Misspecification and Weak Identification in the Nontraded Factor Zoo

Amedeo Andriollo, Cesare Robotti, Xinyi Zhang *

January 27, 2024

Preliminary draft, comments are welcome!

Abstract

To explain the cross-section of asset returns, a "*zoo*" of economic factors that are not portfolios (nontradables) have been proposed. In contrast to traded factors, the non-traded factors tend to have smaller correlations with the asset returns, risk premia inference tends to be more fragile, and the issue of weak identification might be exacerbated by the degree of model misspecification. Yet, robust inference has often been overlooked in many empirical works, and limited efforts have been devoted to "*domesticate*" such factors. With respect to the most commonly used asset returns, this paper aims at providing a comprehensive re-examination of the non-traded factor zoo published in top academic journals, with a focus on the aforementioned fragilities. We confirm that the vast majority of the original model specifications suffer such problems, and robust inference leads to strong evidences that most of the proposed nontraded factors are unpriced. The findings are more drastic when considering additional the market factor as control. When considering the entirety of the nontradables with PCA, we conclude however that the zoo carries some nonzero pricing ability.

Keywords: Factor Zoo, Nontradable, Misspecification, Weak identification, Risk premium, Asset pricing

JEL Codes: C58, G12

*E-mails: amedeo.andriollo@warwick.ac.uk; cesare.robotti@wbs.ac.uk; xinyi.zhang.2@warwick.ac.uk; A. Andriollo is at the Economics Department of the University of Warwick. C. Robotti is at Warwick Business School. X. Zhang is at Warwick Business School and CY Cergy Paris University. For publicly sharing their data, we thank David Ardia, Oliver Boguth, Yong Chen, Zhi Da, Fernando Duarte, Stefano Giglio, Amit Goyal, Joachim Grammig, Bernard Herskovic, Dashan Huang, Mahyar Kargar, Tim Kroencke, Yukun Liu, Martin Lettau, Sydney Ludvigson, Asaf Manela, Tyler Muir, Jun Pan, Christopher Polk, Ronnie Sadka, Robert Stambaugh, Philipp Schuster, Stijn Van Nieuwerburgh, Maria Vassalou, Jessie Jiaxu Wang, Jeffrey Wurgler, Guofu Zhou. Andriollo would like thank Eric Renault, Pablo Beker and Giulio Rossetti for the useful discussions.

1 Introduction

Throughout the past decades, a universe or "*zoo*" of pricing factors (over 400+) have been documented as proxies of systemic risk dimensions, or in other words, deemed to be explaining the cross-sectional variation of asset returns. However, their statistical properties have been questioned only recently.

After Cochrane (2011), a considerable number of empirical asset pricing studies examine the universe of traded factors (or tradables), i.e., factors that are portfolio excess returns or return spreads. Such investigations take different angles: from Harvey et al. (2016), which propose to use a higher hurdle for significance to mitigate the problem of multiple testing, Hou et al. (2020) that examine the out-of-sample performance with the t-statistics of long-short portfolios, and to apply machine learning techniques to capture the marginal explanatory power, such as the case of double-selection LASSO procedure in Feng et al. (2020). However, given the heterogeneity of the outlooks, the studies do not direct to a unanimous conclusion: for instance, Hou et al. (2020) unveil disappearing performance for most of the traded factors, while Chen and Zimmermann (2021) conclude the opposite.

Although the numerous scrutinies of the traded factor zoo, a limited attention has been paid to the economic factors that are not portfolios or return spreads (e.g., consumption growth, unexpected variation in inflation, news index, and others[1]), and only few notable examples (see Section 2.1) provide a limited sketch of their statistical properties. Given the lack of a systematic scrutiny of the "*nontraded factor zoo*", our ambition is thus to offer an extensive reexamination of the deemedly pricing nontradables that, on one hand, extends and consolidates to a larger scope the previous analysis, and on the other, complements the recent literature about scientific replication of empirical asset pricing studies. Our contribution is along four aspects.

In a similar fashion of Harvey et al. (2016), we are the first to offer a census[2] the nontraded or nontradable factors, which have been proposed as proxies of some systemic risk dimensions in the context of asset pricing. The nontraded factor zoo consists into 104 distinct non-traded factors from 52 papers, covering up to Februrary 2023. For further details, please refer to Section 4.1.

[1] More about the non-traded factors can be found in Table 1-6 in the Appendix B.

[2] A considerable number of authors share the factors on their websites or through the replication package of the journal website. We wish them to thank for publicly sharing the data. For completeness, we provide all the non-traded factors in a unique source.

Secondly, we investigate the pricing ability of such zoo by examining their explanatory power in the cross-sections of the commonly used classes of portfolios. Our goal is to scientifically replicate the model specifications in asset-pricing literature and, by using established econometric techniques, to conduct identification and/or misspecification robust inference on the risk premia, so to understand whether the original nonzero pricing conclusions of the non-tradables remain unaltered.
Specifically, with respect to commonly used asset returns that are supposed to represent the financial markets (equity and non-equity), we individually examine each model specification (right-hand-side variables) used by the papers which originally propose the non-traded factors, testing i) the presence of weak identification and/or misspecification ("*are the original models prone to suffer from risk premia identification issues?*"), and ii) whether the risk premia inference is fragile to such problems ("*after accounting for such issues, can we still conclude that the proposed factor(s) is priced?*"). The aim of this design is, first, to motivate why robust risk premia inference matters, second, to draw credible pricing conclusions with respect to the portfolios which are standard representatives of the asset return universe.

Third, after examining the original model specifications, we consider a plethora of specifications which are relevant to further investigate the magnitude of weak identification and misspecification problems. Regarding weak identification, we inspect single-factor models while controlling for the market factor (MKT). This exercise aims to understand how much marginal (linear) explanatory contribution is added by the non-traded zoo with respect to the natural traded benchmark. Regarding the degree of misspecification, we also study the sensitivity of the results to the presence/absence of the cross-sectional intercept, i.e., with or without constraining the zero beta rate.[3]

Lastly, we shift our assessment from model-specific to global nontraded zoo. Parallel to Giglio and Xiu (2021)'s methodology, we extract the first relevant principal components from a balanced panel of the nontradables, after being mimicked onto portfolios. Thus, we run the same machinery as before to test the cross-sectional pricing power of these representative factors, in the pursuit of understanding the linear information entailed in the nontraded zoo.

Our findings point out that weak identification is a pervasive issue, since around $40 - 45\%$ of the original model specifications, on average, fall in such category. While examining the nontraded factors individually, the shares do not change much, as the single-

[3]The benchmark results in Section 5.3 are with the cross-sectional intercept.

factor models prone to be weakly identified are about 35% for equity and 45% for non-equity. Once controlling for the MKT factor, the proportion goes up to 50%, thus strongly suggesting that a non-negligible portion of the proposed nontraded factors are capturing similar linear information entailed by the MKT factor. The overall picture is more drastic if one considers exclusively the Kleibergen and Zhan (2018)'s GRS-FAR test and the Kleibergen and Zhan (2020)'s Identification-robust mimicking-portfolio test, which suggest that, on average, more than 80% of the original models suffer of weak identification (and/or misspecification[4]).

By comparing identification tests, we are prone to conclude that the Kleibergen (2009)'s KP and Kleibergen and Zhan (2020)'s F-test tend to have low rejection rates (lower than Gospodinov et al. (2017)'s CD-test and Chen and Fang (2019)'s test), which might be motivated by low power of the tests and/or strong heteroskedasticity of the covariance matrix of the exposures.

Less agreeing conclusions can be drawn from the analysis on the issue of misspecification. While the specification test based on the HJ-distance points out that around 65% of the original models might suffer from such issues, the Hansen J-test suggests rather that 90% of such models have no mispricing. The divergence between statistics is not surprising: first, in line with Kleibergen and Zhan (2022)'s discussion, we indeed find that, in more than 80% of the original models, there is a negligible difference between the Anderson Rubin statistic and the J-statistic; second, the J-test correlates more with the Chen and Fang (2019)'s test when the model is dubbed as rank deficient, or in other words, the models that are dubbed as correctly specified by the J-test are more prone to suffer of weak identification issues.

In view of these findings, we conduct inference on the pricing of the nontraded factors by regarding methodologies robust to the aforementioned issues. As we expect, non-robust methodologies tend to inflate the number of priced factors by almost 3-4 times with respect to the robust ones. For the original model specifications, the standard methodologies (Fama and MacBeth (1973)'s and Shanken (1992)'s) suggest that about $30-40\%$ of the proposed nontraded factors are priced, whereas the robust procedures indicate that less than 10% are priced. In particular, for the large sample size ($T > 400$), although the overall picture for the latter ones remains unchanged, the percentage of priced nontraded factors increases for the former. Conversely, a higher proportion of nontraded factors ($+15/25\%$) turns out to be priced once removing the zero-beta rate from the model specifications.

[4] The confidence sets are not bounded, which means that the models is suffering one of the two aforementioned problems.

For a constructive appraisal, we report the survived[5] models whose nonzero pricing abilities are robust to the aforementioned issue: Campbell and Vuolteenaho (2004)'s news factors, for the US anomalies portfolios; Baker and Wurgler (2006)'s sentiment factors, for the US anomalies and ME/BM double-sorted Fama-French 25 portfolios; Sadka (2006)'s liquidity factors, for the He et al. (2017)'s US corporate bonds portfolios; Bali et al. (2017)'s uncertainty factors, for the Hou et al. (2020)'s anomalies portfolios; Chen et al. (2018)'s liquidity factors, for the He et al. (2017)'s option portfolios; Boons et al. (2020)'s macro factors, for the ME/MOM double-sorted Fama-French 25 portfolios; Ardia et al. (2022) climate change factor, for the He et al. (2017)'s US corporate bonds portfolios; Chen et al. (2023a)'s sentiment factors, for the ME/INV double-sorted Fama-French 25 portfolios.

Finally, in the attempt to globally assess the non-traded factor zoo, we use a balanced panel of 37 nontradables to construct a five-factor model by means of PCA. Most of the extracted factors correlates significantly with some of the Fama and French (2015)'s and Lettau and Pelger (2020)'s factors, suggesting they might capture similar exposures. Vice versa, these do not correlate with Giglio and Xiu (2021)'s factors, which hints that the latter seems not to capture the risk dimensions associated to the nontradables.
Parallel to Bryzgalova et al. (2023), we find positive evidence about the pricing abilities of such PCs with respect to most of the equity portfolios, which supports the intuition that the non-traded factors indeed captures some relevant risk dimensions.

2 Literature Review

This section provides a brief overview of the two main strands of literature we add to. First, we contribute to the large body of asset pricing literature that aims at identifying the risk factors, in particular to the ones that are considered to be non-traded in the financial markets. Second, we add to the recent literature whose goal is to "*tame the factor zoo*", by critically reexamine the existing findings in the empirical asset pricing literature.

2.1 The Nontradables

Intuitively, a tradable (or traded) risk factor proxies a source of systematic risk which can be traded in the financial markets. In other words, there exists a set of traded assets through which the risk factor exactly prices such systematic risk. Examples are the

[5]The model specifications that are not considered rank deficient by Chen and Fang (2019)'s test, and priced by at least two misspecification-robust t-test based on GLS/HJ/CU-GMM estimates (refer Section 3).

renowned Fama and French (1993)'s three factors, Carhart (1997)'s momentum factor, and in general, factors associated either to firm characteristics or common sources of risk (e.g., portfolio excess returns or return spreads, see Harvey et al. (2016), Barillas and Shanken (2018)).

Inversely, a nontraded/nontradable factor is a risk factor that cannot be traded.[6] The literature commonly hints at them as macro or economic-based factors (e.g., see Balduzzi and Robotti (2010), Kleibergen and Zhan (2018)). Examples are garbage growth, industrial production, and labor tightness.

Yet, the class of macroeconomic or economic factors do not exhaust the class of nontradable factors. In fact, parallel to the tradables, the nontradables might often constitute an equally valid alternative, as capturing the same risk dimensions. For instance, He et al. (2017) proposes a proxy of financial intermediary risk in both nontraded (*intermediary capital ratio*) and traded (*intermediary value-weighted investment return*) terms. Theoretically, since "*many theoretical models are silent on what assets are traded in equilibrium [...] some priced factors may not be reflected in many of the assets that are traded.*"(Giglio et al. (2022)), we cannot draw an exact line between tradables and nontradables.

While *ex ante* unclear, through the lens of a defined asset pricing model, the distinction between these two classes becomes more apparent. Under a (correct) asset pricing model, if the factor is tradable, then its risk premia concides with the expected excess returns on the (exact mimicking) portfolio, or in other words, with its time-series expected value. Vice versa, if nontradable, without further theoretical stance on the pricing, multiple pricing kernels can potentially define the pricing of such risk dimension. In this case, it is then common practice to replace it by its associated maximally correlated mimicking portfolios, which consists in the linear projection of the nontraded factors on a set of base assets that span the asset space (e.g., Vassalou (2003), Adrian et al. (2014)).[7]

Although clearer, the separation is conditional on the model: it depends eventually on the validity of the asset pricing model (i.e., its correct specification). In other words, the presence of any mispricing can invalidate any conclusive classification between traded and nontraded, thus making the previous definitions *de facto* impractical.

For the sake of a sharper distinction that meets practical needs, and light of the ne-

[6]We maintain an agnostic position about why such risk is not traded (market incompleteness, market frictions, theoretical-based model,..)

[7]Because economic risk factors are encoding key insight about the general state of the economy but they might not be perfectly correlated with the asset price movements (e.g., Bai and Ng (2006)), conventionally one appeals to footnote 7 in Breeden (1979). For the proper definitions of the aforementioned objects, please refer to Section 3. Additionally, see Kleibergen and Zhan (2018).

cessity to discipline the factor zoo, we offer a classification of the existing nontraded risk factors, based on the criterion that a nontraded factor is not derived directly as function of weighted asset returns or portfolios.

We recognize they capture 6 risk dimensions: News, Sentiment, Consumption[8], Macroeconomics, Intermediary and Aggregate Firm-level Risk. We provide a summary of the classification in Section 4.1. Appendix B contains the full list of the nontraded factors, with their associated description and reference. Since the nontradables have a coherent picture in terms of statistical properties, following Harvey (2017), we refer to it as the Nontraded Factor Zoo.

