_{\mathrm{z},t-1}$ which is obtained by regressing the asset returns $\mathbf{R}_t$ on each of the anomaly $\mathbf{Z}_{t-1}$ at the time. Colums 2–4 use instead the time series of $\gamma_{\mathrm{z},t-1}$ obtained by regressing the asset returns $\mathbf{R}_t$ on each of the anomaly $\mathbf{Z}_{t-1}$ at the time, but including also the betas estimated from the CAPM (Column 2), the FF3 (Column 3) and FF5 (Column 4) in the regression specification. Once again, in many cases and in all the six categories, we find evidence of a significant shift in the mean of the parameters' distribution.

TABLE V HERE

In conclusion, all these results raise some doubts about the reliability of the main assumptions required by the traditional approach and provide the ground for our novel methodology.

10.4 Dissecting Time-Varying Anomalies

In this section we apply our theoretical results derived in Section 5 to our data. Our primary objective is to investigate the potential time-varying effect of asset-specific characteristics over time and to assess their contribution in explaining the cross-section of stock returns.

In the following analysis, we use balanced panels over fixed-time windows of two years (i.e., $T = 24$). Specifically, at each point in time t, we first obtain the matrix $\hat{\mathbf{B}}$ of factor betas using (26), employing the past two years of data, up to time t. We then run a cross-sectional regression, using the feasible representation in (30), in which all the anomalies in $\mathbf{Z_{t-1}}$ have been previously orthogonalized to the model risk factors, using (28). Finally, we obtain the time-varying premia estimates using our new estimator in (32), together with their corresponding standard errors derived in Theorem 2. We then shift the time window month by month over the 1986-2020 period, and obtain the rolling time series of the premia estimates and their corresponding t-statistics.

We first start the analysis by considering univariate regressions of asset returns on market beta and one anomaly at the time (Section 10.4.1). We then consider the case of multivariare regressions in Section 10.4.2, where we use more than one anomaly in each regression. Other robustness checks will be presented in the Online Appendix.

10.4.1 Univariate Analysis

To gain a preliminary insight into the impact of the anomalies, in this section we consider simple univariate regressions, focusing on a single anomaly at a time and using the market factor betas estimated through the CAPM specification.[49] The t-statistics associated with each anomaly premium at each month in the sample is reported in the heatmap in Figure 4. The figure shows the heatmap of the t-statistics distribution obtained for each univariate regression (vertical axis) and for each month (horizontal axis). Each cell in the map represents the degree of statistical significance of the t-statistics with a different color, from gray (non-significant t-stat), to yellow (significance at only 10% level), orange (significance at 5% level), and red (significance at 1% level).

FIGURE 4 HERE

[49] A similar exercise has been also conducted using the betas estimated through the FF3 and FF5 factor models. Results are provided in the online Appendix.

While no clear pattern emerges within the six categories, a distinct structure becomes apparent when examining the distribution of t-statistics over time. Notably, the majority of the red data points in Figure 4 cluster within specific time intervals. What is even more interesting is that these periods of high-significance concentration appear to coincide with major financial crises. For example, they align with significant events such as the early 1990-91 recession, the dot-com bubble from 1999 to 2000, the financial crisis of 2007-2009, and the recent stock market crash in early 2020 caused by the outbreak of the COVID-19 pandemic.

This result is further corroborated by Figure 5, in which we depict the percentage of anomalies found to be statistically significant at the 5% confidence level, across the various points in time. The light gray bands on the graph correspond to NBER recession dates as well as different economic and financial crises. Notably, the figure illustrates that higher percentages of statistically significant anomalies frequently coincide with periods of heightened market uncertainty, with a peak exceeding 70% during the 2007-2009 financial crisis. To statistically reinforce this finding, we also conducted an OLS regression of the percentage of statistically significant anomalies against a time dummy variable, which is set to one if the period t aligns with a crisis period and zero otherwise. Our analysis revealed a substantial and positive slope coefficient of 4.29, with a corresponding t-statistic of 2.75.

FIGURE 5 HERE

In summary, this simple analysis offers evidence of substantial time-variation in regression estimates. However, such time-varying signal would be overlooked if we solely concentrated on *average* estimators, as is conventionally practiced in the literature.

All the results presented above have been derived from simple univariate regressions. However, given that univariate regressions are seldom employed in empirical applications, in the next section, we apply our time-varying methodology through multiple regressions. This will involve a careful selection of a dynamic set of representative anomalies at each time point, as we explain in the next section.

10.4.2 Multivariate Analysis

In this section, we apply our time-varying methodology using multiple cross-sectional regressions. Dealing with hundreds of firm characteristics can lead to several technical issues. First, using all the characteristics simultaneously will dramatically reduce the number of assets, as it is quite challenging to find a balanced panel in which all assets have all the characteristics available for each time interval. Second, many of the characteristics are likely to be highly correlated with each other, especially those within the same category, which can lead to multicollinearity problems in the regression models. To circumvent these issues and to identify the "best" representative set of variables, we select, at each point in time, the six anomalies (one from each category) that have produced the highest R_z^2 values in the univariate regressions from the previous section. The time-varying set of anomalies is graphically depicted in Figure 6, where each red point represents the variable selected in each category (vertical axis) for each month t (horizontal axis).[50]

FIGURE 6 HERE

To assess the predictive ability of the selected models, we present the time series of the R_z^2 statistics in Figure 7a, alongside the total variance decomposition depicted in Figure 7b, obtained from each multivariate regression. As observed in Figure 7a, the portion of total cross-sectional variation in asset returns jointly explained by these anomalies exhibits substantial temporal variability, spanning from a minimum of 0.7% to a maximum of 46%. In line with the results of the previous section, higher R_z^2 values tend to coincide with periods of economic or financial crises, denoted by the gray regions in the figure. This result reaffirms the concept that anomalies exert a more pronounced influence during times of heightened uncertainty.

FIGURE 7 HERE

The market beta also appears to hold significance (as indicated by the green bars in Figure 7b), contributing on average to almost 8% of the total variance, with occasional peaks reaching up to 40%. Unlike anomalies, the highest contribution of the market beta does not appear to be linked

[50] In some cases, while extracting the balanced panel of asset returns along with the chosen six anomalies, there might be instances where only a very limited number of observations remains available for analysis. To mitigate this concern, in such scenarios, we explore alternative combinations of regressors for which we have a sufficiently large sample size (N ¿ 100) and opt for the combination that yields the highest (in-sample) R_z^2.

to periods of crises. On average, we find that when considering anomalies and betas together, they can explain over 20% of the total cross-sectional variation in asset returns. Of this average 20%, anomalies contribute 60%, while betas account for the remaining 40%.

Finally, we evaluate whether the joint contribution of anomalies to the overall R-squared of each model is statistically significant or not. Our null hypothesis is defined as $H_0 : \boldsymbol{\gamma}_z = \mathbf{0}_{K_z}$, against the alternative that at least one anomaly is different from zero, i.e., $H_1 : \boldsymbol{\gamma}_z \neq \mathbf{0}_{K_z}$, with $K_z = 6$. To test this hypothesis, we make use of the limiting results of Theorem 8, part (i), wherein we tabulate the asymptotic distribution of the statistics $\mathcal{T}_z$ under H_0 using 10,000 random draws from six i.i.d. χ^2_1, weighted with the estimated values $(\hat{c}_1, ..., \hat{c}_6)$ obtained in each model. The resulting time series of p-values associated with the $\mathcal{T}_z$ statistics for each model at every point in time is presented in Figure 8. The yellow bands in the figure correspond to p-values representing periods in which there is no compelling evidence to reject the null hypothesis ($p > 0.05$). On the other hand, the blue regions indicate p-values ≤ 0.05, identifying periods where we can confidently reject the null hypothesis, implying a non-zero contribution from anomalies to the total variability of the model. Our findings indicate that the joint contribution of anomalies to the total R-squared of the model is statistically distinguishable from zero in approximately 40% of the cases. Even though this percentage might seem quite low, it is noteworthy that in 71% of these instances, a significant contribution of anomalies aligns again with periods of financial downturns.

We would like to conclude this section by acknowledging that, although our analysis allows for a very general form of time-variation, it comes with certain inherent limitations. In fact, it could suffer from high levels of idiosyncratic risk, which, in turn, can influence the time-varying R-squared of our regressions. Naturally, averaging data over a more extended time period would yield significantly higher R-squared values, albeit at the expense of obscuring most of the time-varying patterns.

11 Conclusion

We extend the two-pass methodology for estimating and testing the effect of anomalies in asset pricing models with time-varying premia. Our methodology is designed for when large cross-sections of N assets are available but the number of time-series observations T is fixed and possibly very small, but applies also when N and T are both very large. We develop the method for ordinary and weighted least-squares estimation, and consider both cases of correct specification and global misspecification of the candidate asset pricing model. Inference relies on asymptotically valid standard errors for the premia estimators, derived in closed-form. A cross-sectional R-squared test to dissect anomalies is proposed, establishing its limiting properties under the null hypothesis of no effect of anomalies and its alternative. Using a dataset of $20,000$ individual US stock returns, we find that although anomalies are statistically significant in about half the cases (out of 170 anomalies), they explain a small fraction (less than 10%) of the cross-sectional variation of expected returns. Anomalies tend to be more important during economic and financial crises.

References

Acharya, V. V., and L. H. Pedersen (2005): "Asset pricing with liquidity risk," *Journal of Financial Economics*, 77, 375–410.

Ahn, S. C., A. R. Horenstein, and N. Wang (2018): "Beta matrix and common factors in stock returns," *Journal of Financial and Quantitative Analysis*, pp. 1–24, forthcoming.

Anatolyev, S., and A. Mikusheva (2020): "Factor models with many assets: Strong factors, weak factors, and the two-pass procedure," *Working Paper, MIT*.

Ang, A., R. Hodrick, Y. Xing, and X. Zhang (2006): "The cross section of volatility and expected return," *Journal of Finance*, 11, 259–299.

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

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

Avramov, D., and T. Chordia (2006): "Asset pricing models and financial market anomalies," *The Review of Financial Studies*, 19, 1001–1040.