Two works formalize econometrically the issue of the nontraded factors: Balduzzi and Robotti (2008) and Balduzzi and Robotti (2010). The first one considers the two alternative formulation of the beta model with non-traded (traditional and mimicking portfolios) and studies their theoretical properties and eventually when they are equivalent. The second one offers a theoretical decomposition of the traditional risk premia into three components: the premium on the mimicking portfolio, the covariance of the non-traded components between pricing kernel and factors, and the mispricing of the mimicking portfolio. Lastly, in their empirical analysis, their findings indicate the fragility of the traditional estimates of the risk premia, as they tend to be exhibit large biases and standard errors.

After the seminal work of Kan and Zhang (1999), the first work that presents evidence of the link between nontraded factors and the issue of weak identification is Kleibergen (2009). The author proposes several statistics that are robust to weak identification and to large-N assets, and reexamines the macroeconomic factors proposed by Jagannathan and Wang (1996) and Lettau and Ludvigson (2001), in their respective CAPM model specifications. In particular, while they find little support for the human CAPM, they find no evidence for the consumption CAPM, despite unable to reject it either.

Proposing a formal methodology to study pricing model performance (based on the cross-sectional R^2), Kan et al. (2013) finds that misspecification-robust standard errors are substantially larger for nontradables (Lettau and Ludvigson (2001), Parker and Julliard (2005), Yogo (2006)), pointing at an association between macroeconomic nontraded factors and (stronger) misspecification problems. Along these lines, Gospodinov et al. (2014) finds sharper conclusions: the durable and non durable consumption factors (Yogo (2006)) and the *cay* factor (Lettau and Ludvigson (2001)) tend to be unpriced, once taken into account model misspecification and weak identification. Similarly, Gospodinov et al. (2019)

[8] We list the consumption factors in a separate category due to the diverse proxies of consumption and the attempt to test if to what extends the consumption-based model empirically holds.

carries on an analogous analysis in the maximum likelihood-based context, confirming the spurious nature of their pricing. Notably, in the context of revisiting the validity of the nontraded factors (and their associated model specification) with respect to the traditional market factor, we refer also to Gospodinov and Robotti (2021a), Gospodinov and Robotti (2021b).

Parallel to Kan et al. (2013), Kleibergen and Zhan (2015) derives the asymptotic distribution of the cross-sectional R^2 when the risk factors might be weak, and offer a reexamination of the findings of Jagannathan and Wang (1996), Lettau and Ludvigson (2001), Lustig and Van Nieuwerburgh (2005), Santos and Veronesi (2006), Yogo (2006). Their findings suggests that the performance of most of the models is inflated, given the "*considerable unexplained factor structure in the first pass residuals*" (Kleibergen and Zhan (2015)). Restricted to consumption-related CAPM literature, Kleibergen and Zhan (2020) studies the weak identification and misspecification issues of the proposed factors of Parker and Julliard (2005), Jagannathan and Wang (2007), Savov (2011), Kroencke (2017). What they find is that in most cases the hypothesis of weak identification cannot be rejected. Similarly, Kleibergen et al. (2023) revisits Lettau and Ludvigson (2001) and Kroencke (2017), with a focus on the relation between sample sizes and weak identification.

Regarding robust inference for mimicking portfolio approach, Kleibergen and Zhan (2018) proposes a test statistic for the risk premia (on the mimicking portfolio) in the case of weak identification. They revisit Adrian et al. (2014), and provide substantial evidence about the weakness of the proposed leverage factor.

We highlight that we do not innovate this literature by proposing a new methodology, but rather our goal is to extend their analysis at more comprehensive and larger scale since the aforementioned works reach their conclusions for a small number of factors.

2.2 Taming the Factor Zoo

In response to Cochrane (2011), a growing number of recent works in the asset pricing literature has put considerable effort to address the (*traded*) factor zoo, to appraise and reconcile the results about the hundreds of published papers that propose risk factors to explain the cross-section of expected returns.

As first attempt, Harvey et al. (2016) focuses on cataloging the papers and associated factors ("*navigate the zoo*" of 313 published works that study cross-sectional return patterns), and provides some guidance about the appropriate level of the t-statistic for conducting risk premia inference.

In the quest of understanding whether this zoo is a byproduct of data mining, several works study the significance of the zoo of the characteristic-based factors with respect to different strategies (from out-of-sample performance, e.g. McLean and Pontiff (2016), Yan and Zheng (2017) and Linnainmaa and Roberts (2018), to multiple hypothesis testing correction, e.g. Chordia et al. (2020), Harvey and Liu (2020) and Giglio et al. (2021)). Their findings indicate that most of the anomalies might be indeed artifacts of data snooping.

Together with the skepticism regarding the validity of the proposed factors, financial economics faces what seems to be a *replication crisis*. Hou et al. (2020) replicates 452 anomaly variables that are published in the literature and finds that *"most anomalies fail to hold up to currently acceptable standards for empirical finance"*[9], hence casting a doubt on their replicability. On the contrary, Chen and Zimmermann (2021) shows that the almost all of the characteristics-based factor zoo can be successfully reproduced and similarly, Jensen et al. (2021) claims that the majority of asset pricing factors can be replicated and represent significant parts of the tangency portfolio.

While we share a similar aspiration of Harvey et al. (2016) in terms of the nontradables, our objective does not consist in purely replicating/reproducing the factors and the associated results, for the sake of providing correction for inference, or in general testing for data snooping.
Our approach to replicate the studies is rather focused on scientific replication[10]: while aiming to match the proposed factors and model specifications (RHS), we assess them using a wide range of commonly used portfolios, which spans equity and non-equity portfolios. In fact, to conclude that a nontradable is unpriced, that is not associated ex ante with any pecularia classes of assets, it requires to conduct robust inference on whether it is an anomaly for the universe of asset returns, so with respect to any potential asset classes.

Contrary to Feng et al. (2020), which proposes a LASSO-based model selection methodology to test whether a proposed risk factor adds explanatory power beyond the existing (*traded*) factor zoo, our main evaluation is by individual model. Albeit several papers examines cross-sectional properties of the traded factor zoo (Avramov et al. (2021), Dong et al. (2022), Engelberg et al. (2023), among others), only few works conduct robust inference on both tradables and nontradables. From a Bayesian perspective, Bryzgalova et al. (2023) conducts robust inference on the risk premia analysing 2.25 quadrillion models that stem from 51 factors, and finds that the latent SDF is likely to contain 23 to 25

[9] In numbers, they find that 65% are not significant with respect to the usual t-value of 1.96

[10] See Hamermesh (2007) for the distinction between pure and scientific replication. Alternatively, refer to Welch et al. (2019)

observable factors, both traded and nontraded. Another eminent example is Kleibergen and Zhan (2022), where they propose a new statistic robust to misspecification and weak identification for infer the risk premia, and revisit 6 nontradables and the factor zoo of Feng et al. (2020).[11] Lastly, Zhang et al. (2021) investigates the difference between tradable and nontradable pricing by comparing the Hansen–Jagannathan (HJ) distance.

We differentiate from the aforementioned both in terms of scale and number of robust statistics employed for conducting inference in the three relevant instances of the econometrics of asset-pricing models: identification, misspecification, and pricing. In doing so, we remark that in our benchmark (scientific) replication analysis we do not consider all possible grouping of specifications, because it would represent a departure from the reexamination of the results of the published literature[12].

3 Robust Inference and Test Procedures

This section presents a concise overview of the main statistics employed in our analysis for conducting robust inference on the parameters of interest. The scope is limited to briefly explain the rationale behind such testing strategies, together with a focus on the main assumptions and asymptotics. Hence, for further details, we ask the reader to directly refer to the papers.

3.1 The Linear Asset-pricing Model

Let us consider the data $\{r_t, f_t\}_{t=1,..,T}$, with r_t a $N \times 1$ vector of excess asset returns (LHS variables), and f_t a $K \times 1$ vector of proposed risk factors (RHS variables). A linear factor-pricing model (or beta model) is formulated as follows:

$$r_t = \alpha + \beta\gamma + \beta(f_t - \mathbb{E}\{f_t\}) + u_t \tag{1}$$

where the object of interest, γ, is the $K \times 1$ vector of risk premia, while β is a $N \times K$ matrix of factor exposures, and u_t is an $N \times 1$ vector of zero-mean idiosyncratic errors,

[11] Those are Jagannathan and Wang (1996), Yogo (2006), Lettau and Ludvigson (2001), Savov (2011), Adrian et al. (2014), Kroencke (2017), He et al. (2017).

[12] For instance, Kleibergen and Zhan (2022) consider all possible specifications with 6 factors drawn from a zoo of 150 factors, meaning 15 billions models. This clearly diverges from the original economic/financial-motivated model specifications of the published papers.

uncorrelated with the factors. The model is said to be correctly specified if it holds with $\alpha = 0_N$, or equivalently when $\mathbf{E}\{r_t\} = \beta\gamma$.

Traditional methodology to estimate the risk premia is the two-pass procedure (Fama and MacBeth (1973), Shanken (1992)): the first step is a time-series regression of excess returns onto a constant and the factors; the second step consists in a cross-sectional regression of the time-average excess returns on the estimated risk exposures (produced by the former step).

Following the Stochastic Discount Factor (SDF) approach, we write equivalent to eq.(1):

$$\mathbb{E}[r_t[1 \;\; (f_t - \mathbb{E}\{f_t\})']\lambda - \mathbf{1}_N] = \alpha \tag{2}$$

where we have $\lambda_0 = \gamma_0$ and $\lambda_1 = Var[f_t]^{-1}\gamma_1$. Because of this formulation (asset-pricing restrictions as moments), the typical estimation strategies in this literature are based on either GMM or maximum likelihood.

As aforementioned, common in the literature related to the nontradables, the mimicking portfolio approach consists in considering the pricing the maximum correlation portfolios rather the actual factors. In other words, instead of using the nontradables directly in the asset-pricing model, the standard is to use as risk factor the time-series linear projection of these nontradables on a set of base assets that span the asset space (i.e., the maximum correlation portfolios). The analysis then remains unchanged.

Among the many underlying assumptions that guarantees the reliability of the existing estimators of the risk premia in eq.(1) or (2), two are crucial in our analysis: the full rank of the betas/exposures (i.e., $rk(\beta) = K$), and the correct specification of the model (i.e., $\alpha = 0_N$).

In our analysis, for checking the strength of identification ($\mathcal{H}_0^{rk} : rk(\beta) = K - 1$), we employ two sets of test statistics, proposed by Kleibergen and Paap (2006), Gospodinov et al. (2017), and Chen and Fang (2019).

For testing whether the model is correctly specified ($\mathcal{H}_0^{spec} : \alpha = 0_N$), we use the conventional Sargan-Hansen J-test and HJ-distance test (Hansen and Jagannathan (1997)).

Finally, for testing the pricing performance of the model ($\mathcal{H}_0^{rp} : \gamma = \gamma_0$), we use the statistics proposed by Kleibergen (2009), Kleibergen and Zhan (2020), Kan et al. (2013), Gospodinov et al. (2014), Burnside (2011), Giglio et al. (2022), Kleibergen and Zhan (2018).

For what concerns the standard methodologies for testing $\mathcal{H}_0^{rp} : \gamma = \gamma_0$, we refer to "FM" as the t-stat associated to the Fama and MacBeth (1973)'s estimation strategy (OLS), and to "Shanken EIV gen" and "Shanken EIV gen gls" as the generalized t-stat associated to Shanken (1992) (respectively, OLS and GLS).

3.1.1 Kleibergen and Paap (2006) and Gospodinov et al. (2017)

Starting by assuming a $\sqrt{T}-$asymptotically normal estimator of the (unrestricted) matrix, Kleibergen and Paap (2006) shows that, after applying a decomposing of the matrix, the smallest singular value has a normal limiting distribution. Exploiting this fact, they propose a rank statistic based on the quadratic form of this orthogonal decomposition.

Studying the problem of spurious factors in linear asset-pricing models, Gospodinov et al. (2017) characterizes the limiting behavior of the specification test based on the Continuously Updated GMM (CU-GMM) estimator when the derivate matrix of the moment conditions is rank deficient. In other words, they propose a rank test *à la* Cragg and Donald (1997), deriving distribution and conservative bounds.[13] The finite-sample version of the test (i.e., the F-stat) is parallel to the one in Kleibergen and Zhan (2020) and Kleibergen et al. (2023). They assume that the time series $\{r_t, f_t\}_{t=1,..,T}$ is jointly stationary and ergodic with positive definite variance and finite fourth moment, which is typical of CU-GMM literature.

We refer to the first one as "KP test", and the other two as "CD rank test" and "F test". The statistics are testing the null hypothesis $\mathcal{H}_0^{rk} : rk(\beta) = K - 1$.

3.1.2 Chen and Fang (2019)

Recently, Chen and Fang (2019) propose a new methodology to test the rank of a matrix, whose focus is shifted to hypotheses on its eigenvalues rather than on its integer-value rank. Their approach overcomes the problem that, in some settings (e.g., heteroskedasticity), the limiting distributions of the standard Kleibergen and Paap (2006)'s test statistics under the null may not stochastically dominate those that are considering lower rank than the hypothesised one. Given the nonstardard asymptotic distribution of the statistic under the null, their procedure is via bootstrap.

We refer to it as "ChenFang" or CF. The statistics test the null hypothesis $\mathcal{H}_0^{rk} : \phi_r(\beta) = 0$, where $\phi_r(\cdot)$ is the sum of $k - r$ the smallest squared singular values of β.

[13]The behavior depends on the dimension of the null space. They characterize the distribution when the null space is unidimensional, while the upper bound when the null space has dimension larger than one. In particular, it is important to realize that when the null space is of dimension 1, we possibly have two cases: the model is correctly specified and identified vs. the model is misspecified with a spurious factor.

3.1.3 Kleibergen (2009) and Kleibergen and Zhan (2020)

Kleibergen (2009) and Kleibergen and Zhan (2020) share analogous settings. Under the assumption of i.i.d. error terms (along the time-series) and usual finite variance conditions (Shanken (1992)), they propose a variety of statistics that are testing the pricing of the factors, which are robust to repackaging[14] and to all possible value of the betas.[15] Among the many, Kleibergen (2009) proposes a statistic *à la Anderson and Rubin (1949)*, the Factor AR statistic (FAR).

To avoid pre-test bias, Kleibergen and Zhan (2020) introduces an extension of the GRS-FAR statistic, which is a (joint) test on the risk premia, assuming the model is correctly specified. Peculiarly, the confidence sets need to be derived from inverting the test. Such confidence sets might take three different forms: 1) the confidence intervals are unbounded, which suggests weak identification; 2) the confidence intervals are empty, that implies the model is prone to be misspecified; 3) the confidence intervales are bounded, so the risk premium is identified and model is correctly specified. Only in latter case, if the (bounded) confidence interval does not include zero, then the factors is said to be priced.

We refer to the last one as "GRS-FAR", which tests the null hypothesis $\mathcal{H}_0^{rp} : \gamma = \gamma_0$.

3.1.4 Kan et al. (2013) and Gospodinov et al. (2014)

Kan et al. (2013) and Gospodinov et al. (2014) relax the assumption of correct specification towards allowing for potential global misspecification, or in other words, population deviations from the model (i.e., $\alpha \neq 0$). Standard to the GMM literature (Hansen (1982)), they assume the time series $\{r_t, f_t\}_{t=1,..,T}$ to be jointly stationary and ergodic, with positive definite variance and finite fourth moment.

In these settings, Kan et al. (2013) derive the asymptotically normal distribution of the sample cross-sectional R^2 (under three scenarios: $\{R^2 = 1\}$, $\{0 < R^2 < 1\}$, $\{R^2 = 0\}$), and of the t-statistic associated to the risk premia estimates. In particular, they show that the asymptotic distribution of the latter does not depend only on the asymptotic variance and the errors-in-variables adjustment (Shanken (1992)), but also on a misspecification term. Once adjusting for it, the statistics are then robust to potential misspecification.[16]

[14]"*[They] assume that the moment equation also applies to the returns on any repacked portfolio of assets*" (Kleibergen and Zhan (2020)). Thus, it allows to consider a model with or without the zero-beta return.

[15]The main intuition is that the misspricing errors (i.e., the α) and the betas (i.e., the β) are independently distributed when the sample size becomes large $T \to \infty$.

[16]While it is more explicit the robustness to misspecification of the t-stat, for the other set of statistic they claim that: "*Although it is possible that some of the GMM sample moment conditions are not asymptotically*

Gospodinov et al. (2014) generalizes the previous results for a SDF framework, allowing for the presence of one (or more) useless factor(s). Thus, in this case, they provide a t-stat for testing the risk premia that is robust to both identification failure and potential global misspecification. Also, they derive these results with respect the HJ-distance.

We refer to the first t-stats as "t-statm CSR" and "t-statm CSR GLS", while to the last two t-stats as "t-statm HJ" and "t-statm CU-GMM". The statistics test the null hypothesis $\mathcal{H}_0^{rp} : \gamma = \gamma_0$.

3.1.5 Burnside (2011)

For accounting the potential model misspecification, Burnside (2011) proposes a pairwise bootstrap procedure: by sampling with replacement jointly the factors and the test asset returns, one can generate the bootstrapped distribution of the risk premia. In particular, their approach is designed to account for the generated-regressor issue (i.e, the Shanken (1992)'s correction), and can be easily accommodated for heteroskedasticity by block bootstrap. Moreover, it shows that it is robust against the problem of spurious factors.

We refer to it as "Burnside CI". The statistics tests the null hypothesis $\mathcal{H}_0^{rp} : \gamma = \gamma_0$.

3.1.6 Giglio and Xiu (2021)

Differently to the previous framework, Giglio and Xiu (2021) presume that the observed excess returns are not priced by the observed factors but rather by unobservables. Hence, bridging the literature of latent factor models and of mimicking portfolios, they propose to price the observed factor via a three-step procedure that estimates first the risk premia associated with the latent asset-pricing factor model and then to project the observed factors onto the latent ones (extracted via PCA), in order to obtain the risk premia associated to the mimicking portfolio. The combination of the two leads to the final estimator of the risk premia of the observed factor. The inference of the risk premia is claimed to be robust with respect to the issues of spurious, omitted, or bad-proxy[17] factors.

The results are derived under assumptions *à la* Bai (2003) (up to some relaxations), which can be summarized as: i) the latent factors are pricing correctly the assets, ii) the pricing factors are strong and pervasive, and finally the asymptotic is both in N and T (i.e., $N, T \to \infty,\ N/T \to c \in \mathbb{R}^+$).

normally distributed (see Gospodinov et al. (2012) for details), our results on the asymptotic distribution of [the R^2] are not affected by this problem" (Kan et al. (2013)).

[17]Factors that are proxy and their measurement error components is not negligible.

We refer to the their t-stat as "GX", which tests the null hypothesis $\mathcal{H}_0^{rp} : \gamma = \gamma_0$.

3.1.7 Kleibergen and Zhan (2018)

Under similar settings as in Kleibergen and Zhan (2020), Kleibergen and Zhan (2018) propose a novel test for the risk premia when employing the mimicking portfolio approach to price a factor. Their statistic is designed to be robust to issues of weak identification. Likewise the GRS-FAR statistic, this statistic construct the $100 \times (1 - \alpha)\%$ confidence intervals for the risk premia by performing a grid search for the hypothesized value of risk premium with values at which the null cannot be rejected.

Since the mimicking portfolio approach requires to select the spanning portfolios, to avoid eventual biases due to such choice we remain agnostic by following the intuition of Giglio and Xiu (2021): instead of hand picking the test assets, we consider the spanning portfolios with respect to a class of test assets to be some latent factors (or rather portfolios) that we estimate from the asset returns via PCA (more details in Section 4.3).

4 Data and Model specifications

In this section, we first provide a brief description of our data: the non-traded factor zoo (right-hand-side/RHS variables) and test portfolios (left-hand-side/LHS variables) at different frequencies (monthly, quarterly, annual). Secondly, we outline how we specify the different models and provide a summary of our procedure.

4.1 The Nontraded Factor Zoo

We collect 104 distinct non-traded factors from 52 papers published in top economic, finance, and management academic journals in the last 40 years (from July 1986 to February 2023): *The Journal of Finance* (19), *The Journal of Financial Economics* (19), *Journal of Political Economy* (4), *Journal of Financial and Quantitative Analysis* (2), *Review of Financial Studies* (2), *American Economic Review* (1), *Journal of Business* (1), *Journal of Economic Dynamics and Control* (1), *Journal of Monetary Economics* (1), *Management Science* (1), *Review of Accounting Studies* (1).

The factors are grouped into 6 categories based on the risk dimensions the factors are measuring: News (11), Sentiment (17), Liquidity (16), Consumption (25), Macroeconomics (28), Intermediary and aggregate firm-level risk (7).

Table 1-6 in the Appendix B tabulate the factors with short name, brief description, and source in the academic literature. Similarly to Harvey et al. (2016), our focus is restricted to published works that propose new factors. Up to our knowledge, our nontraded factor zoo represents the first attempt to classify the nontradables in terms of breadth.

While we do not include works that studied the same nontradables in different contexts (with respect to assets and/or time periods), we include papers that propose different empirical proxies for the same type of economic risk. In fact, it is the latter that discloses which risk dimension is likely to be fundamental for the systematic risk.
Also, we do not narrow our attention on factors that have been exclusively proposed to price the cross-section of expected returns: out of 52 papers, we include 11 studies whose analysis does not explicitly price the cross-section of asset returns. We include them in the zoo since those nontraded factors, deemed to be useful in the time-series dimension, could be eventually used for cross-sectional purposes.

Regarding the nontradables that are not publicly shared, we replicate them using macroeconomic series from FRED. In particular, we used CPI, labor income (as the difference between total US income and dividend income), Personal Consumption Expenditures (PCE), Personal Consumption Expenditure on Nondurable Goods (PCEND), monetary aggregates M2 and M3 (seasonally adjusted and unadjusted), total US population (Seasonally Adjusted), Industrial Production. We also use the Producer Price Index (Crude Petroleum series), University of Michigan Inflation Expectation, and Real Gross National Product/Gross Domestic Product (RGDP) from Survey of Professional Forecasters of Federal Reserve Bank of Philadelphia. Also, from Amit Goyal's website, we employ long-term government bond yield, dividends on S&P500 index, term spread and default spread. Overall, we include 66 factors in monthly, 28 in quarterly and 14 in annual frequency. Lastly, the additional control factor MKT and risk-free rate are from Kenneth French's website.

4.2 The Test Portfolios

By definition, the nontraded factors are not associated ex ante to a particular class of assets, and Kan et al. (2013) point out that the model performance can vary with respect to the test assets when misspecified. In this paper, we consider 13 sets of test portfolios[18] covering 7 major asset classes as representative benchmarks to individually test the cross-sectional pricing ability of nontraded zoo at different frequencies[19] (monthly, quarterly, annual).

Regarding equity portfolios, we use seven groups of portfolios[20]: 1) 25 Fama-French size and book-to-market portfolios (FF25 ME/BM), 2) 25 Fama-French book-to-market and operating profitability portfolios (FF25 BEME/OP), 3) 25 Fama-French size and investment portfolios (FF25 ME/INV), 4) 25 Fama-French size and momentum portfolios (FF25 ME/MOM), 5) 24 portfolios formed with the bottom and top portfolios of 11 US anomalies [21] listed on Kenneth French's library as well as two long-short portfolios of "quality-minus-junk" (Asness et al. (2019)) and "betting-against-beta" (Frazzini and Pedersen (2014)) from AQR's data library[22]; 6) 48 portfolios formed with the bottom and top portfolios of 24 anomalies [23]studied by Hou et al. (2020); 7) 10 equity portfolios sorted on cash-flow duration used by Weber (2018).

Regarding non-equity portfolios, following Lettau et al. (2014), He et al. (2017) and Lettau et al. (2019), we consider: 1) 20 US bond portfolios [24] spanning Jan. 1975 to Dec. 2011; 2) 6 sovereign bonds from Jan. 1995 to April 2011; 3) 18 S&P 500 index option portfolios from April 1986 to Jan. 2012; 4) 12 foreign exchange portfolios from March 1976 to Jan. 2010; 5) 23 commodity portfolios from Sept. 1986 to Dec. 2012; 6) 20 CDS portfolios from Feb. 2001 to Dec. 2012.

[18]We use excess portfolio returns in the tests except Kleibergen and Zhan (2020) and Gospodinov et al. (2014) misspecification-robust statistics which use gross returns.

[19]For all test portfolios used, we compound the monthly returns to match the sample of factors at the quarterly and annual frequencies.

[20]For more details, please see the Appendix A.1 of Bandi et al. (2021).

[21]The 11 anomalies are size, value, profitability, investment, accruals, cash flow to price, earnings to price, long-term reversal, net share issuance, residual variance and short-term reversal.

[22]Data is available at https://www.aqr.com/Insights/Datasets

[23]The 24 anomalies are book-to-market equity (bm), cash-flow-to-price (cp), enterprise multiple (em), earnings-to-price (ep), long-term reversal (rev_1) and sales-to-price (sp), operating accruals (oa), composite equity issuance (cei), discretionary accruals (dac), net operating assets (noa), change in net operating assets (dnoa), change in PPE and inventory-to-assets (dpia), investment-to-assets (ia), investment growth (ig), and net stock issues (nsi), return on equity (roe_6), change in the return on equity (droe_6) and operating profits-to-book equity (ope), organizational capital-to-assets (oca), market equity (me), market beta (beta_1), idiosyncratic volatility (ivff_1), short-term reversal (srev) and total volatility (tv_1). Data is available at http://global-q.org/testingportfolios.html

[24]10 maturity-sorted government bond portfolios from CRSP Fama Bond Portfolios file and 10 corporate bond portfolios sorted on yield spread.

4.3 Replicating the Model Specifications

In conducting robust inference, our first priority is to scientifically replicate the proposed model specifications (RHS) in the original papers. Hence, we refer to "Original Models" as the batch of model specifications that share the same RHS variables with the respective published works. For the papers that run cross-sectional analysis at individual-asset level, we include the familiar factors that proxy for the anomalies/characteristics used in the paper. We collect 139 models that are originally used in the papers.

As the first stage, we analyze the pervasiveness of the problems that might jeopardize the validity of the risk premia estimates in the original model specifications. Consequently, we conduct inferences that are robust to these problems in order to get reliable estimates of the risk premia.

In the second stage, we hand pick the nontraded factors and run two different model setups: nontradables only, and nontradables controlling for the market factor (MKT). The purpose of this first part of the analysis is to understand if the issue of weak identification starts before the original specification of the model, and so is a distinctive statistical property of the proposed factor. Secondly, i n a similar spirit of Gospodinov and Robotti (2021a), we then test whether the nontraded factors capture the same linear information entailed by the market factor. A visual example is provided in Figure 1.

As mentioned in the previous sections, one way to conduct inference on the nontradables is via mimicking portfolios. Necessary step for implementing the mimicking portfolio approach, involves the selection of the spanning test assets. In light of the intuition of Giglio and Xiu (2021), we consider two sets of spanning portfolio: 1) the Principal Component extracted from the pool of either the equity portfolios or the nonequity portfolios[25] 2) the factors proposed by Giglio and Xiu (2021) (GX), and Lettau and Pelger (2020) (RP-PC, SMV).

[25]In particular, we are considering the first 7 PCs for monthly frequency, the first 5 PCs for quarterly frequency, and the first 3 PCs for annual frequency. The decreasing number of latent components is due the precision in terms of number of observations.

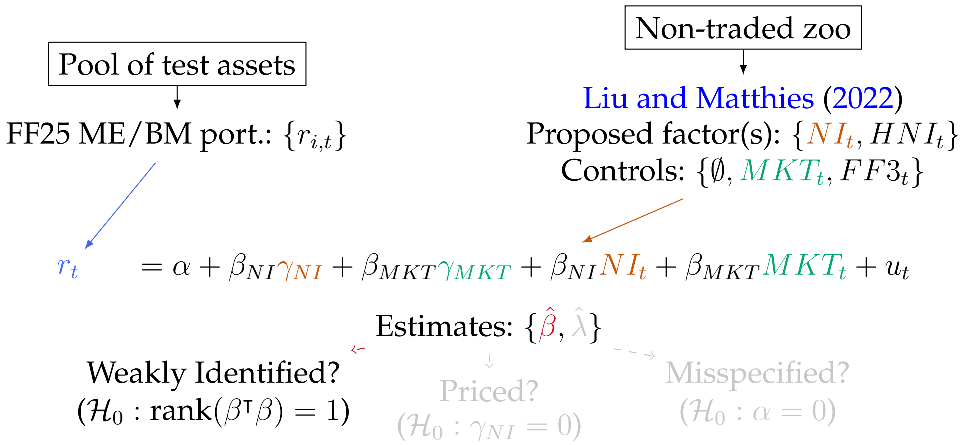


Figure 1: A visual example of the evaluation of the model specifications. In this specific case, we hand pick the News Index factor (orange), NI_t, proposed by Liu and Matthies (2022) (see their table IV) and testing for weak identification controlling by the market factor (green), MKT_t, with respect the FF25 ME/BM portfolios (blue). In Other hypotheses of interest are whether the model is correctly specified, and if the News Index factor is pricing the cross-section of the FF25 ME/BM portfolios (grey).