Back, K., N. Kapadia, and B. Ostdiek (2015): "Testing Factor Models on Characteristic and Covariance Pure Plays," Discussion paper, http://dx.doi.org/10.2139/ssrn.2621696.

Barras, L., O. Scaillet, and R. Wermers (2010): "False discoveries in mutual fund performance: Measuring luck in estimated alphas," *The Journal of Finance*, 65, 179–216.

Brennan, M., T. Chordia, and A. Subrahmanyam (1998): "Alternative factor specifications, security characteristics, and the cross-section of expected stock returns," *Journal of Financial Economics*, 49, 345–373.

Brennan, M., T. Chordia, A. Subrahmanyam, and Q. Tong (2012): "Sell-order liquidity and the cross-section of expected stock returns," *Journal of Financial Economics*, 105, 523–541.

Bryzgalova, S. (2014): "Spurious factors in linear asset pricing models," *Working Paper, Stanford University*.

Bryzgalova, S., S. Lerner, M. Lettau, and M. Pelger (2022): "Missing Financial Data," *Available at SSRN 4106794*.

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

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

Cattaneo, M., R. Crump, M. Farrell, and S. Schaumburg (2020): “Characteristic-Sorted Portfolios: Estimation and Inference,” *Review of Economics and Statistics*, 102, 531–551.

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

Chen, A. (2021): “The Limits of p-hacking: Some thought experiments,” *Journal of Finance*, 76, 2447–2480.

Chen, A. Y., and T. Zimmermann (2019): “Publication Bias and the Cross-Section of Stock Returns,” *The Review of Asset Pricing Studies*, 10, 249–289.

Chordia, T., A. Goyal, and J. Shanken (2015): “Cross-sectional asset pricing with individual stocks: Betas versus characteristics,” *Working Paper, Emory University.*

Chu, C.-S. J., K. Hornik, and C.-M. Kaun (1995): “MOSUM tests for parameter constancy,” *Biometrika*, 82(3), 603–617.

Chu, C.-S. J., K. Hornik, and C.-M. Kuan (1995): “The moving-estimates test for parameter stability,” *Econometric Theory*, 11(4), 699–720.

Connor, G., and R. Korajczyk (2022): “Semi-strong factors in asset returns,” *Journal of Financial Econometrics*, Forthcoming.

Connor, G., and O. Linton (2007): “Semiparametric estimation of a characteristic-based factor model of common stock returns,” *Journal of Empirical Finance*, 14, 694 – 717.

Conrad, J., M. Cooper, and G. Kaul (2003): “Value versus glamour,” *The Journal of Finance*, 58, 1969–1995.

Daniel, K., L. Mota, S. Rottke, and T. Santos (2020): “The Cross-Section of Risk and Returns,” *The Review of Financial Studies*, 33, 1927–1979.

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

DeMiguel, V., A. Martin-Utrera, F. J. Nogales, and R. Uppal (2020): "A transaction-cost perspective on the multitude of firm characteristics," *The Review of Financial Studies, forthcoming*.

Dong, X., L. Yan, D. E. Rapach, and G. Zhou (2021): "Anomalies and the expected market return," *Journal of Finance, forthcoming*.

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

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

——— (2008): "Dissecting anomalies," *The Journal of Finance*, 63, 1653–1678.

——— (2015): "A five-factor asset pricing model," *Journal of Financial Economics*, 116, 1–22.

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

Fama, G., and K. French (2020): "Comparing Cross-Section and Time-Series Factor Models," *The Review of Financial Studies*, 33, 1891–1926.

Fan, J., Y. Liao, and W. Wang (2016): "Projected principal component analysis in factor models," *The Annals of Statistics*, 44, 219–254.

Ferson, W. E., and C. R. Harvey (1999): "Conditioning variables and cross-section of stock returns," *Journal of Finance*, 54, 1325–1360.

Freyberger, J., B. Höppner, A. Neuhierl, and M. Weber (2021): "Missing data in asset pricing panels," *Available at SSRN*.

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

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

Giglio, S., D. Xiu, and D. Zhang (2021): "Test assets and weak factors," Discussion paper, National Bureau of Economic Research.

Gospodinov, N., R. Kan, and C. Robotti (2014): "Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors," *The Review of Financial Studies*, 27, 2139–2170.

Green, J., J. R. M. Hand, and X. F. Zhang (2017): "The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns," *The Review of Financial Studies*, 30, 4389–4436.

Hansen, L. P., and R. Jagannathan (2007): "Assessing specification errors in stochastic discount factor models," *The Journal of Finance*, 52, 557–590.

Hansen, L. P., and S. F. Richard (1987): "The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models," *Econometrica*, 55, 587–613.

Harvey, C. R., Y. Liu, and H. Zhu (2016): "... and the cross-section of expected returns," *The Review of Financial Studies*, 29, 5–68.

Hou, K., X. Chen, and L. Zhang (2015): "Digesting anomalies: An investment approach," *The Review of Financial Studies*, 28, 650–705.

——— (2020): "Replicating anomalies," *The Review of Financial Studies*, 12, 2019–2133.

Hou, K., and R. Kimmel (2006): "On the estimation of risk premia in linear factor models," *Working Paper, Ohio State University.*

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

Jegadeesh, N., and S. Titman (1993): "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *The Journal of Finance*, 48(1), 65–91.

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

Kan, R., and C. Zhang (1999a): "GMM tests of stochastic discount factor models with useless factors," *Journal of Financial Economics*, 54, 103–127.

——— (1999b): "Two-pass tests of asset pricing models with useless factors," *The Journal of Finance*, 54, 203–235.

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

Kim, S., R. A. Korajczyk, and A. Neuhierl (2021): "Arbitrage Portfolios," *The Review of Financial Studies*, 34, 2813–2856.

Kim, S., V. Raponi, and P. Zaffaroni (2020): "Testing for Spurious Factors in Conditional Asset Pricing," *Working Paper, Imperial College London.*

——— (2023): "Testing Weak Factors in Asset Pricing," Discussion paper, KAIST, IESE, and Imperial College London.

Kleibergen, F. (2009): "Tests of risk premia in linear factor models," *Journal of Econometrics*, 149, 149–173.

Kleibergen, F., and Z. Zhan (2014): "Mimicking portfolios of macroeconomic factors," *Working Paper, Aarhus University.*

——— (2020): "Robust inference for consumption-based asset pricing," *The Journal of Finance*, 75, 507–550.

Knapp, T. R. (1978): "Canonical correlation analysis: A general parametric significance-testing system," *Psychological Bulletin*, 85, 410–416.

Koijen, R. S. J., T. J. Moskowitz, L. H. Pedersen, and E. B. Vrugt (2018): "Carry," *Journal of Financial Economics*, 27, 197–225.

Kozak, S., and S. Nagel (2023): "When Do Cross-Sectional Asset Pricing Factors Span the Stochastic Discount Factor?," Discussion paper, http://dx.doi.org/10.2139/ssrn.4443643.

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

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

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

Lewellen, J. (2015): "The Cross-section of Expected Stock Returns," *Critical Finance Review*, 4.

Lintner, J. (1965): "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *The Review of Economics and Statistics*, 47, 13–37.

Lo, A. W., and A. C. MacKinlay (1990): "Data-snooping biases in tests of financial asset pricing models," *The Review of Financial Studies*, 3, 431–467.

McLean, R., and J. Pontiff (2016): "Does academic research destroy stock return predictability?," *The Journal of Finance*, 71, 5–32.

Pástor, L., and R. F. Stambaugh (2003): "Liquidity risk and expected stock returns," *Journal of Political Economy*, 111, 642–685.

Ploberger, W., W. Krämer, and K. Kontrus (1989): "A new test for structural stability in the linear regression model," *Journal of Econometrics*, 40(2), 307–318.

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

Raponi, V., and P. Zaffaroni (2023): "Dissecting Anomalies in Conditional Asset Pricing - Online Appendix," Discussion paper, Imperial College London.

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

Shanken, J. (1990): "Intertemporal Asset Pricing: An Empirical Investigation," *Journal of Econometrics*, 45, 99–120.

Shanken, J. (1992): "On the estimation of beta-pricing models," *The Review of Financial Studies*, 5, 1–33.

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

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

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

Zaffaroni, P. (2022): "Factor Models for Conditional Asset Pricing," Discussion paper, Imperial College London.

Zeileis, A., F. Leisch, K. Hornik, and C. Kleiber (2002): "strucchange: An R Package for Testing for Structural Change in Linear Regression Models," *Journal of Statistical Software*, 7(2), 1–38.

Appendix

A.1 Assumptions

In this section, we present the main assumptions required for the validity of our large-N asymptotic theory, without further comments (see Section OA.2 for detailed comments). All the moments below are assumed to hold conditionally on the factors $\mathbf{F}$, even if not written explicitly, and all the limits below hold as $N \to \infty$.

It is useful to recall the $N \times K_{\mathrm{z}}(T-1)$ matrix of anomalies $\mathbf{Z} \equiv (\mathbf{z}_1, ..., \mathbf{z}_N)'$, where $\mathbf{z}_i$ defines the $K_{\mathrm{z}}(T-1) \times 1$ vector $\mathbf{z}_i \equiv \left(z_{i,1}^{(1)}, \cdots, z_{i,T-1}^{(1)}, \cdots, z_{i,1}^{(K_{\mathrm{z}})}, \cdots, z_{i,T-1}^{(K_{\mathrm{z}})}\right)'$. The $N \times K_{\mathrm{z}}$ matrix of anomalies at time $t-1$ is defined as $\mathbf{Z}_{t-1} = (\mathbf{z}_{1,t-1}, \cdots, \mathbf{z}_{N,t-1})'$, while the $(T-1) \times K_{\mathrm{z}}$ matrix of anomalies specific for the i-th asset is $\mathbf{Z}_i = (\mathbf{z}_{i,1}, \cdots, \mathbf{z}_{i,T-1})'$, setting $\mathbf{z}_{i,t-1} = \left(z_{i,t-1}^{(1)}, \cdots, z_{i,t-1}^{(K_{\mathrm{z}})}\right)'$.