5 Empirical Results

This section provides the results on the statistical properties of the original model specifications in terms of identification, misspecification and pricing, by means of the robust statistics described in Section 3. We also assess the identification strength in the single-factor models, and in the specifications where the market factor is added as control.

We offer our results across classes of test assets, factor categories, and time series sample size to distinguish between potentially different DGPs and between finite and large samples, thus accounting for the different asymptotics of the employed statistics.

We present some results in terms of identification strength for the benchmark models, namely CAPM and Fama and French (1993)'s three-factor model in the Appendix A.

5.1 Weak Identification and Spuriousness

Conventional inference procedures are established on the implicit assumptions that the model specification is well-identified. As already pointed out (Section 2.1), their validity can be jeopardized in the case of spurious or weak factors, and further distorted by the degree of model misspecification.

Despite these potential concerns, we observe that a vast majority of empirical research often seems to ignore these problems, thus implicitly conducting fragile inference on the pricing of the proposed factors. This is especially relevant for the non-traded factors, which tend to have lower correlations with test assets than traded factors. This section is thus providing the results of testing the identification strength of the model specifications associated to the nontraded factors.

5.1.1 Original models

In this section, we discuss the issue of identification for the original model specifications presented in the published papers. We collect 139 original models and test them individually with respect to 13 sets of test portfolios.

Table 9 summarizes the percentages of model rank deficiency, i.e. the percentages of model specifications for which we fail to reject the null. The percentages are with respect original model specification per set of test assets (i.e., the total number of models is total number of right-hand side specifications by total sets of test portfolios). As the tests perform differently with respect to asymptotic regimes, to reconcile the results, we present the percentages for two additional scenarios. We refer to "conservative", when all of test statistics can reject the null of rank deficiency (unanimity), and to "favorable" when any of test statistics can reject. In other words, the conservative scenario should signal the upper bound of identification failure, while the favorable one provides the lower bound.[26]

In Panel A in Table 9, according to KP test and the F test, at 5% significance level, we are prone to conclude that 89.78% and 73.74%, respectively, of model specifications are under-identified with the commonly used test portfolios. The CD test and CF test tend to reject more often, suggesting that about 50% of model specifications suffer from rank deficiency. Only 11.47% [27] of model specifications are rejected from the all four rank tests. Given the large number of hypotheses we are testing, we also control for the Type I error via multiple hypothesis correction. As the classical method of familywise error rate (e.g., Bonferroni's) is too strict to retain power of the tests, we rather control the false discovery rate *à la* Benjamini and Hochberg (1995), fixing it at 5% for specifications in each test. After correction, the KP test and the F test suggest that 98.52% and 83% respectively, of the models are prone to suffer from weak identification. The CD test and CF test show a

[26]The upper and lower bounds are obtained by outer and inner joins of the results, thus we do not consider as alternative definitions the minimum and maximum proportion of rejection from each test.

[27]The percentage is calculated by subtracting 88.53%, i.e., the percentage of identification failure.

less severe picture: around half of the models seems to be affected by such issue (55.25% and 51.43%, respectively).
All things considered, the issue of identification failure seems to be very pervasive, and represents a strong signal to be cautious while conducting risk premium inference, as the traditional methodologies are fragile to such problems.

Since the ability of the tests to reject the null of rank deficiency depends on the power of such tests and the power can be correlated with the number of time series observations, we study the impact of sample size on our results. Panel B, C and D of Table 9 separate the percentages for three batches with different length of the sample size: $T \leq 100$ (small), $100 < T \leq 300$ (moderate) and $T > 300$ (large). Figure 6 also illustrates the percentage of identification failure of the original models by rank tests, and with respect to sample sizes. For the extremely short sample case ($T \leq 100$), there is a large difference between the CD test and the other three tests. The former indicates that only about 30% of the specifications fail to reject the null, in stark contrast with the KP test and F test, which indicates that more than 90% of models are rank deficient. In particular, all annual macroeconomics[28] factors and consumption factors tend to be labelled by the KP test, F test and CF test as weakly identified. With moderate length of the time series ($100 < T \leq 300$), 95% of model specifications are deemed as weakly identified for the KP test, whereas about 77% in F test, 52% in CD test and 55% in CF test. Most of such models mainly consist of quarterly factors, and a few of them are models with liquidity factors at monthly frequency. The different between F test and CD test is narrowed down with $T > 300$. The former shows 64.60% rank deficiency and the latter indicates 57.44%. At the two opposites, the KP test is still the least powerful in rejecting the null with 82.53% rank deficiency, while the CF test rejects more frequently, about 39% of model specifications are deemed as weakly identified.

Table 11 displays the proportion of models that tend to suffer from identification failure across different test asset portfolios.
Remarkably, the pervasiveness of such problem is quite similar across the equities and non-equities. The percentages range from the minimum of around 32% with the HXZ anomalies portfolios to the maximum of 73% with the HKM sovereign bonds.[29] For the most popular FF25 ME/BM portfolios, we still observe identification failure in at least 41% of models. Lastly, we report in Table 12 the pervasiveness of weak identification across risk dimensions, or in other words, by non-traded factor category (see Section 4.1). On an aggregate level, we find that the models associated with proxies of liquidity, sentiment

[28]The only exceptions are the Da et al. (2015)'s models with the FEAR index.
[29]The percentages that are discussed are the ones corresponding to the Chen and Fang (2019).

and volatility are less likely to be dubbed as weakly identified. However, if one considers exclusively the CD test and CF test, we notice that the consumption models and macro models (see Table 4-5 in the Appendix) show low percentages in terms of identification failure. Those models tend to be at a quarterly or yearly frequency, and therefore, due to the shorter time series are likely to suffer finite-sample distortions.

5.1.2 Identification with and without the Market Factor

The lack of identification can occur when the factors are uncorrelated of the returns of test assets. However, Gospodinov et al. (2014) point out that it can also result from two or more factors that are noisy proxies of the same underlying factor. In Appendix A, we show that the market factor is generally identified for most of test portfolios, especially at the monthly and quarterly frequencies. Therefore, we examine if the percentage of rank deficiency varies if we add the market factor in the models.[30] Figure 8 shows the change in percentage of identification failure from single-factor models to the models with the same factor plus the market factor. In each of the rank tests, the percentage of lack of identification increases. In CF test and F test with and without intercept, the rank deficiency goes up by 13% after including market factor. There are also around 6% and 9% increase in KP and CD tests respectively compared to the single-factor benchmark analysis. This points towards the conclusion that the linear information entailed in at least the part of non-traded factors is not statistically different from the market factor.

5.2 Model Misspecification

As emphasized in Section 2.1, the literature has already showed that the implicit assumption of correct specification of the model might lead to over-rejections of the asset-pricing models, in favor of the nonzero pricing of the proposed factors. As this effect might play a crucial role, if not exacerbate the distortions of weak identification, in pricing the nontradables, we report in this section the results in terms of testing the absence of mispricing.

Table 9, together with Figure 12, shows what seems to be a contradiction. On one hand,

[30] We examine 53 factors: Annual_Gar, Annual_IMC, Annual_IS_PI, Annual_Q4Con, Attention_PLS, Attention_sPCA, CIV, Dsv, FEARS30, Fin_uncert, HKMcapital, HKMcapital_Old, Illiq_Perm, Illiq_Trans, Inv_sent, Inv_sent_orth, LCT, LTR, LaborTightness, Liq, Liq_supply_shock, M2C_SA, M2C_SU, M3C_SA, M3C_SU, MCCI, Macro_uncert, NVIX, Noise, Quart_CC, Quart_Dlogks, Quart_HIFac, Quart_HKMcapital, Quart_HKMcapital_Old, Quart_PAG, Real_uncert, Svix2_index, Svix2_vw, WSJ_AR1, adj_PC, adj_logamireal, adj_logamito, adj_logcorwin, adj_logfht, adj_loghm, adj_logps, adj_logroll, adj_logtick, adj_logzeros, chneg_AR1, cons_PCEND_2_est, innov_MCCI, unexp_disrisk.

the J test indicates that one generally would conclude that only 13.58% of models are misspecified. For the other large majority of model specifications, one can easily conclude that the model is perfectly specified, thus consider standard procedures (that assume the correct specification) to conduct inference. On the other hand, the misspecification test based on Hansen-Jagannathan distance displays a different picture with a substantially higher percentage of the models (around 70.26%) being misspecified. In particular, Figure 13 shows that the HJ test indicates that more than 90% of the models tend to be misspecified for FF25 ME/INV, HXZ anomalies, FF25 ME/MOM portfolios.

However, three remarks are in order. First, as discussed in Kleibergen and Zhan (2022), the J test suffers from severe distortions in the presence of weak identification. In particular, we find that for more than 80% of the original models the J test and the Anderson Rubin test are close in magnitude, thus questioning the reliability of the J test. Second, we find that the J-test correlates more with the CF test when the model is dubbed as rank deficient. Third, one needs to keep in mind that, as pointed out by Gospodinov et al. (2017), in the presence of spurious factors, one might not be able to distinguish between correct specification and identification failure (see footnote 13).
Figure 14 displays indeed that the models associated with the non-equities portfolios are more likely to be concluded by the J test as correctly specified. For instance, if one considers the FX portfolios, only for 5% of the specifications we reject the null of correct specification with the J test, while the HJ test indicates more than 60% of them are misspecified.

5.3 Robust inference on the Risk premia

In the presence of weak identification and/or misspecification, the standard methodologies to conduct inference on the risk premium tend to be fragile. Given the previous conclusions about the ubiquity of such issues, we apply a battery of robust statistics to conduct robust inference on the risk premia, to investigate the validity of the conclusions on the risk premia of the non-traded factors. Additionally, we study whether using the mimicking portfolio procedure can lead to a different picture. Lastly, we analyze the sensitivity of the results with respect to the zero-beta rate.

First and foremost, we find none of non-traded factors in the original model specifications has universal nonzero pricing ability across test assets.

Regarding equities, Figure 15 presents the number of nontraded factors with non-zero price of risk.[31] Employing standard FM approach, one can conclude that a vast majority of

[31]We consider a factor to be priced when has a non-zero risk premium in at least one of the originally

factors can explain the cross-sectional variation in the equity portfolios: (out of 137)[32] from 58 to 99 nontraded factors are priced in the FF25 ME/MOM and ME/INV portfolios, respectively. Over 50 factors are suggested to be priced in the HXZ portfolios, US anomalies portfolios and the Weber's 10 duration portfolios. However, when we apply the robust methodologies, the proportion of factors with statistically non-zero risk premium drastically reduces.[33] Marginally, we observe more priced factors for HXZ anomalies portfolios, US anomalies portfolio, and Weber's portfolios than for the Fama French portfolios, which are the most used equity portfolios in empirical asset pricing. Accounting for multiple hypothesis testing, the percentage of non-zero risk premia almost disappears, except in HXZ anomalies portfolios, with only 10 factors priced with respect to the t-statm CSR GLS.

Regarding the non-equity portfolios, similar to equity portfolios, the standard FM procedure over-rejects: 113 (out of 137) nontraded factors are priced in the option portfolios, and about 90 in the CDS, FX and US bonds portfolios. A different picture appears when looking at the robust t-test, where a large number of nontraded factors are unpriced. Robust statistics suggest only $16-30$ priced factors for CDS, $2-15$ for commodity, $1-14$ for FX, $8-17$ for options and $7-22$ for US bonds. It is worth mentioning that, for sovereign bonds, the t-statm CSR GLS, t-statm HJ, and GX all indicate that none of the non-traded factors are priced. After accounting for the multiple hypothesis correction, the significance of the nontraded factor zoo completely disappears. Given the large number of model specifications we test in the paper, we only list the factors with significant risk premia suggested by all of t-statm CSR GLS, t-statm HJ, t-statm CU-GMM in the table 13.

Figures 17 and 18 compare percentages across the length of the sample size. Marginally speaking, it seems that factors with larger sample size ($T > 300$) are more likely to be priced in the cross-sections of equity portfolios.

We summarize the nontraded factors that we deem as priced in some cross-sections of assets: Campbell and Vuolteenaho (2004)'s cash flow news factor for US anomalies portfolios, Sadka (2006)'s liquidity factors for US bonds, Bali et al. (2017)'s uncertainty factors for the HXZ anomalies portfolios, Chen et al. (2018)'s liquidity factors for option portfolios, Boons et al. (2020)'s macro factors, for the FF 25 ME/MOM portfolios, Ardia et al.

proposed RHS specifications.

[32] The 137 factors include the lags of some factors.

[33] We include in our benchmark analysis the standard FM t-statistic (FM), the FM t-statistic estimated with GLS (FM GLS), Kan et al. (2013)'s misspecification-robust t-statistic in the CSR (t-statm CSR and t-statm CSR GLS), Giglio and Xiu (2021)'s three-pass approach (GX), Gospodinov et al. (2014)'s misspecification-robust t-statistic based on Hansen-Jagannathan distance (t-statm HJ) , Gospodinov et al. (2017)'s misspecification-robust t-statistic in linear SDF estimated using continuous updating GMM (t-statm CU-GMM), and Burnside (2011)'s bootstrap approach (Burnside CI).

(2022) climate change factor, for the He et al. (2017)'s US corporate bonds portfolios. This summary is based on the identification tests and the consensus of three misspecification-robust risk premia t-tests (t-statm CSR GLS, t-statm HJ and t-statm CU-GMM).

In addition to the robust test statistics already discussed, we carry on the same analysis with respect to the Mimicking Portfolio Anderson-Rubin (MPAR) statistic in Kleibergen and Zhan (2018) and GRS Factor Anderson-Rubin (GRS-FAR) statistic in Kleibergen and Zhan (2020). Figure 19 and 21 present the number of priced nontraded factor for the equity portfolios and the non-equity portfolios, respectively. For the mimicking portfolio approach we use the following sets of base assets: 1) the principal components extracted from the equity portfolios as well as 2) the ones extracted from non-equity portfolios, in the style of Giglio and Xiu (2021); 3) the 7 latent factors of Giglio and Xiu (2021); 4) the principal components extracted from extreme deciles of 37 characteristic-based anomalies via RP-PCA and 5) sparse mean-variance PCA.