Assumption 3 (*risk factors and anomalies*)**.** Set $\tilde{\mathbf{Z}}_i \equiv \mathbb{M}_{1_{T-1}} \mathbf{Z}_i$, and $\mathbf{D} \equiv (\mathbf{1}_{T-1}, \mathbf{F})$. Then, for every T, the $(T-1) \times (K+1)$ matrix $\tilde{\mathbf{D}}_i = (\mathbf{D}, \tilde{\mathbf{Z}}_i)$ satisfies

$$\tilde{\mathbf{D}}_i' \tilde{\mathbf{D}}_i > 0 \quad \text{for every } i = 1, ..., N.$$

Assumption 4 (*loadings*)**.**

$$\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\beta}_i \to \boldsymbol{\mu}_\beta \qquad \text{and} \qquad \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\beta}_i \boldsymbol{\beta}_i' \to \boldsymbol{\Sigma}_\beta,$$

such that the matrix

$$\boldsymbol{\Sigma}_{\mathrm{X}} \equiv \begin{bmatrix} 1 & \boldsymbol{\mu}_\beta' \\ \boldsymbol{\mu}_\beta & \boldsymbol{\Sigma}_\beta \end{bmatrix} > 0.$$

Remark 1. *Assumption 4 states that the limiting cross-sectional averages of the betas, and of the squared betas, exist. The second part of Assumption 4 rules out the possibility of spurious, weak, and semi-strong factors, which are known to be problematic, in terms of asymptotic behaviour, for the two-pass estimator (see Anatolyev and Mikusheva (2020) for the large N, T case, and Kim, Raponi, and Zaffaroni (2023) for the large N case, akin to the sampling scheme considered in this paper).*

and situations in which at least one of the elements of $\boldsymbol{\beta}_i$ is cross-sectionally constant. It implies that $\mathbf{X} = (\mathbf{1}_N, \mathbf{B})$ has full (column) rank for N sufficiently large. To simplify the exposition, we *assume that the $\boldsymbol{\beta}_i$ are non-random.*[51]

Assumption 5 (*asset-specific components*)**.** The $N \times 1$ vector of error terms $\boldsymbol{\epsilon}_t$ is independently and identically distributed (i.i.d.) over time with

$$\mathrm{E}[\boldsymbol{\epsilon}_t] = \mathbf{0}_N \tag{A.1}$$

and with the $N \times N$ variance-covariance matrix satisfying

$$\mathrm{Var}\left[\boldsymbol{\epsilon}_t\right] = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_N^2 \end{pmatrix} \equiv \boldsymbol{\Sigma} > 0, \tag{A.2}$$

where σ_{ij} denotes the (i,j)-th element of $\boldsymbol{\Sigma}$, for every $i, j = 1, \ldots, N$, and with $\sigma_i^2 \equiv \sigma_{ii}$.

Assumption 6 (*cross-sectional moments of asset-specific components*)**.** (i)

$$\frac{1}{N}\sum_{i=1}^{N}\left(\sigma_i^2 - \sigma^2\right) = o\left(\frac{1}{\sqrt{N}}\right), \tag{A.3}$$

for some $0 < \sigma^2 < \infty$.

(ii)

$$\sum_{i,j=1}^{N} \mid \sigma_{ij} \mid \mathbb{1}_{\{i \neq j\}} = o\left(N\right). \tag{A.4}$$

(iii)

$$\frac{1}{N}\sum_{i=1}^{N}\mu_{4i} \to \mu_4, \tag{A.5}$$

for some $0 < \mu_4 < \infty$, where $\mu_{4i} \equiv E[\epsilon_{i,t}^4]$.

(iv)

$$\frac{1}{N}\sum_{i=1}^{N}\sigma_i^4 \to \sigma_4, \tag{A.6}$$

for some $0 < \sigma_4 < \infty$.

(v)

$$\sup_i \mu_{4i} \leq C < \infty, \tag{A.7}$$

for a generic constant C.

[51] See Gagliardini, Ossola, and Scaillet (2016) and Raponi, Robotti, and Zaffaroni (2020) for the analysis of asset pricing models with random betas.

(vi)

$$E[\epsilon_{i,t}^3] = 0. \tag{A.8}$$

(vii)

$$\frac{1}{N}\sum_{i=1}^{N}\kappa_{4,iiii} \to \kappa_4, \tag{A.9}$$

for some $0 \leq |\kappa_4| < \infty$, where $\kappa_{4,iiii} \equiv \kappa_4[\epsilon_{it}, \epsilon_{it}, \epsilon_{it}, \epsilon_{it}]$ denotes the fourth-order cumulant of the asset-specific component $\{\epsilon_{i,t}, \epsilon_{i,t}, \epsilon_{i,t}, \epsilon_{i,t}\}$.

(viii) For every $3 \leq h \leq 8$, all the following mixed cumulants of order h satisfy

$$\sup_{i_1}\sum_{i_2,\ldots,i_h=1}^{N}|\kappa_{h,i_1 i_2 \ldots i_h}| = o\left(N\right), \tag{A.10}$$

for at least one i_j $(2 \leq j \leq h)$ different from i_1, where $\kappa_{h,i_1,i_2 \ldots i_h}$ is the mixed cumulant in the $\{\epsilon_{i_1,s}, \epsilon_{i_2,s}, \cdots, \epsilon_{i_h,s}\}$ of order h.

Assumption 7 (*CLT of asset-specific component*)**.** (i)

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\epsilon}_i \to_d \mathcal{N}\left(\mathbf{0}_{T-1}, \sigma^2\mathbf{I}_{T-1}\right). \tag{A.11}$$

(ii)

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathrm{vec}\left(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' - \sigma_i^2\mathbf{I}_{T-1}\right) \to_d \mathcal{N}\left(\mathbf{0}_{(T-1)^2}, \mathbf{U}_\epsilon\right), \tag{A.12}$$

where $\mathbf{U}_\epsilon \equiv \lim \frac{1}{N}\sum_{i,j=1}^{N} E\Big[\mathrm{vec}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' - \sigma_i^2\mathbf{I}_{T-1})\mathrm{vec}(\boldsymbol{\epsilon}_j\boldsymbol{\epsilon}_j' - \sigma_j^2\mathbf{I}_{T-1})'\Big]$.

(iii) For any $T \times 1$ vector $\mathbf{c}$,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{c}' \otimes \begin{pmatrix}1\\ \boldsymbol{\beta}_i\end{pmatrix}\right)\boldsymbol{\epsilon}_i \to_d \mathcal{N}\left(\mathbf{0}_{K_\mathrm{f}+1}, (\mathbf{c}'\mathbf{c})\sigma^2\boldsymbol{\Sigma}_\mathrm{X}\right). \tag{A.13}$$

Remark 2. *The expression for* $\mathbf{U}_\epsilon$ *in* (A.12) *can be derived in closed form. In particular, Raponi,* *Robotti, and Zaffaroni (2020) established that the $T^2 \times T^2$ matrix $\mathbf{U}_\epsilon$ has the following form*

$$\mathbf{U}_\epsilon = \left[\begin{array}{c:c:c:c:c} U_{11} & \cdots & U_{1t} & \cdots & U_{1T} \\ \hdashline \vdots & \ddots & \vdots & \vdots & \vdots \\ \hdashline U_{t1} & \cdots & U_{tt} & \cdots & U_{tT} \\ \hdashline \vdots & \vdots & \vdots & \ddots & \vdots \\ \hdashline U_{T1} & \cdots & U_{Tt} & \cdots & U_{TT} \end{array}\right].$$

Each block of U_ϵ is a $T \times T$ matrix. The blocks along the main diagonal, denoted by U_{tt}, $t = 1, 2, \ldots, T$, are themselves diagonal matrixes with $(\kappa_4 + 2\sigma_4)$ in the (t,t)-th position and σ_4 in the (s,s) position for every $s \neq t$. The blocks outside the main diagonal, denoted by U_{ts}, $s, t = 1, 2, \ldots, T$ with $s \neq t$, are all made of zeros except for the (s,t)-th position that contains σ_4; that is,

$$U_{tt} = \begin{array}{cc} & \begin{array}{c} \downarrow \\ t\text{-th column} \end{array} \\ \underset{\substack{t\text{-th} \\ \text{row}}}{\rightarrow} & \begin{bmatrix} \sigma_4 & \cdots & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \sigma_4 & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & (\kappa_4 + 2\sigma_4) & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & \sigma_4 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & \sigma_4 \end{bmatrix} \end{array}, \quad U_{ts} = \begin{array}{cc} & \begin{array}{c} \downarrow \\ t\text{-th column} \end{array} \\ \underset{\substack{s\text{-th} \\ \text{row}}}{\rightarrow} & \begin{bmatrix} 0 & \cdots & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & \sigma_4 & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 \end{bmatrix} \end{array}.$$

Assumption 8 (*moments and CLT of anomalies*)**.** *Define the $K_{\mathrm{z}}(T-1)^2 \times 1$ vector $\mathbf{u}_i \equiv \boldsymbol{\epsilon}_i \otimes \mathbf{z}_i$.*

(i)

$$\frac{\mathbf{Z}'\mathbf{1}_N}{N} \to_p (\boldsymbol{\mu}_{\mathrm{z}} \otimes \mathbf{1}_{T-1}) \equiv \boldsymbol{\mu}_{\mathrm{z,T-1}}$$

for a finite $K_{\mathrm{z}} \times 1$ vector $\boldsymbol{\mu}_{\mathrm{z}} = \left(\mu_{\mathrm{z}}^{(1)}, \ldots, \mu_{\mathrm{z}}^{(K_{\mathrm{z}})}\right)' \equiv \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\mu}_{\mathrm{z}_i}$, setting $\boldsymbol{\mu}_{\mathrm{z}_i} \equiv \mathrm{E}[\mathbf{z}_{i,s}]$.

(ii)

$$\frac{\mathbf{Z}'\mathbf{Z}}{N} \to_p \boldsymbol{\Sigma}_{\mathrm{Z}},$$

for a finite $K_z(T-1)\times K_z(T-1)$ matrix $\mathbf{\Sigma}_Z$, such that $\mathbb{J}'\mathbf{\Sigma}_Z\mathbb{J} > 0$ and $\mathbb{J}_{t-1}'\mathbf{\Sigma}_Z\mathbb{J}_{t-1} > 0$, for every $2 \leq t \leq T$.

(iii)

$$\frac{\mathbf{Z}'\mathbf{B}}{N} \to_p \mathbf{\Sigma}_{ZB},$$

for a finite $K_z(T-1)\times K_f$ matrix $\mathbf{\Sigma}_{ZB}$.

(iv) *Setting* $\boldsymbol{\mu}_{u_i} \equiv \mathrm{E}[\mathbf{u}_i]$,

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\mu}_{u_i} = o\left(\frac{1}{\sqrt{N}}\right).$$

(v) *Setting* $\mathbf{\Sigma}_{u,ij} \equiv \mathrm{Cov}[\mathbf{u}_i, \mathbf{u}_j]$, *for* $i, j = 1, ..., N$,

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Sigma}_{u,ii} \to \mathbf{\Sigma}_U \equiv (\sigma^2\mathbf{I}_{T-1}\otimes\mathbf{\Sigma}_Z) \text{ and } \sum_{i,j=1}^{N}\mathbf{\Sigma}_{u,ij}\mathbb{1}_{i\neq j} = o(N).$$

(vi) For any $i, j = 1, ..., N$,

$$\mathrm{Cov}\left[\mathbf{z}_{i,t}, \boldsymbol{\epsilon}_j'\otimes\boldsymbol{\epsilon}_j'\right] = \mathbf{0}_{K_z\times(T-1)^2}, \quad \mathrm{Cov}\left[\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j'\otimes(\boldsymbol{u}_j - \mathrm{E}[\boldsymbol{u}_j])'\right] = \mathbf{0}_{T-1\times K_z(T-1)^3}.$$

(vii)

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\boldsymbol{u}_i - \boldsymbol{\mu}_{u_i}) \to_d \mathcal{N}\left(\mathbf{0}_{K_z(T-1)^2}, \mathbf{\Sigma}_U\right).$$

(viii) *Setting* $\mathbf{\Sigma}_{u\epsilon,ij} \equiv \mathrm{Cov}\left[\boldsymbol{\epsilon}_i\otimes\boldsymbol{\epsilon}_i, \mathbf{u}_j'\right]$,

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Sigma}_{u\epsilon,ii} \quad\to\quad \mathbf{\Sigma}_{u\epsilon} = \mathbf{0}_{(T-1)^2\times K_z(T-1)^2} \text{ and } \frac{1}{N}\sum_{i,j=1}^{N}\mathbf{\Sigma}_{u\epsilon,ij} \to \mathbf{0}_{(T-1)^2\times K_z(T-1)^2}.$$

(ix)

$$\frac{1}{N^2}\sum_{i,j=1}^{N}\mathrm{Cov}\left[\boldsymbol{u}_i\otimes\boldsymbol{u}_i, \boldsymbol{u}_j'\otimes\boldsymbol{u}_j'\right] \to \mathbf{0}_{K_z^2(T-1)^4\times K_z^2(T-1)^4}.$$

(x) *Let* $\mathbb{P}_{\tilde{Z}_i} = \tilde{\mathbf{Z}}_i(\tilde{\mathbf{Z}}_i'\tilde{\mathbf{Z}}_i)^{-1}\tilde{\mathbf{Z}}_i'$, *with its generic* (t, s) *element denoted by* $\mathbb{p}_{i,ts}$, *for* $t, s = 1, ..., T-1$, *where* $\tilde{\mathbf{Z}}_i = \mathbb{M}_{1_{T-1}}\mathbf{Z}_i$. *Then, for every* $1 \leq t+1, s+1, v_a, u_a \leq (T-1)$, *with* $a = 1, ..., 4$, *the* *following hold:*

$$(x.1) \qquad \frac{1}{N}\sum_{i=1}^{N}\mathbb{P}_{\tilde{Z}_i} \to_p \mathbb{P}_{\tilde{Z}}, \text{ for a finite matrix } \mathbb{P}_{\tilde{Z}},$$

$$(x.2) \qquad \frac{1}{N}\sum_{i=1}^{N}(\mathbb{P}_{\tilde{Z}_i} \odot \mathbb{P}_{\tilde{Z}_i}) \to_p \mathbb{P}_{\tilde{Z}}^{(2)}, \text{ for a finite matrix } \mathbb{P}_{\tilde{Z}}^{(2)},$$

$$(x.3) \qquad \frac{1}{N}\sum_{i=1}^{N}\mathbb{P}_{\tilde{Z}_i}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' - \sigma_i^2\mathbf{I}_{T-1}) = \mathbb{P}_{\tilde{Z}}\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' - \sigma_i^2\mathbf{I}_{T-1}) + o_p\left(\frac{1}{\sqrt{N}}\right),$$

$$(x.4) \qquad \frac{1}{N^2}\sum_{i,j=1}^{N}\kappa_4\left[\prod_{a=1}^{4}\mathbb{p}_{i,t-1u_a}, \prod_{a=1}^{4}\mathbb{p}_{j,s-1v_a}, \prod_{a=1}^{4}\epsilon_{i,u_a+1}, \prod_{a=1}^{4}\epsilon_{j,v_a+1}\right] = o(1),$$

$$(x.5) \qquad \frac{1}{N^2}\sum_{i,j=1}^{N}\kappa_3\left[\prod_{a=1}^{4}\mathbb{p}_{i,t-1u_a}, \prod_{a=1}^{4}\mathbb{p}_{j,s-1v_a}, \prod_{a=1}^{4}\epsilon_{i,u_a+1}\right] = o(1),$$

$$(x.6) \qquad \frac{1}{N^2}\sum_{i,j=1}^{N}\kappa_3\left[\prod_{a=1}^{4}\mathbb{p}_{i,t-1u_a}, \prod_{a=1}^{4}\epsilon_{i,u_a+1}, \prod_{a=1}^{4}\epsilon_{j,v_a+1}\right] = o(1),$$

$$(x.7) \qquad \frac{1}{N^2}\sum_{i,j=1}^{N}\text{Cov}\left[\prod_{a=1}^{4}\mathbb{p}_{i,t-1u_a}, \prod_{a=1}^{4}\epsilon_{j,v_a+1}\right] = o(1),$$

$$(x.8) \qquad \frac{1}{N^2}\sum_{i,j=1}^{N}\text{Cov}\left[\mathbb{p}_{j,su_1}\mathbb{p}_{i,tv_1}, \epsilon_{i,t+1}\epsilon_{j,s+1}\epsilon_{iu_1+1}\epsilon_{jv_1+1}\right] = o(1),$$

$$(x.9) \qquad \frac{1}{N}\sum_{i=1}^{N}\text{Cov}\left[\prod_{a=1}^{4}\mathbb{p}_{i,t-1u_a}, \prod_{a=1}^{4}\epsilon_{i,v_a+1}\right] = o(1).$$

where $\kappa_3[\cdot,\cdot,\cdot]$ and $\kappa_4[\cdot,\cdot,\cdot,\cdot]$ denote the mixed cumulants of order 3 and 4, respectively.

(xi) For every $3 \leq h \leq 8$, all the following mixed cumulants of order h satisfy

$$\sup_{i_1}\sum_{i_2,\dots,i_h=1}^{N}|\kappa_{h,i_1 i_2\dots i_h}^{\mathbb{p}}| = o\left(N\right), \tag{A.14}$$

for at least one i_j $(2 \leq j \leq h)$ different from i_1, where $\kappa_{h,i_1,i_2\dots i_h}^{\mathbb{p}}$ is the mixed cumulant in the $\{\mathbb{p}_{i_1,t_1-1u_1}, \mathbb{p}_{i_2,t-_21u_2}, \cdots, \mathbb{p}_{i_h,t_h-1u_h},\}$ of order h, for every $2 \leq t_1, \cdots, t_h, u_1, \cdots, u_h \leq T$.

A.1.1 Additional assumptions required for the WLS estimation

In this Section, we introduce additional assumptions that are required for the validity of the WLS estimation described in Section 7. Before stating the main assumptions, it is useful to introduce some preliminary notation. In the following, we denote by $\mathbf{w}_{i.} \equiv (\mathrm{w}_{i,1}, \cdots, \mathrm{w}_{i,T-1})'$ the $(T-1) \times 1$

vector of weights specific for the i-th asset, and by $\mathbf{w}_{.t-1} \equiv (\mathrm{w}_{1,t-1}, \cdots, \mathrm{w}_{N,t-1})'$ the $N \times 1$ vector of weights at time $(t-1)$, for every $2 \leq t \leq T$, with the $N \times T$ matrix $\mathbf{W} = (\mathbf{w}_{.1}, \cdots, \mathbf{w}_{.T-1}) = (\mathbf{w}_{1.}, \cdots, \mathbf{w}'_{N.})$.