We find that, for the GRS-FAR test, only 13 models have a priced nontraded factor for FF25 BM/OP, and 10 for FF25 ME/BM, 6 for HXZ anomalies portfolios, and less than 5 for the remaining equity portfolios. For non-equity portfolios, less than 15 factors are identified and non-zero risk premium for CDS, commodity and sovereign bonds and less than 5 factors in the rest of non-equity portfolios. With MPAR approach, no more than 10 nontraded factors are deemed to be priced, regardless of selection of the spanning test assets and classes of assets. The vast majority of original models lead to unbounded confidence interval for the nontraded factor risk premia.

Lastly, we explore the impact of restricting the zero-beta rate equal to risk-free rate, or in other words, the sensitivity of the risk premia estimates to not include the intercept in the cross-sectional regression. Indeed, as Kleibergen and Zhan (2020) and Kroencke and Thimme (2021) point out, excluding intercept may improve identification (conditional on the correct specification). Figure 23 compares the percentages of nonzero risk premia for all test portfolios in the two scenarios, i.e., with and without intercept. Figure 24 shows the previous comparison once taking into account multiple hypothesis testing. We observe that removing the intercept makes higher proportion of factors be priced (see the t-statm CSR, with about 90% of the priced factors). However, this must be understood rather as a cautionary tale: we argue that, when there is little variation in the beta estimates, excluding the intercept may wrongly attribute the explanatory power to the factor exposures. Therefore, it is essential to include intercept when estimating the risk premium, unless the model is correctly specified.

5.4 The Pricing of the Zoo of Nontradables

While our analysis in the previous sections is at individual model specification level, this section concerns rather the evaluation of the pricing of nontradables at aggregate level. In particular, the focus is for equity portfolios at monthly frequency.

We study the explanatory power of the non-traded factor universe with respect to the cross-section of test portfolios by means of Giglio and Xiu (2021)'s methodology. First, we project each non-traded factor to the mimicking portfolios (i.e., the first 7 PCs extracted from the pool of equity portfolios), whose construction is motivated in Section 4.3. In order to have a balanced panel, we consider 37 nontradables, spanning from Jan. 1970 to Dec. 2015, at monthly frequency. Second, in light of Giglio and Xiu (2021), we summarize the linear information entailed by these projected portfolios by extracting the first five principal components. It is important to realize that extracting the principal components without projecting them firstly might be misleading in economic terms, as the nontraded factors are not (weighted) portfolios, and indeed a linear combination of such factors does not have a direct economic connotation.

To understand what linear information is summarized in those PCs from the non-traded factor zoo, Table 14 shows the correlation with three popular pricing (traded) factor models: Fama and French (2015)'s five factors, Giglio and Xiu (2021)'s seven factors (GX1-7) and Lettau and Pelger (2020)'s five factors (RP-PC1-5).

We notice that, while the PCs of nontradables do correlate with the Fama and French (2015)'s and Lettau and Pelger (2020)'s factors, they do not correlate much with Giglio and Xiu (2021)'s, which implies that they are not capturing the risk dimensions that are captured by the non-traded factor zoo. This needs to be read as indirect evidence on why Giglio and Xiu (2021)'s findings about the zero pricing of some nontradables. Generally speaking, we conclude that: 1) the first PC ($eqPC1$) correlates mildly with RP-PC2; 2) the second PC ($eqPC2$) correlates strongly with the MKT and RP-PC1; 3) the third PC ($eqPC3$) correlates strongly with RP-PC3, and mildly with SMB and HML; 4) the last PC ($eqPC5$) correlates strongly with SMB and the RP-PC4. Interestingly, this partially aligns with Lettau and Pelger (2020), where they shows that their proposed latent factors capture some relevant risk dimensions: RP-PC1, RP-PC3, and RP-PC5 are highly correlated, respectively, with the CRSP value-weighted market return, the value factor, and reversal portfolios. RP-PC2 correlates mildly with a number of anomalies, and RP-PC4 is mainly driven by momentum, value interaction and trading frictions portfolios. Lastly, regarding the fourth PC ($eqPC4$), it seems that is not strongly collinear with any of the aforemen-

tioned traded factors.
For what concerns the loading with respect to the 43 non-traded factors, those weights seem to remain similar across nontradables. The only exceptions are $eqPC4$ and $eqPC5$, which place considerably more weights on sentiment, consumption and macroeconomic factors.[34]

Table 15 displays the model diagnosis tests regarding identification and misspecification. As we expect, the identification strength of the model $eqPC1-5$ is stronger in the equity portfolios (especially HXZ anomalies and US anomalies portfolios), rather than the non-equity portfolios.
Table 16 and 17 present the risk premium estimates[35] of the model $eqPC1-5$ calculated from the Fama and MacBeth (1973) procedure, Giglio and Xiu (2021) methodology, and the misspecification-robust CUGMM (Gospodinov et al. (2017)), as well as the Burnside (2016)'s bootstrap confidence interval for the risk premium. We see that, across different equity portfolios, $eqPC1$ tends to be priced, in particular we can uniformly conclude that it is priced in the HXZ anomalies and US anomalies portfolios, while the picture is more grey for the other PCs. However, we highlight that $eqPC4$ is significantly pricing the FF25_ME_MOM portfolios, which becomes relevant especially once considering that the correlation between $eqPC4$ and the Carhart (1997)'s momentum factor is very mild (≈ -0.23), signaling that the two factors are capturing different linear information.

6 Conclusion

After collecting 104 distinct non-traded factors from 52 papers published in top economic and finance academic journals of the last 40 years, we offer a comprehensive investigation of the pricing of the non-traded factor zoo, by conducting inference robust to weak identification and misspecification. We offer the first study to extensively reexamine at this large scope the statistical properties of the nontradables, evaluating the pricing for both equity and non-equity portfolios.

Remarkably, we find that at least 45% of originally proposed models tend to be weakly identified. Only a small minority (around 11.47%) of the models are to be considered identified. Accounting for multiple hypothesis testing, the issue of weak identification becomes more pervasive (KP test even suggests more than 98% of the model suffer lack

[34] The full list is: $IndProdY$, Oil, $MICHI\_inf\_exp$, $LaborGrowth$, $LaborGrowth^2$, $M2C\_SU$, $M3C\_SU$ and $cons\_PCE$. Please refer to Appendix B for the references.

[35] Please note that the risk premia is identified up to a sign because the factors are PCs.

of identification). The problem of misspecification has a similar pervasiveness. We document that the Hansen J-test can deliver misleading conclusion: in our empirical study, the J-test rejects around 15% of models, in stark contrast with the HJ-test. Furthermore, although a large number of nontraded factors are priced by the lens of the standard FM approach, a variety of robust statistics show the opposite picture: very few factors are in fact priced in the test portfolios. Using the original model specifications, we do not find any non-traded factor that has universal pricing ability in all test asset classes.

Providing a constructive appraisal, we list the nontraded factors that are likely pricing some cross-sections of assets: Campbell and Vuolteenaho (2004)'s cash flow news factor for US anomalies portfolios, Sadka (2006)'s liquidity factors for US bonds, Bali et al. (2017)'s uncertainty factors for the Hou et al. (2020)'s anomalies portfolios, Chen et al. (2018)'s liquidity factors for option portfolios, Boons et al. (2020)'s macro factors, for the FF 25 ME/MOM portfolios, Ardia et al. (2022) climate change factor, for the He et al. (2017)'s US corporate bonds portfolios. The list is based by considering a consensus of both the identification and misspecification-robust risk premia tests.

In the attempt to summarize the linear information entailed in the non-traded factor zoo, by means of PCA, we summarize a balanced panel of 37 nontradables in a five-factor model. We find some positive evidence about the pricing abilities of the nontradables with respect to most of the equity portfolios, which confirms Bryzgalova et al. (2023) in highlighting the salience of nontradable factors for pricing.

References

ADRIAN, T., E. ETULA, AND T. MUIR (2014): "Financial intermediaries and the cross-section of asset returns," *The Journal of Finance*, 69, 2557–2596.

ANDERSON, T. W. AND H. RUBIN (1949): "Estimation of the parameters of a single equation in a complete system of stochastic equations," *The Annals of mathematical statistics*, 20, 46–63.

ANDRADE, S. C., A. EKPONON, AND A. JEANNERET (2023): "Sovereign risk premia and global macroeconomic conditions," *Journal of Financial Economics*, 147, 172–197.

ARDIA, D., K. BLUTEAU, K. BOUDT, AND K. INGHELBRECHT (2022): "Climate change concerns and the performance of green vs. brown stocks," *Management Science*.

ASNESS, C. S., A. FRAZZINI, AND L. H. PEDERSEN (2019): "Quality minus junk," *Review of Accounting Studies*, 24, 34–112.

ATANASOV, V., S. V. MØLLER, AND R. PRIESTLEY (2020): "Consumption fluctuations and expected returns," *The Journal of Finance*, 75, 1677–1713.

AVRAMOV, D., S. CHENG, L. METZKER, AND S. VOIGT (2021): "Integrating factor models," *Journal of Finance, forthcoming*.

BAI, J. (2003): "Inferential theory for factor models of large dimensions," *Econometrica*, 71, 135–171.

BAI, J. AND S. NG (2006): "Evaluating latent and observed factors in macroeconomics and finance," *Journal of Econometrics*, 131, 507–537.

BAKER, M. AND J. WURGLER (2006): "Investor sentiment and the cross-section of stock returns," *The journal of Finance*, 61, 1645–1680.

BALDUZZI, P. AND C. ROBOTTI (2008): "Mimicking portfolios, economic risk premia, and tests of multi-beta models," *Journal of Business & Economic Statistics*, 26, 354–368.

——— (2010): "Asset pricing models and economic risk premia: A decomposition," *Journal of Empirical Finance*, 17, 54–80.

BALI, T. G., S. J. BROWN, AND Y. TANG (2017): "Is economic uncertainty priced in the cross-section of stock returns?" *Journal of Financial Economics*, 126, 471–489.

BANDI, F. M., S. E. CHAUDHURI, A. W. LO, AND A. TAMONI (2021): "Spectral factor models," *Journal of Financial Economics*, 142, 214–238.

BARILLAS, F. AND J. SHANKEN (2018): "Comparing asset pricing models," *The Journal of Finance*, 73, 715–754.

BENJAMINI, Y. AND Y. HOCHBERG (1995): "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal statistical society: series B (Methodological)*, 57, 289–300.

BERKMAN, H., B. JACOBSEN, AND J. B. LEE (2011): "Time-varying rare disaster risk and stock returns," *Journal of Financial Economics*, 101, 313–332.

BIRRU, J. AND T. YOUNG (2022): "Sentiment and uncertainty," *Journal of Financial Economics*, 146, 1148–1169.

BOGUTH, O. AND L.-A. KUEHN (2013): "Consumption volatility risk," *The Journal of Finance*, 68, 2589–2615.

BOGUTH, O. AND M. SIMUTIN (2018): "Leverage constraints and asset prices: Insights from mutual fund risk taking," *Journal of Financial Economics*, 127, 325–341.

BOONS, M., F. DUARTE, F. DE ROON, AND M. SZYMANOWSKA (2020): "Time-varying inflation risk and stock returns," *Journal of Financial Economics*, 136, 444–470.

BREEDEN, D. T. (1979): "An intertemporal asset pricing model with stochastic consumption and investment opportunities," *Journal of financial Economics*, 7, 265–296.

BRYZGALOVA, S., J. HUANG, AND C. JULLIARD (2023): "Bayesian solutions for the factor zoo: We just ran two quadrillion models," *The Journal of Finance*, 78, 487–557.

BURNSIDE, C. (2011): "The cross section of foreign currency risk premia and consumption growth risk: Comment," *American Economic Review*, 101, 3456–3476.

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

CAMPBELL, J. Y. (1996): "Understanding risk and return," *Journal of Political economy*, 104, 298–345.

CAMPBELL, J. Y., S. GIGLIO, C. POLK, AND R. TURLEY (2018): "An intertemporal CAPM with stochastic volatility," *Journal of Financial Economics*, 128, 207–233.

CAMPBELL, J. Y. AND T. VUOLTEENAHO (2004): "Bad beta, good beta," *American Economic Review*, 94, 1249–1275.

CARHART, M. M. (1997): "On persistence in mutual fund performance," *The Journal of finance*, 52, 57–82.

CHAN, K., S. FORESI, AND L. H. LANG (1996): "Does money explain asset returns? Theory and empirical analysis," *The Journal of Finance*, 51, 345–361.

CHAPMAN, D. A. (1997): "The cyclical properties of consumption growth and the real term structure," *Journal of Monetary Economics*, 39, 145–172.

CHEN, A. Y. AND T. ZIMMERMANN (2021): "Open source cross-sectional asset pricing," *Critical Finance Review, Forthcoming*.

CHEN, J., G. TANG, J. YAO, AND G. ZHOU (2023a): "Employee sentiment and stock returns," *Journal of Economic Dynamics and Control*, 149, 104636.

CHEN, N.-F., R. ROLL, AND S. A. ROSS (1986): "Economic forces and the stock market," *Journal of business*, 383–403.

CHEN, Q. AND Z. FANG (2019): "Improved inference on the rank of a matrix," *Quantitative Economics*, 10, 1787–1824.

CHEN, Y., G. W. EATON, AND B. S. PAYE (2018): "Micro (structure) before macro? The predictive power of aggregate illiquidity for stock returns and economic activity," *Journal of Financial Economics*, 130, 48–73.

CHEN, Y., B. HAN, AND J. PAN (2021): "Sentiment trading and hedge fund returns," *The Journal of Finance*, 76, 2001–2033.

CHEN, Z., Z. DA, D. HUANG, AND L. WANG (2023b): "Presidential economic approval rating and the cross-section of stock returns," *Journal of Financial Economics*, 147, 106–131.

CHORDIA, T., A. GOYAL, AND A. SARETTO (2020): "Anomalies and false rejections," *The Review of Financial Studies*, 33, 2134–2179.

COCHRANE, J. H. (2011): "Presidential address: Discount rates," *The Journal of finance*, 66, 1047–1108.

CRAGG, J. G. AND S. G. DONALD (1997): "Inferring the rank of a matrix," *Journal of econometrics*, 76, 223–250.

DA, Z., J. ENGELBERG, AND P. GAO (2015): "The sum of all FEARS investor sentiment and asset prices," *The Review of Financial Studies*, 28, 1–32.

DITTMAR, R. F. (2002): "Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns," *The Journal of Finance*, 57, 369–403.

DONG, X., Y. LI, D. E. RAPACH, AND G. ZHOU (2022): "Anomalies and the expected market return," *The Journal of Finance*, 77, 639–681.