Assumption 9. *(CSR WLS weights)*

(i)
$$\frac{\mathbf{1}'_N \mathbf{W}_{t-1} \mathbf{1}_N}{N} \rightarrow_p 1.$$

(ii) For any real number $h > 1$ then,
$$\frac{\mathbf{1}'_N \mathbf{W}^h_{t-1} \mathbf{1}_N}{N} \rightarrow_p \mu^h_{\mathrm{w},t-1}$$

(iii)
$$\frac{1}{N} \sum_{i=1}^{N} \mathbf{w}_{i.} \mathbf{w}'_{i.} \rightarrow_p \mathbf{\Sigma}_{\mathrm{W}}.$$

Assumption 10. *(Weighted loadings)* Let $\mathbf{W}_{t-1}$ satisfy Assumption 9 and let the loadings $\boldsymbol{\beta}_i$ be a non-random sequence. As $N \rightarrow \infty$, then

$$\frac{1}{N} \mathbf{B}' \mathbf{W}_{t-1} \mathbf{1}_N \rightarrow_p \boldsymbol{\mu}_\beta \qquad \text{and} \qquad \frac{1}{N} \mathbf{B}' \mathbf{W}_{t-1} \mathbf{B} \rightarrow_p \mathbf{\Sigma}_\beta, \tag{A.15}$$

such that
$$\mathbf{\Sigma}_{\mathrm{X}} \equiv \begin{bmatrix} 1 & \boldsymbol{\mu}'_\beta \\ \boldsymbol{\mu}_\beta & \mathbf{\Sigma}_\beta \end{bmatrix} > 0. \tag{A.16}$$

Assumption 11. *(Weighted cross-sectional moments of returns' innovations)* As $N \rightarrow \infty$,

(i) Let $0 < \sigma^2 < \infty$.Then, for every $2 \leq t \leq T$:
$$\frac{1}{N} \sum_{i=1}^{N} \mathrm{w}_{i,t-1} \left(\sigma_i^2 - \sigma^2 \right) = o_p \left(\frac{1}{\sqrt{N}} \right), \tag{A.17}$$

(ii)
$$\sum_{i,j=1}^{N} \mathrm{w}_{i,t-1} \mid \sigma_{ij} \mid \mathbb{1}_{\{i \neq j\}} = o_p \left(N \right). \tag{A.18}$$

(iii) Let $0 < \mu_4 < \infty$, and let $\mu_{4i} = \mathrm{E}[\epsilon_{it}^4]$. Then, for every $2 \leq t \leq T$:

$$\frac{1}{N}\sum_{i=1}^{N} \mathrm{w}_{i,t-1}\mu_{4i} \rightarrow_p \mu_4, \tag{A.19}$$

(iv) Let $0 < \sigma_4 < \infty$. Then, for every $2 \leq t \leq T$:

$$\frac{1}{N}\sum_{i=1}^{N} \mathrm{w}_{i,t-1}\sigma_i^4 \rightarrow_p \sigma_4, \tag{A.20}$$

(v) *Let $\kappa_3(a,b,c)$ denote the third-order cumulant of the random variables $a, b,$ and c. Then,*

$$\kappa_3[\epsilon_{i,t}, \epsilon_{j,s}, \mathrm{w}_{j,h}] = 0, \quad \textit{and} \quad \kappa_3[\epsilon_{i,t}, \epsilon_{j,s}, \mathbf{z}_{j,h}] = \mathbf{0}_{K_{\mathrm{z}}}. \tag{A.21}$$

(vi) Let $\kappa_{4,iiii} = \kappa_4[\epsilon_{i,t}, \epsilon_{i,t}, \epsilon_{i,t}, \epsilon_{i,t}]$ denote the fourth-order cumulant of the asset-specific error $\{\epsilon_{i,t}, \epsilon_{i,t}, \epsilon_{i,t}, \epsilon_{i,t}\}$. Then, for some $0 \leq |\kappa_4| < \infty$ and for every $2 \leq t \leq T$:

$$\frac{1}{N}\sum_{i=1}^{N} \mathrm{w}_{i,t-1}\kappa_{4,iiii} \rightarrow_p \kappa_4. \tag{A.22}$$

(vii) For every $3 \leq h \leq 8$, all the following mixed cumulants of order h satisfy

$$\sup_{i_1} \sum_{i_2,\ldots,i_h=1}^{N} |\kappa_{h,\mathrm{w}_{i_1,t-1}i_2\ldots i_h}| = o\left(N\right), \tag{A.23}$$

and

$$\sup_{i_1} \sum_{i_2,\ldots,i_h=1}^{N} |\kappa_{h\mathrm{w}_{i_1,t-1},\mathbf{z}_{i_2,r},i_3\ldots i_h}| = o\left(N\right), \tag{A.24}$$

for at least one i_j $(2 \leq j \leq h)$ different from i_1, where $\kappa_{h,\mathrm{w}_{i_1,t-1}i_2\ldots i_h}$ is the mixed cumulant in the $\{\mathrm{w}_{i_1,t-1}, \epsilon_{i_2,s}, \cdots, \epsilon_{i_h,s}\}$ of order h, and $\kappa_{h,\mathrm{w}_{i_1,t-1},\mathbf{z}_{i_2,r},i_3\ldots i_h}$ is the mixed cumulant in the $\{\mathrm{w}_{i_1,t-1}, \mathbf{z}_{i_2,r}, \epsilon_{i_3,s}, \cdots, \epsilon_{i_h,s}\}$ of order h.

Assumption 12. *(Weighted moments and CLT of anomalies) We define the $(T-1)^2 \times 1$ vector $\mathbf{v}_i \equiv (\boldsymbol{\epsilon}_i \otimes \mathbf{w}_i)$ and the corresponding $N \times (T-1)^2$ matrix $\mathbf{V} \equiv (\mathbf{v}_1, \cdots, \mathbf{v}_N)'$, such that $\mathrm{E}[\mathbf{v}_i] \equiv \boldsymbol{\mu}_{\mathrm{v}_i} < \infty$, and $\boldsymbol{\Sigma}_{\mathrm{v},ij} \equiv \mathrm{Cov}\left[\mathbf{v}_i, \mathbf{v}_j\right]$.*

(i)

$$\frac{\boldsymbol{\epsilon}\left(\mathbf{W}_{t-1} - \mathrm{E}[\mathbf{W}_{t-1}]\right)\boldsymbol{\epsilon}'}{N} \rightarrow_p \mathbf{0}_{(T-1)\times(T-1)}.$$

(ii)

$$\frac{\mathbf{Z}_{t-1}'\mathbf{W}_{t-1}\mathbf{1}_N}{N} \quad \to_p \quad \boldsymbol{\mu}_{\mathrm{z},t-1} \text{ and } \frac{\mathbf{Z}_{t-1}'\mathbf{W}_{t-1}\mathbf{Z}_{t-1}}{N} \to_p \boldsymbol{\Sigma}_{\mathrm{Z},t-1}.$$

(iii) *Let* $\boldsymbol{\Sigma}_{\mathrm{ZW}}$ *be a finite* $K_{\mathrm{z}}(T-1)\times(T-1)$ *matrix. Then,*

$$\frac{\mathbf{Z}'\mathbf{W}}{N} \to_p \boldsymbol{\Sigma}_{\mathrm{ZW}}.$$

(iv)

$$\frac{1}{N}\left(\mathbf{Z}_{t-1} - \mathrm{E}[\mathbf{Z}_{t-1}]\right)'\left(\mathbf{W}_{t-1} - \mathrm{E}[\mathbf{W}_{t-1}]\right)\boldsymbol{\epsilon}' \to_p \mathbf{0}_{K_{\mathrm{z}}\times(T-1)}.$$

(v)

$$\frac{1}{N}\left(\mathbf{Z}_{t-1} - \mathrm{E}[\mathbf{Z}_{t-1}]\right)'\mathbf{W}_{t-1}\boldsymbol{\epsilon}' - \frac{1}{N}\left(\mathbf{Z}_{t-1} - \mathrm{E}[\mathbf{Z}_{t-1}]\right)'\boldsymbol{\epsilon}' = o_p\left(N^{-\frac{1}{2}}\right).$$

(vi)

$$\frac{\mathbf{X}'\mathbb{M}_{1_N}\mathbf{V}}{N} = o_p\left(N^{-\frac{1}{2}}\right), \quad \text{and} \quad \frac{\mathbf{Z}'\mathbb{M}_{1_N}\mathbf{V}}{N} = o_p\left(N^{-\frac{1}{2}}\right).$$

(vii)

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Sigma}_{\mathrm{v},ii} \to \boldsymbol{\Sigma}_{\mathrm{V}} \equiv \sigma^2\mathbf{I}_{T-1}\otimes\boldsymbol{\Sigma}_{\mathrm{W}}, \quad \text{and} \quad \sum_{i=1}^{N}\boldsymbol{\Sigma}_{\mathrm{v},ij}\mathbb{1}_{i\neq j} = o(N)$$

(viii)

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{v}_i - \boldsymbol{\mu}_{\mathrm{v}_i}\right) \to_d N\left(\mathbf{0}_{(T-1)^2}, \boldsymbol{\Sigma}_{\mathrm{V}}\right) \quad \text{and} \quad \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\mu}_{\mathrm{v_i}} = o\left(N^{-\frac{1}{2}}\right).$$

(ix)

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{z}_i - \boldsymbol{\mu}_{\mathrm{z}_i}) \to_d N(\mathbf{0}_{K_z(T-1)}, \boldsymbol{\Sigma}_{ZZ}) \text{ and } \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{\mu}_{\mathrm{z}_i} - \boldsymbol{\mu}_z) = o(N^{-\frac{1}{2}}).$$

A.1.2 Additional assumptions required for estimation under model misspecification

Assumption 13. *(mixed-moments of pricing errors)*

(i)

$$\frac{1}{N}\boldsymbol{\epsilon}\mathbf{m}_{t-1} \to_p \begin{bmatrix} \boldsymbol{\theta}_{t-1,m} \\ \mathbf{0}_{T-t+1} \end{bmatrix},$$

with $\boldsymbol{\theta}_{t-1,m} \equiv (\theta_{t-3,m}, \theta_{t-4,m}, \ldots, \theta_{0,m})'$, defined as, for every $2 \leq s, t \leq T$,

$$\frac{1}{N}\sum_{i=1}^{N} \epsilon_{i,s} m_{i,t-1} \to_p \theta_{t-1-s,m}, \text{ such that } \theta_{u,m} = 0 \text{ for } u < 0.$$

(ii)

$$\frac{1}{N}\mathbf{m}_{t-1}'\mathbf{m}_{t-1} \to_p \sigma_{t-1mm}.$$

(iii)

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{P}_{\tilde{D}_i}\boldsymbol{\epsilon}_i m_{i,t-1} \to_p \mathbf{0}_{T-1}.$$

A.1.3 Additional assumptions required for the cross-sectional R-squared test

In this Section we introduce additional assumptions that are required to derive the R-squared test described in Section 9.