ELTON, E. J., M. J. GRUBER, AND C. R. BLAKE (1995): "Fundamental economic variables, expected returns, and bond fund performance," *The Journal of Finance*, 50, 1229–1256.

ENGELBERG, J., R. D. MCLEAN, J. PONTIFF, AND M. C. RINGGENBERG (2023): "Do cross-sectional predictors contain systematic information?" *Journal of Financial and Quantitative Analysis*, 58, 1172–1201.

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.

——— (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.

FENG, G., S. GIGLIO, AND D. XIU (2020): "Taming the factor zoo: A test of new factors," *The Journal of Finance*, 75, 1327–1370.

FRAZZINI, A. AND L. H. PEDERSEN (2014): "Betting against beta," *Journal of financial economics*, 111, 1–25.

GIGLIO, S., Y. LIAO, AND D. XIU (2021): "Thousands of alpha tests," *The Review of Finan-*

cial Studies, 34, 3456–3496.

GIGLIO, S. AND D. XIU (2021): "Asset pricing with omitted factors," *Journal of Political Economy*, 129, 1947–1990.

GIGLIO, S., D. XIU, AND D. ZHANG (2022): "Test assets and weak factors," Tech. rep., National Bureau of Economic Research.

GOLDBERG, J. AND Y. NOZAWA (2021): "Liquidity supply in the corporate bond market," *The Journal of Finance*, 76, 755–796.

GOSPODINOV, N., R. KAN, AND C. ROBOTTI (2012): "Further results on the limiting distribution of GMM sample moment conditions," *Journal of Business & Economic Statistics*, 30, 494–504.

——— (2014): "Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors," *The Review of Financial Studies*, 27, 2139–2170.

——— (2017): "Spurious inference in reduced-rank asset-pricing models," *Econometrica*, 85, 1613–1628.

——— (2019): "Too good to be true? Fallacies in evaluating risk factor models," *Journal of Financial Economics*, 132, 451–471.

GOSPODINOV, N. AND C. ROBOTTI (2021a): "Capital Share Risk in US Asset Pricing: A Reappraisal," *Journal of Finance, Replications and Comments*.

——— (2021b): "Common pricing across asset classes: Empirical evidence revisited," *Journal of Financial Economics*, 140, 292–324.

GRAMMIG, J. AND S. JANK (2016): "Creative destruction and asset prices," *Journal of Financial and Quantitative Analysis*, 51, 1739–1768.

HAMERMESH, D. S. (2007): "Replication in economics," *Canadian Journal of Economics/Revue canadienne d'économique*, 40, 715–733.

HANSEN, L. P. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica: Journal of the econometric society*, 1029–1054.

HANSEN, L. P. AND R. JAGANNATHAN (1997): "Assessing specification errors in stochastic discount factor models," *The Journal of Finance*, 52, 557–590.

HARVEY, C. R. (2017): "Presidential address: The scientific outlook in financial economics," *The Journal of Finance*, 72, 1399–1440.

HARVEY, C. R. AND Y. LIU (2020): "False (and missed) discoveries in financial economics," *The Journal of Finance*, 75, 2503–2553.

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

He, Z., B. Kelly, and A. Manela (2017): "Intermediary asset pricing: New evidence from many asset classes," *Journal of Financial Economics*, 126, 1–35.

Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh (2016): "The common factor in idiosyncratic volatility: Quantitative asset pricing implications," *Journal of Financial Economics*, 119, 249–283.

Hou, K., C. Xue, and L. Zhang (2020): "Replicating anomalies," *The Review of financial studies*, 33, 2019–2133.

Hu, G. X., J. Pan, and J. Wang (2013): "Noise as information for illiquidity," *The Journal of Finance*, 68, 2341–2382.

Huang, D., F. Jiang, J. Tu, and G. Zhou (2015): "Investor sentiment aligned: A powerful predictor of stock returns," *The Review of Financial Studies*, 28, 791–837.

Huang, D., J. Li, and L. Wang (2021): "Are disagreements agreeable? Evidence from information aggregation," *Journal of Financial Economics*, 141, 83–101.

Huynh, T. D. and Y. Xia (2021): "Climate change news risk and corporate bond returns," *Journal of Financial and Quantitative Analysis*, 56, 1985–2009.

Jagannathan, R. and Y. Wang (2007): "Lazy investors, discretionary consumption, and the cross-section of stock returns," *The Journal of Finance*, 62, 1623–1661.

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

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

Jensen, T. I., B. T. Kelly, and L. H. Pedersen (2021): "Is there a replication crisis in finance?" Tech. rep., National Bureau of Economic Research.

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

Kargar, M. (2021): "Heterogeneous intermediary asset pricing," *Journal of Financial Economics*, 141, 505–532.

Kleibergen, F. (2009): "Tests of risk premia in linear factor models," *Journal of economet-*

rics, 149, 149–173.

KLEIBERGEN, F., L. KONG, AND Z. ZHAN (2023): "Identification robust testing of risk premia in finite samples," *The Journal of Financial Econometrics*, 21, 263–297.

KLEIBERGEN, F. AND R. PAAP (2006): "Generalized reduced rank tests using the singular value decomposition," *Journal of econometrics*, 133, 97–126.

KLEIBERGEN, F. AND Z. ZHAN (2015): "Unexplained factors and their effects on second pass R-squared's," *Journal of Econometrics*, 189, 101–116.

——— (2018): "Identification-robust inference on risk premia of mimicking portfolios of non-traded factors," *Journal of Financial Econometrics*, 16, 155–190.

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

——— (2022): "Misspecification and Weak Identification in Asset Pricing," *arXiv preprint arXiv:2206.13600*.

KROENCKE, T. A. (2017): "Asset pricing without garbage," *The Journal of Finance*, 72, 47–98.

KROENCKE, T. A. AND J. THIMME (2021): "A skeptical appraisal of robust asset pricing tests," in *Proceedings of Paris December 2021 Finance Meeting EUROFIDAI-ESSEC*.

KUEHN, L.-A., M. SIMUTIN, AND J. J. WANG (2017): "A labor capital asset pricing model," *The Journal of Finance*, 72, 2131–2178.

LETTAU, M. AND S. LUDVIGSON (2001): "Consumption, aggregate wealth, and expected stock returns," *the Journal of Finance*, 56, 815–849.

LETTAU, M., S. C. LUDVIGSON, AND S. MA (2019): "Capital share risk in US asset pricing," *The Journal of Finance*, 74, 1753–1792.

LETTAU, M., M. MAGGIORI, AND M. WEBER (2014): "Conditional risk premia in currency markets and other asset classes," *Journal of Financial Economics*, 114, 197–225.

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

LINNAINMAA, J. T. AND M. R. ROBERTS (2018): "The history of the cross-section of stock returns," *The Review of Financial Studies*, 31, 2606–2649.

LIU, Y. AND B. MATTHIES (2022): "Long-Run Risk: Is It There?" *The Journal of Finance*, 77, 1587–1633.

LUSTIG, H. N. AND S. G. VAN NIEUWERBURGH (2005): "Housing collateral, consumption insurance, and risk premia: An empirical perspective," *The Journal of Finance*, 60, 1167–1219.

MANELA, A. AND A. MOREIRA (2017): "News implied volatility and disaster concerns," *Journal of Financial Economics*, 123, 137–162.

MARTIN, I. W. AND C. WAGNER (2019): "What is the Expected Return on a Stock?" *The Journal of Finance*, 74, 1887–1929.

MCLEAN, R. D. AND J. PONTIFF (2016): "Does academic research destroy stock return predictability?" *The Journal of Finance*, 71, 5–32.

PAPANIKOLAOU, D. (2011): "Investment shocks and asset prices," *Journal of Political Economy*, 119, 639–685.

PARKER, J. A. AND C. JULLIARD (2005): "Consumption risk and the cross section of expected returns," *Journal of Political Economy*, 113, 185–222.

PÁSTOR, L. AND R. F. STAMBAUGH (2003): "Liquidity risk and expected stock returns," *Journal of Political economy*, 111, 642–685.

REICHENBACHER, M. AND P. SCHUSTER (2022): "Size-adapted bond liquidity measures and their asset pricing implications," *Journal of Financial Economics*, 146, 425–443.

SADKA, R. (2006): "Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk," *Journal of Financial Economics*, 80, 309–349.

SANTOS, T. AND P. VERONESI (2006): "Labor income and predictable stock returns," *The Review of Financial Studies*, 19, 1–44.

SAVOV, A. (2011): "Asset pricing with garbage," *The Journal of Finance*, 66, 177–201.

SHANKEN, J. (1992): "On the estimation of beta-pricing models," *The review of financial studies*, 5, 1–33.

SWEENEY, R. J. AND A. D. WARGA (1986): "The pricing of interest-rate risk: evidence from the stock market," *The Journal of Finance*, 41, 393–410.

VASSALOU, M. (2003): "News related to future GDP growth as a risk factor in equity returns," *Journal of financial economics*, 68, 47–73.

VASSALOU, M. AND Y. XING (2004): "Default risk in equity returns," *The journal of finance*, 59, 831–868.

WEBER, M. (2018): "Cash flow duration and the term structure of equity returns," *Journal of Financial Economics*, 128, 486–503.

WELCH, I. ET AL. (2019): "Reproducing, extending, updating, replicating, reexamining, and reconciling," *Critical Finance Review*, 8, 301–304.

YAN, X. S. AND L. ZHENG (2017): "Fundamental analysis and the cross-section of stock returns: A data-mining approach," *The Review of Financial Studies*, 30, 1382–1423.

YOGO, M. (2006): "A consumption-based explanation of expected stock returns," *The Journal of Finance*, 61, 539–580.

ZHANG, X., Y. LIU, K. WU, AND B. MAILLET (2021): "Tradable or nontradable factors—what does the Hansen–Jagannathan distance tell us?" *International Review of Economics & Finance*, 71, 853–879.

A Appendix A

A.1 Identification for the CAPM and FF3 models

As part of our investigation consists in understanding when the non-traded factors can still be identified after adding controlling for some factors, we provide some benchmark results on the control factors, i.e., the MKT factor and the Fama and French (1993)'s three factors.

We start by table 7 which, together with 2 and 3, shows the results of the identification tests for the CAPM.

For equity portfolios, all of the rank tests suggest that the CAPM is well identified with respect to all of equity portfolios at monthly and quarterly frequencies. It confirms the intuition that the market factor is correlated with most of equity test assets. At annual frequency, with shorter time series, the picture is less uniform. While CD test and Chen-Fang tests signal strong identification for all test assets classes, the F test (robust to finite sample) points at rather rank deficiency, in particular with FF25 BEME/OP, FF25 ME/INV, HXZ 48 portfolios on 24 anomalies. The picture following the KP test is more concerning as it indicates that the model is only identified with FF25 size and momentum portfolios at yearly frequency. This is also in line with Kleibergen and Zhan (2020)'s findings, especially in light of their considerations on the finite sample on the power of tests, see also Kroencke and Thimme (2021).

For non-equity portfolios, the CD test and the F test indicate the model is identified at all sampling frequency (except for the commodity portfolios), and the Chen-Fang test suggest that the CAPM is weakly identified with respect to the FX portfolios (at all sampling frequencies). Lastly, the KP test seems to be the most conservative in terms of rejecting the null of rank deficiency, especially as the sampling frequency decreases (possibly due the power distortions). For the annual frequency, we cannot reject the rank deficiency of CAPM with respect to all non-equity portfolios.

Table 8 and figures 4 - 5 report the results for the identification tests of FF3 model.

For equity portfolios, except for 10 Weber portfolios, the FF3 model is deemed as identified with all of the equity portfolios,at monthly frequency. However, at lower frequencies, the picture is less harmonic. For quarterly data, although being always identified according to Chen-Fang test, the model is considered to be weakly identified for FF25 BM/OP portfolios by the remaining tests. Moreover, the KP test suggests that the model is weakly identified also for FF25 ME/INV and FF25 ME/MOM. At yearly frequency, the KP and F tests the model rarely reject the null, which again may suggest is due to low power given the finite sample.

For non-equity portfolios, both the KP test and F test suggest the model is always weakly identified, at the various frequencies. Regarding the CD test, while it points out that the FF3 model is weakly identified at monthly frequency, it shows that identification strength is regained at lower frequencies, especially for the quarterly and annual CDS portfolios, annual commodity and FX portfolios. However, although seeming puzzling, we want to underscore that the sample size at yearly frequency is extremely small ($T < 40$), and the it clearly affects the reliability of the test.

B Appendix B

B.1 List of non-traded factors

The following tables list the non-traded factors that we include.

We highlight that we did not replicate the factors that are already available on the authors' websites. For further details, we invite the reader to refer to their works or to the personal website of the following authors: David Ardia, Oliver Boguth, Yong Chen, Zhi Da, Fernando Duarte, Stefano Giglio, Amit Goyal, Joachim Grammig, Bernard Herskovic, Dashan Huang, Mahyar Kargar, Tim Kroencke, Yukun Liu, Martin Lettau, Sydney Ludvigson, Asaf Manela, Tyler Muir, Jun Pan, Christopher Polk, Ronnie Sadka, Robert Stambaugh, Philipp Schuster, Stijn Van Nieuwerburgh, Maria Vassalou, Jessie Jiaxu Wang, Jeffrey Wurgler, Guofu Zhou.

The non-traded factors are grouped into 6 broad categories based on the risk dimensions the factors are measuring: News (Table 1), Sentiment (Table 2), Liquidity (Table 3), Consumption (Table 4), Macroeconomics (Table 5), Intermediary and aggregate firm-level risk (Table 6). For each factor, we list its short name[36], brief description, and reference.

[36]The short name is the label used in our dataset and in our replication package.

Table 1: News

Short Name	Factor Description	Reference
N_dr	Discount-rate news	Campbell and Vuolteenaho (2004)
N_cf	Cash-flow news	Campbell and Vuolteenaho (2004)
Nr	Discount-rate news	Campbell et al. (2018)
Ncf	Cash-flow news	Campbell et al. (2018)
Nv	News about market return variance	Campbell et al. (2018)
NVIX	News implied volatility	Manela and Moreira (2017)
wsj_AR1_Inno	Residuals from an AR(1) model of WSJ Climate Change News Index	Huynh and Xia (2021)
chneg_AR1_inno	Residuals from an AR(1) model of CH Negative Climate Change News Index	Huynh and Xia (2021)
NI	Wall Street Journal news index about economic growth prospects	Liu and Matthies (2022)
HNI	Innovations to the HN-index	Liu and Matthies (2022)
inn_MCCI	Unexpected media climate change concerns	Ardia et al. (2022)

Table 2: Sentiment

Short Name	Factor Description	Reference
investor_sent	Investor sentiment	Baker and Wurgler (2006)
investor_sent_orth	Investor sentiment orthogonal to macro variables	Baker and Wurgler (2006)
PLS_SENT_lag	Investor sentiment index lagged one month	Huang et al. (2015)
PLS_SENT_ortmacrolag	Investor sentiment index orthogonal to macro variables lagged one month	Huang et al. (2015)
FEARS30	Market sentiment: daily Internet search	Da et al. (2015)
FEARS35	Market sentiment: daily Internet search	Da et al. (2015)
FEARS25	Market sentiment: daily Internet search	Da et al. (2015)
DisagHigh	Disagreement in high sentiment period	Huang et al. (2021)
DisagLow	Disagreement in low sentiment period	Huang et al. (2021)
Disag	Disagreement index	Huang et al. (2021)
Cons_sentiment	Monthly change in Michigan US consumer sentiment index	Chen et al. (2021)
Inv_sent_X_NVIX_sq	Interaction between BW investor sentiment and M-M news implied volatility	Birru and Young (2022)
Inv_sent_or_X_NVIX_sq	Interaction between BW orthogonal investor sentiment and M-M news implied volatility	Birru and Young (2022)
Attention_PCA	First PC from 12 individual attention proxies using PCA	Chen et al. (2023a)
Attention_PLS	First PC from 12 individual attention proxies using PLS	Chen et al. (2023a)
Attention_sPCA	First PC from 12 individual attention proxies using sPCA	Chen et al. (2023a)
PEAR	Monthly presidential economic approval rating (PEAR) index	Chen et al. (2023b)

Table 3: Liquidity

Short Name	Factor Description	Reference
Illiq_Trans	Transitory illiquidity	Sadka (2006)
Illiq_Perm	Permanent illiqudity	Sadka (2006)
Liq	Liquidity	Pástor and Stambaugh (2003)
Adj_logroll	Break- and volatility-adjusted Roll measure (ROLL)	Chen et al. (2018)
Adj_logcorwin	Break- and volatility-adjusted Corwin and Schultz measure (CS)	Chen et al. (2018)
Adj_logfht	Break- and volatility-adjusted Fong, Holden, and Trzcinka measure (FHT)	Chen et al. (2018)
Adj_logtick	Break- and volatility-adjusted effective tick measure (TICK)	Chen et al. (2018)
Adj_logzeros	Break- and volatility-adjusted zeros measure (ZEROS)	Chen et al. (2018)
Adj_logamireal	Break- and volatility-adjusted Amihud illiquidity measure (AMI)	Chen et al. (2018)
Adj_logamito	Break- and volatility-adjusted Amihud measure based on the ratio of absolute daily returns to daily turnover	Chen et al. (2018)
Adj_logps	Break- and volatility-adjusted Pastor and Stambaugh measure (PS)	Chen et al. (2018)
Adj_loghm	Break- and volatility-adjusted Hou and Moskowitz measure (HM):	Chen et al. (2018)
Adj_PC	Principal component from break- and volatility-adjusted illiquidity measures above	Chen et al. (2018)
Liq_supply_shock	Liquidity supply shock	Goldberg and Nozawa (2021)
Liq_demand_shock	Liquidity demand shock	Goldberg and Nozawa (2021)
Rationorm_mean_cvols	Trade size-adapted Schultz (2001) bond liquidity measure (normalized)	Reichenbacher and Schuster (2022)

Table 4: Consumption

Short Name	Factor Description	Reference
cons_PCE, cons_PCEND	Polynomials of consumption growth	Chapman (1997)
cay	Consumption–aggregate wealth ratio in terms of observable variables	Lettau and Ludvigson (2001)
Pjconsump	(Quarterly/Annual) PJ consumption	Parker and Julliard (2005)
Dur	Durable consumption	Yogo (2006)
Nondur	Nondurable consumption	Yogo (2006)
Q4Con	Annual_Q4-Q4_cons	Jagannathan and Wang (2007)
Gar	Per capita garbage growth	Savov (2011)
D_mu	Changes in beliefs about the conditional mean of consumption growth	Boguth and Kuehn (2013)
D_sig	Changes in beliefs about consumption growth volatility	Boguth and Kuehn (2013)
NIPA-N&S, NIPA-QQ-N&S	(Annual/Quarterly) Reported Nondurables and Services consumption	Kroencke (2017)
NIPA-N, NIPA-QQ-N	(Annual/Quarterly) Reported nondurables consumption	Kroencke (2017)
NIPA-S, NIPA-QQ-S	(Annual/Quarterly) Reported S: Services consumption	Kroencke (2017)
Q4-N	Annual Q4 to Q4 annual consumption of nondurables only	Kroencke (2017)
UNFIL-N&S, UNFIL-QQ-N&S	(Annual/Quarterly) Unfiltered N&S consumption	Kroencke (2017)
UNFIL-N, UNFIL-QQ-N	(Annual/Quarterly) Unfiltered nondurables consumption	Kroencke (2017)
UNFIL-Q4-N&S	Annual Unfiltered Q4 to Q4 annual consumption	Kroencke (2017)
UNFIL-Q4-N	Annual Unfiltered Q4 to Q4 annual consumption of nondurables only	Kroencke (2017)
CC	Cyclical consumption	Atanasov et al. (2020)
exp_con_grow	Expected consumption growth	Andrade et al. (2023)

Table 5: Macroeconomics

Short Name	Factor Description	Reference
LTR	Change in long-term government bond yield	Sweeney and Warga (1986)
IndProdY	Industrial production growth (monthly/yearly)	Chen et al. (1986)
Oil	Oil price growth	Chen et al. (1986)
MICHI_inf_exp	Change in Inflation expectation (michigan survey)	Elton et al. (1995)
Real_GDP	Change in Expectation on GNP(GDP) survey	Elton et al. (1995)
LaborGrowth	Labor income growth	Campbell (1996)
M2C_SU, M2C_SA	Seasonally unadjusted or adjusted per-capita inside money (growth)	Chan et al. (1996)
M3C_SU, M3C_SA	Seasonally unadjusted or adjusted per-capita inside money (growth)	Chan et al. (1996)
LaborGrowth	Labor income growth	Jagannathan and Wang (1996), Jagannathan and Wang (2007)
D_LI_growth	Difference in log labor earnings	Lettau and Ludvigson (2001)
LaborGrowth, LB	Polynomials of labor income growth and returns	Dittmar (2002)
myfar	Fixed-assets-based ratio housing to human wealth	Lustig and Van Nieuwerburgh (2005)
mymor	Mortgage-based ratio housing to human wealth	Lustig and Van Nieuwerburgh (2005)
myrwr	Residential-wealth-based ratio housing to human wealth	Lustig and Van Nieuwerburgh (2005)
crisis	Unexpected disaster risk	Berkman et al. (2011)
exp_crisis	Expected disaster risk	Berkman et al. (2011)
IMC	Investment minus consumption producers innovations	Papanikolaou (2011)
IS_PI	Relative price of new equipment	Papanikolaou (2011)
Noise	Root mean squared distance between the market yields and the Svensson model-implied yields	Hu et al. (2013)
PAG	Net growth rate of patenting activity (Quarterly/Annual)	Grammig and Jank (2016)
Macro_uncert	Macro uncertainty	Bali et al. (2017)
Fin_uncert	Financial uncertainty	Bali et al. (2017)
Real_uncert	Real Uncertainty	Bali et al. (2017)
LaborTightness	LaborTightness	Kuehn et al. (2017)
Dlogks	Growth in capital share	Lettau et al. (2019)
NRC	Nominal-real covariance	Boons et al. (2020)
Inf	Inflation innovation from ARMA(1,1)	Boons et al. (2020)

Table 6: Intermediary, volatility, and aggregate firm-level risk

Short Name	Factor Description	Reference
dsv	Equally weighted average of the firm default likelihood	Vassalou and Xing (2004)
Lev	(Seasonally adjusted) log changes in the level of broker-dealer leverage	Adrian et al. (2014)
CIV	Equally weighted average of firm-level market model residual return variance (common idiosyncratic volatility shock)	Herskovic et al. (2016)
HKMcapital	Intermediary capital risk factor	He et al. (2017)
LCT	Value-weighted average beta of the aggregate stock holdings of all actively managed equity fund (leverage constraint tightness)	Boguth and Simutin (2018)
Svix2_vw	The value-weighted average of stocks' risk-neutral variance	Martin and Wagner (2019)
Svix2_index	Risk-neutral variance of the market index	Martin and Wagner (2019)
Quart_HIFac	The heterogeneous intermediary factor	Kargar (2021)

B.2 Empirical results

Table 7: Model diagnosis of CAPM
This table presents the p values of misspecification tests (HJ distance and J test), identification tests (KP test, F test, CD test and CF test), and the associated number of time series observations for monthly, quarterly and annual market excess return relative to the sets of test assets used in our paper. *** significant at 1% signficiance level. ** significant at 5% level. * significant at 1% level.

ModelName	TestAssets	HJ distance	J test	KP test	F test	CD test	CF test	T	N
MKT	FF25_BEME_OP	0.03**	0.03**	0.00***	0.00***	0.00***	0***	705	25
MKT	FF25_ME_BM	0.00***	0.00***	0.00***	0.00***	0.00***	0***	1143	25
MKT	FF25_ME_INV	0.00***	0.00***	0.00***	0.00***	0.00***	0***	705	25
MKT	FF25_ME_MOM	0.00***	0.00***	0.00***	0.00***	0.00***	0***	1143	25
MKT	HKMCDS	0.00***	0.18	0.01**	0.00***	0.00***	0***	143	20
MKT	HKMCommodity	0.06*	0.06*	0.00***	0.00***	0.00***	0***	316	23
MKT	HKMFX	0.00***	0.11	0.10	0.00***	0.00***	0.07*	407	12
MKT	HKMOptions	0.00***	0.00***	0.00***	0.00***	0.00***	0***	310	18
MKT	HKMSovBonds	0.19	0.23	0.01**	0.00***	0.00***	0***	196	6
MKT	HKMUSBonds	0.00***	0.01**	0.00***	0.00***	0.00***	0***	444	20
MKT	HXZ_24_anoms_porf_48	0.00***	0.00***	0.00***	0.00***	0.00***	0***	660	48
MKT	US_anoms_porf_24	0.00***	0.00***	0.00***	0.00***	0.00***	0***	705	24
MKT	Weber10_duration	0.00***	0.00***	0.00***	0.00***	0.00***	0***	612	10
Quart_MKT	FF25_BEME_OP	0.05**	0.07*	0.00***	0.00***	0.00***	0***	235	25
Quart_MKT	FF25_ME_BM	0.00***	0.00***	0.00***	0.00***	0.00***	0***	381	25
Quart_MKT	FF25_ME_INV	0.00***	0.00***	0.00***	0.00***	0.00***	0***	235	25
Quart_MKT	FF25_ME_MOM	0.00***	0.00***	0.00***	0.00***	0.00***	0***	381	25
Quart_MKT	HKMCDS	0.00***	0.04**	0.39	0.00***	0.00***	0***	48	20
Quart_MKT	HKMCommodity	0.09*	0.24	0.11	0.00***	0.00***	0***	105	23
Quart_MKT	HKMFX	0.06**	0.26	0.10	0.01***	0.00***	0.26	135	12
Quart_MKT	HKMOptions	0.00***	0.00***	0.01***	0.00***	0.00***	0***	103	18
Quart_MKT	HKMSovBonds	0.04**	0.24	0.11	0.00***	0.00***	0***	65	6
Quart_MKT	HKMUSBonds	0.00***	0.01***	0.00***	0.00***	0.00***	0***	148	20
Quart_MKT	HXZ_24_anoms_porf_48	0.00***	0.00***	0.01***	0.00***	0.00***	0***	220	48
Quart_MKT	US_anoms_porf_24	0.00***	0.00***	0.00***	0.00***	0.00***	0***	235	24
Quart_MKT	Weber10_duration	0.00***	0.00***	0.00***	0.00***	0.00***	0***	204	10
Annual_MKT	FF25_BEME_OP	0.00***	0.01**	0.42	0.14	0.00***	0***	58	25
Annual_MKT	FF25_ME_BM	0.00***	0.00***	0.11	0.01**	0.00***	0***	95	25
Annual_MKT	FF25_ME_INV	0.00***	0.01**	0.51	0.31	0.00***	0***	58	25
Annual_MKT	FF25_ME_MOM	0.00***	0.00***	0.01**	0.01**	0.00***	0***	95	25
Annual_MKT	HKMCDS								
Annual_MKT	HKMCommodity	0.00***	0.00***	0.36	0.26	0.00***	0.11	26	23
Annual_MKT	HKMFX	0.00***	0.10	0.28	0.01**	0.00***	0.66	33	12
Annual_MKT	HKMOptions	0.00***	0.00***	0.52	0.02**	0.00***	0***	25	18
Annual_MKT	HKMSovBonds	0.02**	0.00***	0.23	0.01**	0.00***	0.00***	16	6
Annual_MKT	HKMUSBonds	0.00***	0.00***	0.32	0.00***	0.00***	0***	37	20
Annual_MKT	HXZ_24_anoms_porf_48	0.00***	0.00***	0.41	0.35	0.00***	0***	55	48
Annual_MKT	US_anoms_porf_24	0.00***	0.00***	0.14	0.00***	0.00***	0***	58	24
Annual_MKT	Weber10_duration	0.00***	0.06*	0.10	0.03**	0.01**	0***	50	10

Table 8: Model diagnosis of FF3

This table presents the p values of misspecification tests (HJ distance and J test), identification tests (KP test, F test, CD test and CF test), and the associated number of time series observations for monthly, quarterly and annual Fama-French three factors relative to the sets of test assets used in our paper. *** significant at 1% signficiance level. ** significant at 5% level. * significant at 1% level.