Assumption 14. (i)

$$\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta = o\left(N^{-\frac{1}{2}}\right) \text{ and } \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\beta}_i\boldsymbol{\beta}_i' - \boldsymbol{\Sigma}_\beta = o\left(N^{-\frac{1}{2}}\right).$$

(ii)

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} ((\mathbf{z}_i \otimes \mathbf{z}_i) - \mathrm{vec}(\boldsymbol{\Sigma}_\mathrm{Z})) \to_d N(\mathbf{0}_{K_\mathrm{z}^2}, \mathbf{U}_\mathrm{Z}), \quad \text{with}$$
$$\frac{1}{N}\sum_{i=1}^{N} \mathrm{E}\left[(\mathbf{z}_i \otimes \mathbf{z}_i) - \mathrm{vec}(\boldsymbol{\Sigma}_\mathrm{Z})\right] = o\left(N^{-\frac{1}{2}}\right), \ \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}\left[(\mathbf{z}_i \otimes \mathbf{z}_i) - \mathrm{vec}(\boldsymbol{\Sigma}_\mathrm{Z})\right]\left[(\mathbf{z}_i \otimes \mathbf{z}_i) - \mathrm{vec}(\boldsymbol{\Sigma}_\mathrm{Z})\right]' \to \mathbf{U}_\mathrm{Z},$$
$$\sum_{\substack{i,j=1 \\ i \neq j}}^{N} \mathrm{E}\left[(\mathbf{z}_i \otimes \mathbf{z}_i) - \mathrm{vec}(\boldsymbol{\Sigma}_\mathrm{Z})\right]\left[(\mathbf{z}_j \otimes \mathbf{z}_j) - \mathrm{vec}(\boldsymbol{\Sigma}_\mathrm{Z})\right]' = o(N), \text{ and } \frac{1}{N}\sum_{i,j=1}^{N} \mathrm{Cov}\left[(\mathbf{z}_i \otimes \mathbf{z}_i), \mathbf{z}_j'\right] \to \boldsymbol{\Sigma}_{\mathrm{z}\otimes\mathrm{z}}.$$

(iii)

$$\sqrt{N}\left(\frac{\mathbf{Z}'\mathbf{1}_N}{N} - \boldsymbol{\mu}_{\mathrm{z},T-1}\right) \to_d N\left(\mathbf{0}_{K_\mathrm{z}(T-1)}, \boldsymbol{\Sigma}_\mathrm{Z} - \boldsymbol{\mu}_{\mathrm{z},T-1}\boldsymbol{\mu}_{\mathrm{z},T-1}'\right).$$

(iv)

$$\frac{1}{N}\sum_{i,j=1}^{N}\operatorname{Cov}\left((\mathbf{z}_i\otimes\mathbf{z}_i),(\boldsymbol{\epsilon}_j'\otimes\boldsymbol{\epsilon}_j')\right)\to\boldsymbol{\Sigma}_{\mathrm{Z}\otimes\epsilon}=\mathbf{0}_{((T-1)K_z)^2\times(T-1)^2}.$$

(v)

$$\frac{1}{N}\sum_{i,j=1}^{N}\operatorname{Cov}\left((\mathbf{z}_i\otimes\mathbf{z}_i),\mathbf{u}_j'\right)\to\boldsymbol{\Sigma}_{\mathrm{ZU}}=\mathbf{0}_{((T-1)K_z)^2\times(T-1)^2K_z}.$$

A.2 Empirics: Tables and Plots

Table II: The table provides evidence on the correlation between each anomaly and the betas estimated from a given asset pricing factor model. The betas are estimated using a first-pass regression with $T = 120$ months, using the CAPM, the FF3 and FF5 and specifications. The first column of the table shows the average R-squared (i.e., the square of the correlation coefficient), together with its minimum and maximum value (in parenthesis), for each of the six categories and for each of the three model specifications (CAPM in Panel A, FF3 in Panel B and FF5 in panel C). In the second column we report, instead, the average percentage of times in which the correlation coefficient has been found to be statistically different from zero. To assess whether the correlation coefficient is statistically different from zero, we use the standard Pearson correlation test in the CAPM specification, while we use the p-value of the F-test for the FF3 and FF5 models. We report the results using a 5% confidence level.

Panel A: Average correlation between B and $\mathbf{Z}_t$ in the CAPM

	R-squared (min, max)	% of significance
Momentum	0.041 (0.003, 0.104)	60.83%
Value VS Growth	0.032 (0.001, 0.279)	60.06%
Investment	0.008 (0.001, 0.027)	36.35%
Profitability	0.008 (0.001, 0.017)	40.08%
Intangibles	0.029 (0.001, 0.255)	56.03%
Trading Frictions	0.087 (0.001, 0.461)	79.21%

Panel B: Average correlation between B and $\mathbf{Z}_t$ in FF3

	R-squared (min, max)	% of significance
Momentum	0.073 (0.081, 0.177)	71.55%
Value vs Growth	0.076 (0.003, 0.343)	80.93%
Investment	0.019 (0.005, 0.050)	58.12%
Profitability	0.032 (0.005, 0.109)	61.78%
Intangibles	0.065 (0.002, 0.296)	68.36%
Trading Frictions	0.178 (0.009, 0.590)	91.47%

Panel C: Average correlation between B and $\mathbf{Z}_t$ in FF5

	R-squared (min, max)	% of significance
Momentum	0.093 (0.012, 0.208)	76.43%
Value vs Growth	0.098 (0.005, 0.458)	81.65%
Investment	0.027 (0.008, 0.074)	63.16%
Profitability	0.048 (0.009, 0.130)	66.86%
Intangibles	0.083 (0.006, 0.357)	70.86%
Trading Frictions	0.195 (0.012, 0.645)	91.61%

Table III: The table reports the percentage of anomalies (within each category) for which we find a statistically superior fitting ability of a time-varying model (among ARMA(p,q) models, with p and q ranging from 1 to 6) over a white noise (i.e., an ARMA(0,0)). Column 1 of the table refers to the case in which we obtain the time series of the $\gamma_{z,t}$ estimates in (**??**), considering one anomaly at the time in the regression of $\mathbf{R}_t$ on $\mathbf{Z}_t$. Colums 2–4 use instead the time series of the $\gamma_{z,t}$ obtained by regressing the asset returns $\mathbf{R}_t$ on each of the anomaly $\mathbf{Z}_t$ at the time, and including also the betas estimated from the CAPM (Column 2), the FF3 (Column 3) and FF5 (Column 4) in the regression specification, using T=120 months.

Category	$\mathbf{Z}_t$	$\mathbf{Z}_t$+CAPM	$\mathbf{Z}_t$+FF3	$\mathbf{Z}_t$+FF5
Momentum	80.0	93.3	86.7	86.7
Value vs Growth	100.0	100.0	96.6	96.6
Investment	84.2	89.5	84.2	94.7
Profitability	93.3	96.7	93.3	86.7
Intangibles	89.8	95.9	89.8	91.8
Trading Frictions	92.6	100.0	100.0	96.3

Table IV: The table reports the average percent of anomalies (within each category) for which we reject the null hypotesis of no structural breaks in the time series of the premia estimates. Specifically, for each anomaly, we define the process $\Delta\gamma_{z,t} = \delta_1 + \delta_2\gamma_{z,t} + u_t$ and test for structural changes by analyzing the cumulative sums of OLS-residuals (CUSUM) from the estimated linear model (see Ploberger, Krämer, and Kontrus (1989) and Zeileis, Leisch, Hornik, and Kleiber (2002)). in the first column of the table, the time series of the $\gamma_{z,t}$ is obtained by regressing the asset returns $\mathbf{R}_t$ on each of the anomaly $\mathbf{Z}_t$ at the time. Colums 2–4 use instead the time series of the $\gamma_{z,t}$ obtained by regressing the asset returns $\mathbf{R}_t$ on each of the anomaly $\mathbf{Z}_t$ at the time, but including also the betas estimated from the CAPM (Column 2), the FF3 (Column 3) and FF5 (Column 4) in the regression specification, using T=120 months.

Category	$\mathbf{Z}_t$	$\mathbf{Z}_t$+CAPM	$\mathbf{Z}_t$+FF3	$\mathbf{Z}_t$+FF5
Momentum	40.0	40.0	40.0	33.3
Value vs Growth	13.8	17.2	20.7	20.7
Investment	70.0	63.3	66.7	60.0
Profitability	15.8	26.3	21.1	15.8
Intangibles	20.4	24.5	28.6	28.6
Trading Frictions	22.2	22.2	44.4	48.2

Table V: The table reports the average percentage of paired sub-periods (aggregated at the category level) for which we find a statistically significant mean difference. Specifically, using the time series of estimated $\gamma_{z,t}$ (one for each anomaly at the time), we consider all the possible pairs of sub-samples (not necessarily consecutive) of length equal to either three years (Panel A) or five years (Panel B), and calculate the difference in the estimates between the corresponding values in each pair. We then calculate the mean difference between each pair of sub-samples and derive the corresponding t-statistic under the null hypothesis that there is no shift in the mean of the two sub-periods (i.e., the mean difference is equal to zero). Column 1 of the table refers to the time series of $\gamma_{z,t}$ which is obtained by regressing the asset returns $\mathbf{R}_t$ on each of the anomaly $\mathbf{Z}_t$ at the time. Colums 2–4 use instead the time series of $\gamma_{z,t}$ obtained by regressing the asset returns $\mathbf{R}_t$ on each of the anomaly $\mathbf{Z}_t$ at the time, but including also the betas estimated from the CAPM (Column 2), the FF3 (Column 3) and FF5 (Column 4) in the regression specification.

Category	$\mathbf{Z}_t$	$\mathbf{Z}_t$+CAPM	$\mathbf{Z}_t$+FF3	$\mathbf{Z}_t$+FF5
Panel A: Paired sub-period of length $T = 36$ (3 years)				
Momentum	15.3	21.1	21.6	23.8
Value vs Growth	21.3	28.8	32.7	35.4
Investment	29.6	34.1	31.2	31.3
Profitability	12.2	21.4	22.9	26.1
Intangibles	14.1	22.4	26.0	28.4
Trading Frictions	10.6	27.3	35.4	34.0
Panel B: Paired sub-period of length $T = 60$ (5 years)				
Momentum	28.0	29.3	22.7	20.9
Value vs Growth	27.1	34.3	30.6	32.9
Investment	43.3	37.6	29.9	31.8
Profitability	29.1	31.6	28.0	31.1
Intangibles	20.0	26.3	25.8	28.2
Trading Frictions	18.6	44.5	43.0	43.6

Figure 3: **The R-squared distribution across anomalies in the regression of γ_{z} against $\boldsymbol{\delta}_{\text{f}}$.** The figure shows the distribution, across anomalies, of the R-squared values in the time-series regression of each $\boldsymbol{\gamma}_{\text{z}}$ against $\boldsymbol{\delta}_{\text{f}}$. The time series of both $\boldsymbol{\gamma}_{\text{z}}$ and $\boldsymbol{\delta}_{\text{f}}$ are obtained from the (cross-sectional) regression $\mathbf{R}_t = \hat{\mathbf{X}}\boldsymbol{\Gamma}_{\text{f},t-1} + \mathbf{Z}_{t-1}\boldsymbol{\gamma}_{\text{z},t-1} + \boldsymbol{\eta}_t$, $t = 2, .., T-1$, using our time-varying estimator in (32). In each regression, we consider one anomaly at the time (i.e., $K_{\text{z}} = 1$), while the estimated betas in $\hat{\mathbf{X}}$ are obtained using (26), in which the matrix $\mathbf{F}$ contains the Fama-French five-factors (i.e., $K_{\text{f}} = 5$). The purple dotted line represents the average R-squared value, while the two blue dotted lines denote the 5th and 95th percentile values, respectively. The vertical gray bands depict the 95% confidence interval associated to each R-squared value. The confidence intervals (CI) have been calculated using the Cohen et al. (2003) formula, that is $CI = R^2 \pm 2SE_{R^2}$, where $SE_{R^2} = \sqrt{(4R^2(1-R^2)^2(n-k-1)^2)/((n^2-1)(n+3))}$, with n denoting the number of observations and k being the number of independent variables in the regression of $\boldsymbol{\gamma}_{\text{z}}$ against $\boldsymbol{\delta}_{\text{f}}$.

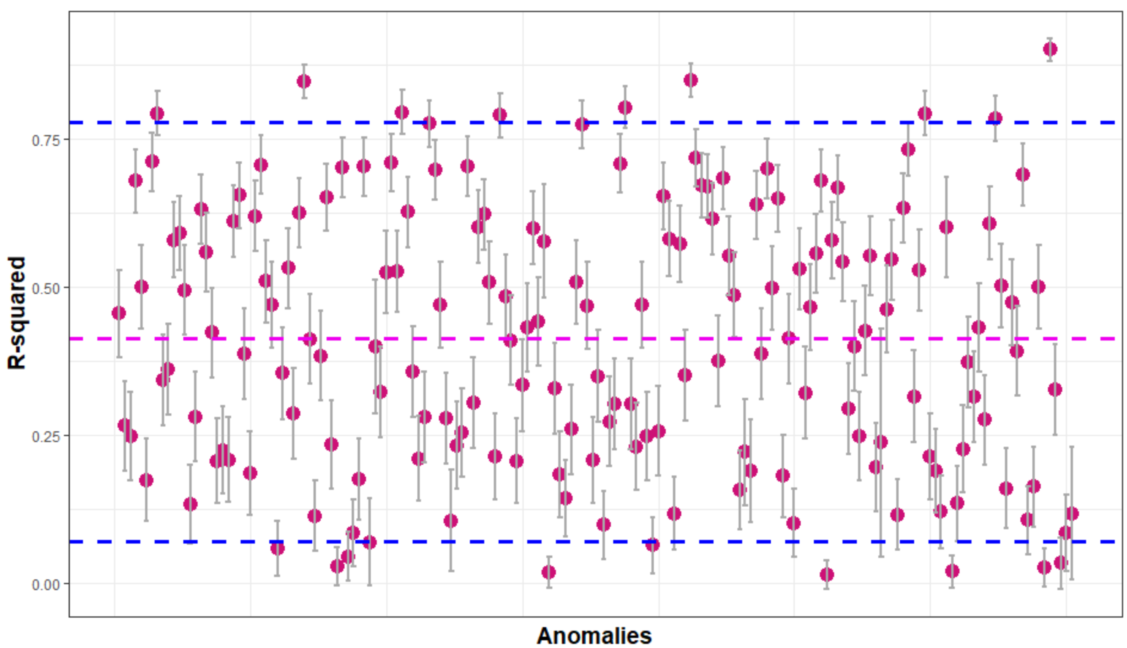


Figure 4: **Heatmap of the t-statistics distribution - time varying case**. The figure shows the heatmap of the t-statistics distribution obtained in each of the 170 univariate model (vertical axis) and for each month (horizontal axis). Each cell in the map represents the degree of statistical significance of the t-statistics with a different color, from gray (non-significant t-stat), to yellow (significance at only 10% level), orange (significance at 5% level), and red (significance at 1% level). The results are obtained by performing univariate regressions of asset returns on the market factor and each of the 170 anomalies, using the theoretical results of Section 5. of time-varying premia and anomalies. The analysis uses balanced panels at each month, with a reference period ranging from January 1986 to December 2020. At each month t, the market beta is obtained by running a first-pass regression using a rolling window on the past two years of data ($T = 24$).

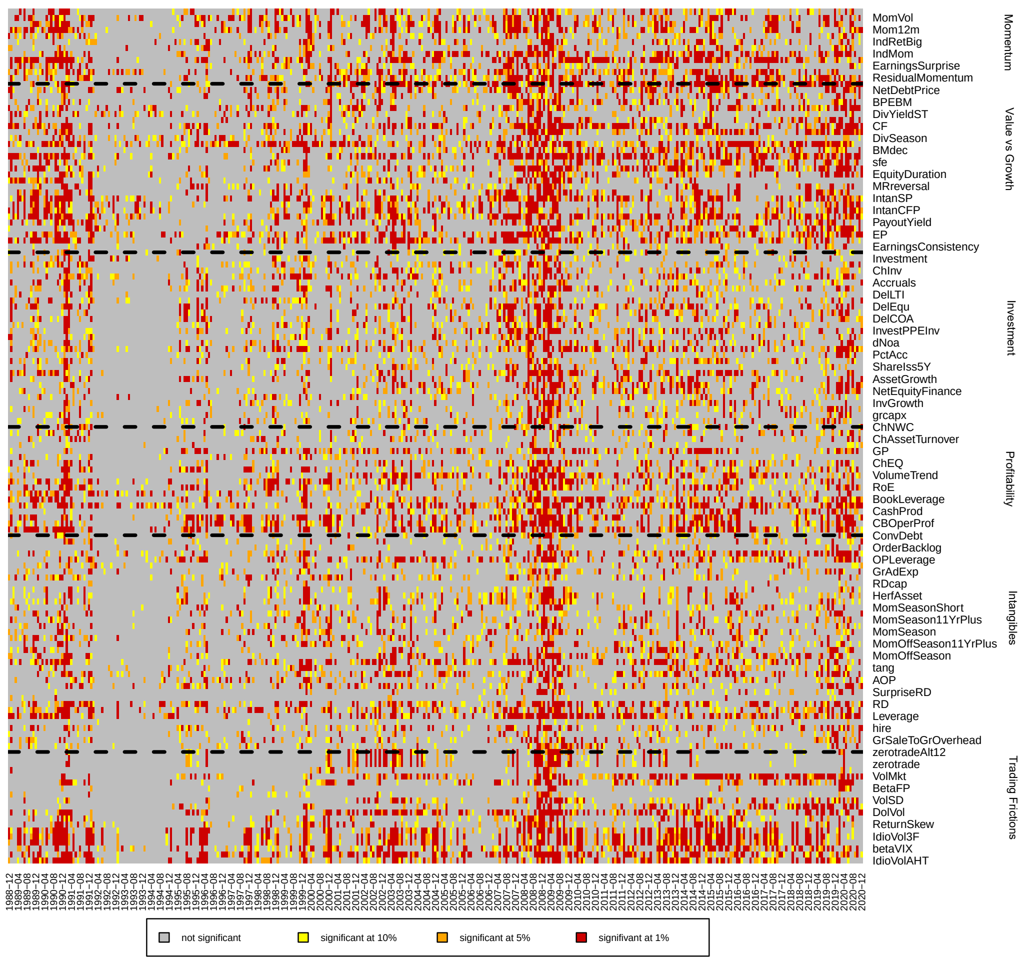


Figure 5: **Anomalies and financial crises - time varying case**. The figure shows the percentage of anomalies found to be significant at 5% (or lower) confidence level at each point in time. The light gray bands correspond to NBER recession dates and to various economic and financial crises. The results are obtained by performing univariate regressions of asset returns on the market factor and each of the 170 anomalies, using the theoretical results of Section 5. of time-varying premia and anomalies. The analysis uses balanced panels for each month, with a reference period ranging from January 1986 to December 2020. At each month t, the market beta is obtained by running a first-pass regression using a rolling window on the past two years of data ($T = 24$).

.

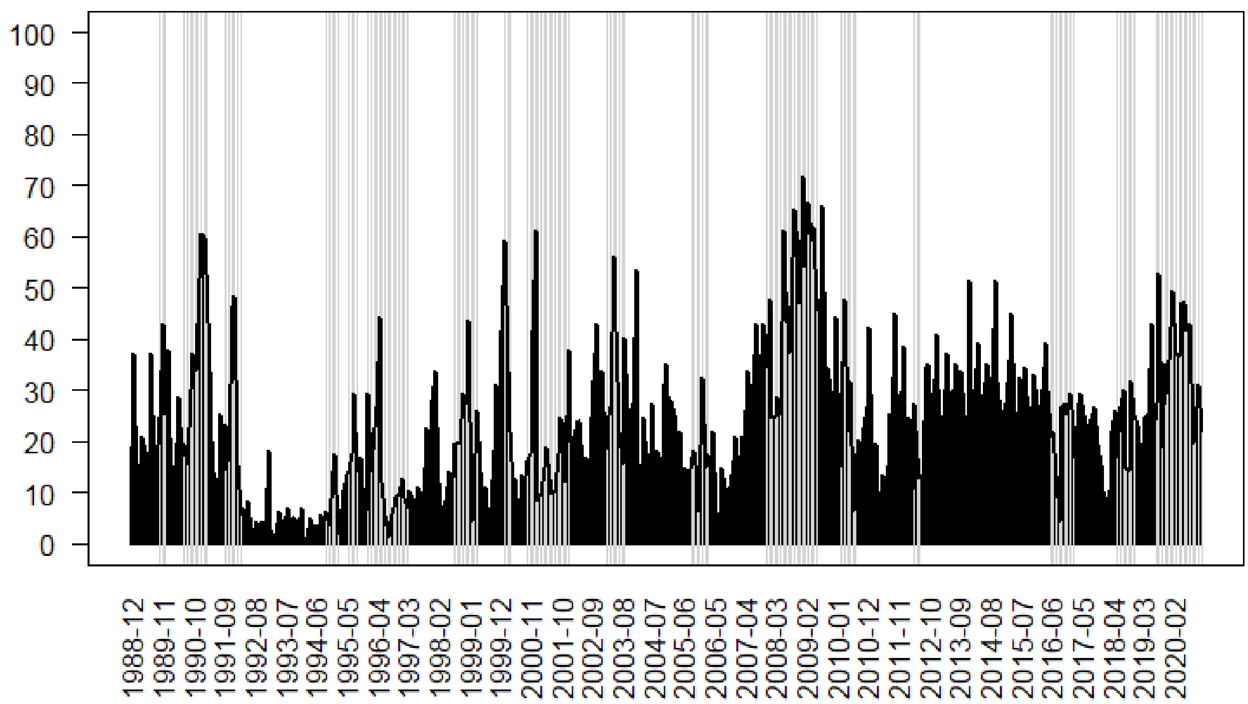


Figure 6: **"Best" representative sets of anomalies in multivariate regressions**. The figure shows the time-varying sets of anomalies that have been used to run multivariate regressions at each month. Each red point denotes the anomaly that has been picked in each category (vertical axis) and in each month (horizontal axis), using the empirical procedure described in Section 10.4.2.

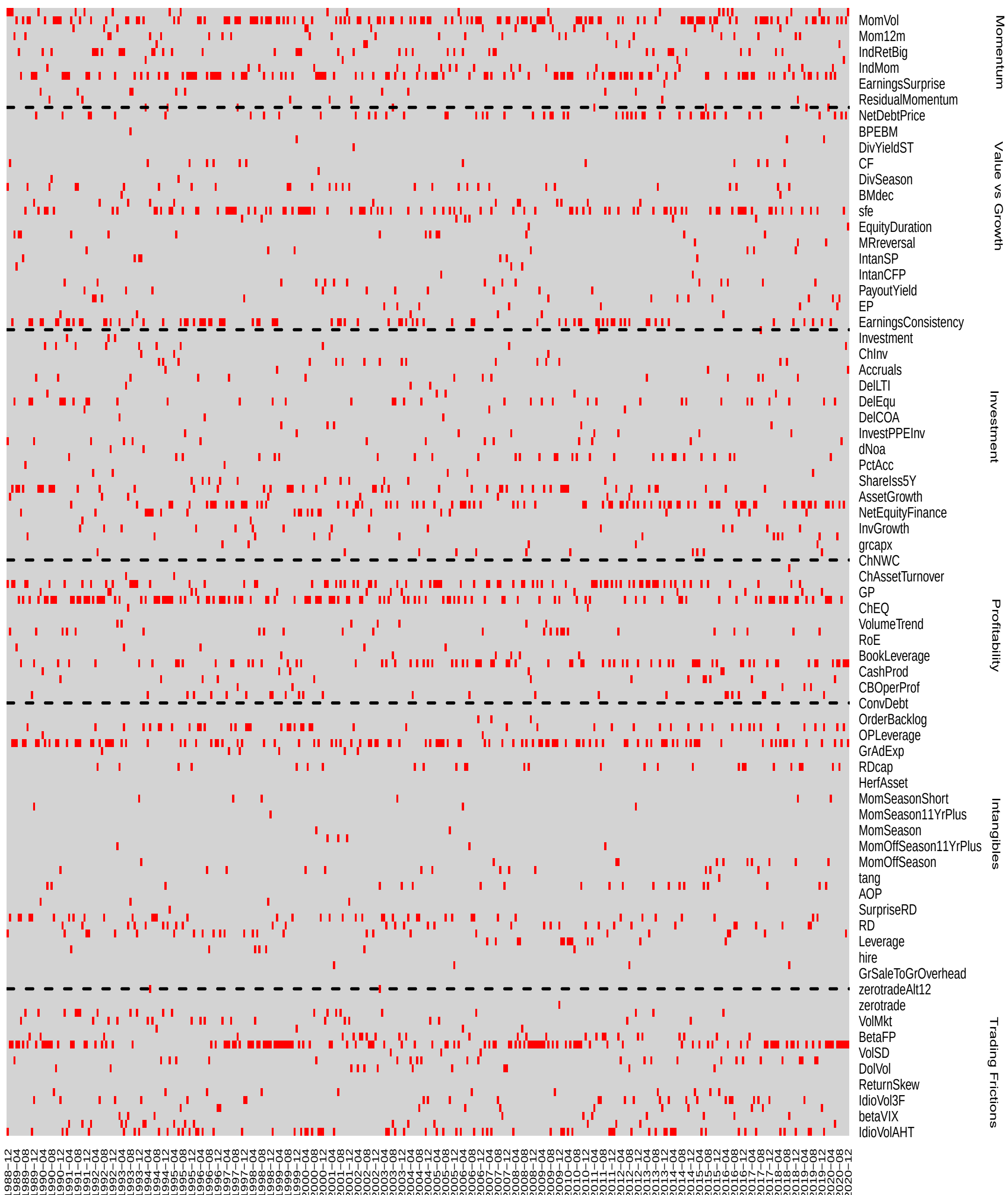
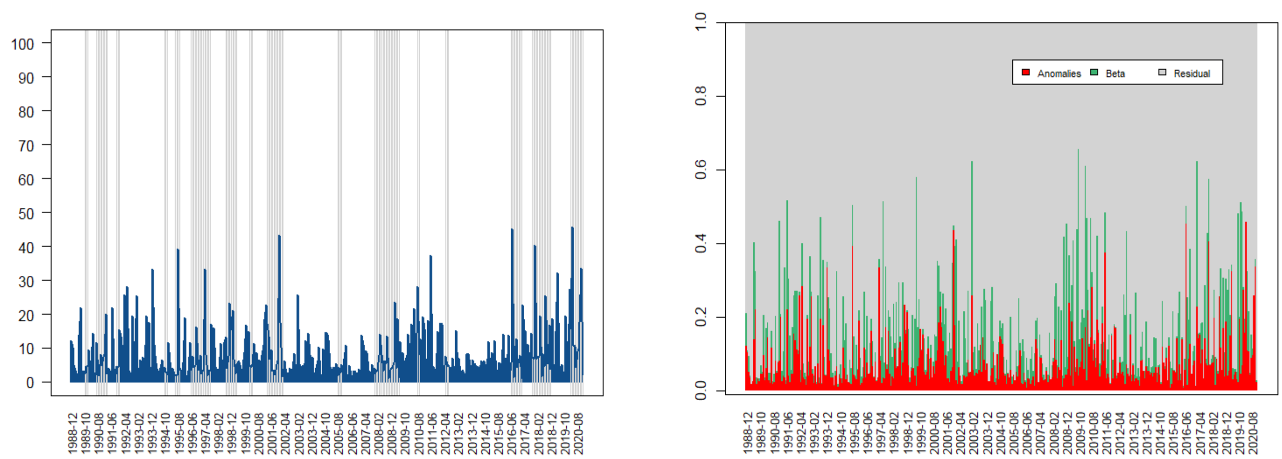


Figure 7: **Anomalies' contribution using time-varying multivariate regressions**. The figure shows the time series of the R_z^2 statistics (a), together with the total variance decomposition (b) obtained in each multivariate regression. The results are obtained by performing multivariate regressions of asset returns on the market factor and a set of six anomalies, selected using the empirical procedure described in Section 10.4.2. The analysis is based on the theoretical results of Sections 5 and 9 for time-varying premia and anomalies. The application uses balanced panels at each month, with a reference period ranging from January 1986 to December 2020. At each month t, the market beta is obtained by running a first-pass regression using a rolling window on the past two years of data ($T = 24$).



(a) Time series of R_z^2

(b) Total variance decomposition

Figure 8: **Testing the joint contribution of anomalies: time series of p-values**. The figure shows the time series of p-values associated with the $\mathcal{T}_z$ statistics for each multivariate model at each point in time. The null hypothesis is that $H_0 : \boldsymbol{\gamma}_z = \mathbf{0}_{K_z}$, against the alternative that at least one anomaly is different from zero, i.e., $H_1 : \boldsymbol{\gamma}_z \neq \mathbf{0}_{K_z}$, with $K_z = 6$. The yellow bands represent the p-values > 0.05 , for which we cannot find evidence to reject the null hypothesis. The blue lines refer to the p-values ≤ 0.05, i.e. all the periods in which we can reject the null hypothesis at the 5% confidence level. The analysis is based on the theoretical results of Theorem 8 (i), where the asymptotic distribution of the statistic $\mathcal{T}_z$ under H_0 has been tabulated using 10,000 random draws from six i.i.d. χ^2_1, weighted with the estimated values $(\hat{c}_1, ..., \hat{c}_6)$ obtained in each multivariate model. The results are obtained by performing multivariate regressions of asset returns on the market factor and a set of six anomalies, selected using the empirical procedure described in Section 10.4.2. The application uses balanced panels at each month, with a reference period ranging from January 1986 to December 2020. At each month t, the market beta is obtained by running a first-pass regression using a rolling window on the past two years of data ($T = 24$).