ModelName	TestAssets	HJ distance	J test	KP test	F test	CD test	CF test	T	N
FF3	FF25_BEME_OP	0.43	0.66	0.04**	0.00***	0.01**	0.00***	705	25
FF3	FF25_ME_BM	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	1143	25
FF3	FF25_ME_INV	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	705	25
FF3	FF25_ME_MOM	0.00***	0.04**	0.00***	0.00***	0.00***	0.00***	1143	25
FF3	HKMCDS	0.01**	0.91	0.78	0.32	0.19	0.82	143	20
FF3	HKMCommodity	0.21	0.76	0.30	0.65	0.09	0.00***	316	23
FF3	HKMFX	0.01**	0.56	0.33	0.45	0.34	0.34	407	12
FF3	HKMOptions	0.00***	0.87	0.68	0.81	0.80	0.80	310	18
FF3	HKMSovBonds	0.87	0.96	0.99	0.98	0.99	0.99	196	6
FF3	HKMUSBonds	0.00***	0.18	0.15	0.06*	0.10	0.02**	444	20
FF3	HXZ_24_anoms_porf_48	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	660	48
FF3	US_anoms_porf_24	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	705	24
FF3	Weber10_duration	0.64	0.77	0.31	0.23	0.27	0.00	612	10
Quart_FF3	FF25_BEME_OP	0.12	0.36	0.47	0.41	0.08	0.00	235	25
Quart_FF3	FF25_ME_BM	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	381	25
Quart_FF3	FF25_ME_INV	0.00***	0.14	0.20	0.00***	0.00***	0.00***	235	25
Quart_FF3	FF25_ME_MOM	0.00***	0.78	0.17	0.00***	0.00***	0.00***	381	25
Quart_FF3	HKMCDS	0.00***	0.38	0.26	0.09*	0.00***	0.53	48	20
Quart_FF3	HKMCommodity	0.19	0.85	0.76	0.80	0.72	0.00***	105	23
Quart_FF3	HKMFX	0.03**	0.71	0.79	0.64		0.16	135	12
Quart_FF3	HKMOptions	0.00***	0.94	0.96	0.99	0.92	0.36	103	18
Quart_FF3	HKMSovBonds	0.83	0.85	0.81	0.87	0.80	0.99	65	6
Quart_FF3	HKMUSBonds	0.00***	0.66	0.67	0.60	0.33	0.02**	148	20
Quart_FF3	HXZ_24_anoms_porf_48	0.00***	0.00***	0.03**	0.00***	0.00***	0.00***	220	48
Quart_FF3	US_anoms_porf_24	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	235	24
Quart_FF3	Weber10_duration	0.20	0.56	0.58	0.36	0.49	0.13	204	10
Annual_FF3	FF25_BEME_OP	0.00***	0.38	0.31	0.76	0.00***	0.07*	58	25
Annual_FF3	FF25_ME_BM	0.00***	0.56	0.72	0.33	0.09*	0.00***	95	25
Annual_FF3	FF25_ME_INV	0.00***	0.46	0.32	0.88	0.01**	0.00***	58	25
Annual_FF3	FF25_ME_MOM	0.00***	0.21	0.06	0.61	0.00***	0.02**	95	25
Annual_FF3	HKMCDS								
Annual_FF3	HKMCommodity	0.00***	0.00***	0.56	0.88	0.00***	0.73	26	23
Annual_FF3	HKMFX	0.07*	0.87	0.91	0.85	0.01**	0.34	33	12
Annual_FF3	HKMOptions	0.02**	0.03**	0.32	0.41	0.00***	0.91	25	18
Annual_FF3	HKMSovBonds	1.00	1.00	0.76	0.80	0.57	0.81	16	6
Annual_FF3	HKMUSBonds	0.00***	0.08*	0.38	0.74		0.23	37	20
Annual_FF3	HXZ_24_anoms_porf_48	0.00***	0.00***	0.67	0.47	0.00***	0.00***	55	48
Annual_FF3	US_anoms_porf_24	0.00***	0.00***	0.08*	0.00***	0.00***	0.00***	58	24
Annual_FF3	Weber10_duration	0.90	0.91	0.51	0.58	0.41	0.34	50	10

Table 9: Identification failure of original models
This table reports the percentage of model identification failure in four tests: KP test, F test, CD test and CF rank test. The null hypothesis of KP, F and CD tests is $\mathcal{H}_0^{rk} : rk(\beta) = K-1$, and the CF statistic is tested on $\mathcal{H}_0^{rk} : rk(\beta) \leqslant K-1$. The upper bound represents the most conservative situation where the model is identified if we can reject in all of four tests and the lower bound represents the most favorable situation where the model is identified if we can reject in any of the tests. The percentage of models that are rejected by the specification tests: Hansen-Jagannathan distance test and over-identifying restriction J test is also reported. Panel A shows the fraction regardless of sample size T and Panel B, C, D show the percentages calculated with respect to $T \leq 100$, $100 < T \leq 300$ and $T > 300$. We also show the results corrected for multiple hypothesis testing.

	KPtest	Ftest	CDtest	CFtest	Upper bound	Lower bound	HJ distance	J test	Models
Panel A: the unconditional fraction of identification failure									
Models Original	89.78	73.74	50.34	48.92	88.53	29.17	70.26	13.58	1752
Models Original (corrected)	98.52	83.33	55.25	51.43	90.07	32.88	67.92	4.62	1752
Panel B: the fraction of identification failure for models in small sample (T<=100)									
Models Original	99.71	90.94	29.53	65.20	95.61	26.90	69.30	27.19	342
Models Original (corrected)	100.00	97.66	32.46	69.01	99.12	29.82	67.25	12.28	342
Panel C: the fraction of identification failure for models in small sample (100<T<=300)									
Models Original	94.79	77.20	52.24	54.76	88.33	34.29	66.25	7.54	557
Models Original (corrected)	100.00	84.56	56.73	56.55	91.20	38.42	63.02	1.44	557
Panel D: the fraction of identification failure for models in large sample (T>300)									
Models Original	82.53	64.60	57.44	38.57	85.81	26.73	73.27	12.08	853
Models Original (corrected)	96.95	76.79	63.42	41.03	85.70	30.48	71.40	3.63	853

Table 10: Model identification and misspecification sorted by group of test assets
This table summarizes the percentage of model identification failure for each set of test assets in four tests: KP test, F test, CD test and CF rank test. The null hypothesis of KP, F and CD tests is $\mathcal{H}_0^{rk}: rk(\beta) = K-1$, and the CF statistic is tested on $\mathcal{H}_0^{rk}: rk(\beta) \leqslant K-1$. The upper bound represents the most conservative situation where the model is identified if we can reject in all of four tests and the lower bound represents the most favorable situation where the model is identified if we can reject in any of the tests. The percentage of models that are rejected by the specification tests: Hansen-Jagannathan distance test and over-identifying restriction J test is also reported. The last column shows the number of models tested for each set of test portfolios.

	KPtest	Ftest	CDtest	CFtest	Upper bound	Lower bound	HJ distance	J test	Models
FF25_BEME_OP	91.37	71.94	50.36	38.85	84.89	23.74	41.73	8.63	139
FF25_ME_BM	86.33	65.47	42.45	41.73	85.61	23.02	87.05	15.83	139
FF25_ME_INV	92.09	77.70	48.20	42.45	90.65	24.46	97.84	15.11	139
FF25_ME_MOM	85.61	60.43	38.85	38.13	81.29	18.71	94.24	15.83	139
HKMCDS	92.38	58.10	34.29	68.57	81.90	29.52	92.38	5.71	105
HKMCommodity	97.69	87.69	49.23	40.00	92.31	19.23	13.85	8.46	130
HKMFX	95.00	90.00	74.29	63.57	96.43	48.57	60.71	5.00	140
HKMOptions	92.09	84.17	56.12	55.40	92.81	32.37	81.29	15.83	139
HKMSovBonds	97.04	88.15	80.00	73.33	97.78	59.26	22.96	10.37	135
HKMUSBonds	87.77	60.43	43.17	55.40	82.01	23.74	92.09	13.67	139
HXZ_24_anoms_porf_48	81.54	62.31	18.46	32.31	80.77	9.23	98.46	33.08	130
US_anoms_porf_24	79.14	60.43	34.53	37.41	86.33	17.27	89.21	20.14	139
Weber10_duration	89.93	88.49	79.14	52.52	96.40	48.92	43.88	7.91	139

Table 11: Model identification and misspecification sorted by group of test assets at 5% false discovery rate

This table summarizes the percentage of model identification failure for each set of test assets in four tests: KP test, F test, CD test and CF rank test. The null hypothesis of KP, F and CD tests is $\mathcal{H}_0^{rk}: rk(\beta) = K - 1$, and the CF statistic is tested on $\mathcal{H}_0^{rk}: rk(\beta) \leqslant K - 1$. The upper bound represents the most conservative situation where the model is identified if we can reject in all of four tests and the lower bound represents the most favorable situation where the model is identified if we can reject in any of the tests. The percentage of models that are rejected by the specification tests: Hansen-Jagannathan distance test and over-identifying restriction J test is also reported. The last column shows the number of models tested for each set of test portfolios.

	KPtest	Ftest	CDtest	CFtest	Upper bound	Lower bound	HJ distance	J test	Models
FF25_BEME_OP	99.28	82.73	53.96	40.29	87.05	25.90	36.69	0.72	139
FF25_ME_BM	97.12	76.26	48.20	43.88	87.05	27.34	86.33	2.16	139
FF25_ME_INV	98.56	85.61	53.96	43.17	89.21	28.78	97.12	2.88	139
FF25_ME_MOM	97.84	75.54	45.32	39.57	82.73	23.02	90.65	2.88	139
HKMCDS	100.00	69.52	39.05	69.52	83.81	34.29	90.48	0.95	105
HKMCommodity	100.00	94.62	53.08	43.08	96.15	20.77	10.00	6.92	130
HKMFX	99.29	95.71	80.00	67.86	99.29	52.86	58.57	1.43	140
HKMOptions	97.84	91.37	61.87	58.27	97.12	36.69	78.42	8.63	139
HKMSovBonds	100.00	94.81	85.93	74.81	98.52	65.19	20.74	6.67	135
HKMUSBonds	97.84	75.54	49.64	62.59	89.21	32.37	90.65	5.04	139
HXZ_24_anoms_porf_48	99.23	73.08	20.77	36.92	80.00	10.77	96.92	13.85	130
US_anoms_porf_24	95.68	71.94	35.97	38.85	82.01	18.71	88.49	6.47	139
Weber10_duration	98.56	93.53	84.89	53.24	97.12	49.64	40.29	1.44	139

Table 12: Model identification and misspecification sorted by factor groups
This table summarizes the percentage of model identification failure for each category of factors in four tests: KP test, F test, CD test and CF rank test. The null hypothesis of KP, F and CD tests is $\mathcal{H}_0^{rk} : rk(\beta) = K - 1$, and the CF statistic is tested on $\mathcal{H}_0^{rk} : rk(\beta) \leqslant K - 1$. The upper bound represents the most conservative situation where the model is identified if we can reject in all of four tests and the lower bound represents the most favorable situation where the model is identified if we can reject in any of the tests. The percentage of models that are rejected by the specification tests: Hansen-Jagannathan distance test and over-identifying restriction J test is also reported. The last column shows the number of models tested for each category of factors. We also report the percentages corrected for multiple hypothesis in Panel B.

	KPtest	Ftest	CDtest	CFtest	Upper bound	Lower bound	HJ distance	J test	Models
Panel A: the percentage of identification failure and misspecification									
News	92.81	82.04	62.28	52.69	93.41	39.52	70.66	7.78	167
Consumption	98.18	82.42	31.82	36.67	87.27	16.97	76.67	23.03	330
Macro	92.82	79.89	53.50	56.19	89.41	33.03	71.27	10.95	557
Liquidity	83.14	62.45	59.00	36.78	89.27	18.77	62.45	10.73	261
Sentiment	82.04	60.82	54.69	51.43	83.27	35.51	66.94	11.84	245
Intermediary	88.46	68.27	31.73	52.88	90.38	20.19	73.08	13.46	104
Volatility	61.54	30.77	23.08	53.85	76.92	19.23	65.38	26.92	26
Aggre	76.14	62.50	61.36	65.91	88.64	54.55	68.18	19.32	88
Panel B: the percentage of identification failure and misspecification at 5% false discovery rate									
News	99.40	90.42	67.07	54.49	93.41	43.11	68.26	0.60	167
Consumption	100.00	90.30	35.76	40.30	92.12	19.09	73.33	9.70	330
Macro	98.74	86.71	58.53	59.07	91.56	37.34	68.40	3.95	557
Liquidity	97.32	75.10	62.45	38.31	90.80	20.31	61.30	1.92	261
Sentiment	99.59	76.33	62.45	53.88	82.45	42.04	65.71	2.86	245
Intermediary	96.15	82.69	38.46	54.81	93.27	26.92	71.15	4.81	104
Volatility	88.46	38.46	30.77	57.69	73.08	23.08	61.54	7.69	26
Aggre	93.18	67.05	63.64	67.05	81.82	55.68	65.91	10.23	88

Table 13: List of priced factors in original models
This table lists all factors with risk premium uniformly significant in t-statm CSR GLS, t-statm HJ, and t-statm CU-GMM. It also reports the p values from FM, GX and Burnside's approaches.

Ref	ModelName	TestAssets	FactorNames	T	NumAssets	FM	Misspecification-robust			GX	Burnside
							GLS	HJ	CUGMM		
1995 Elton Gruber Blake	Quart_inf+rGDP	HKMCDS	MICHI_inf_exp_quart	48	20	0.12	0.01	0.04	0.00	0.86	0.43
2004 Campbell Vuolteenaho	N_dr+N_cf	US_anoms_porf_24	N_cf	462	24	0.00	0.04	0.04	0.01	0.00	0.00
2005 Julliard Parker	Annual_PJCon	HKMCommodity	Con	26	23	0.35	0.01	0.02	0.00	0.58	
2005 Julliard Parker	Quart_PJCon	HXZ_24_anoms_porf_48	Con	208	48	0.00	0.00	0.01	0.00	0.06	
2006 Baker Wurgler	Inv_sent+Inv_sentxMKT	FF25_ME_BM	Inv_sent	680	25	0.98	0.02	0.03	0.01	0.41	0.42
2006 Baker Wurgler	Inv_sent	FF25_ME_BM	Inv_sent	684	25	0.40	0.00	0.00	0.02	0.36	
2006 Baker Wurgler	Inv_sent_orth+Inv_orthxMKT	FF25_ME_BM	Inv_sent_orth	680	25	1.00	0.02	0.03	0.00	0.58	0.43
2006 Baker Wurgler	Inv_sent_orth	FF25_ME_BM	Inv_sent_orth	684	25	0.54	0.01	0.01	0.00	0.54	
2006 Sadka	Illiq_Perm+FF3	HKMUSBonds	Illiq_Perm	345	20	0.01	0.02	0.02	0.04	0.46	0.07
2006 Sadka	Illiq_Perm+MKT	HKMUSBonds	Illiq_Perm	345	20	0.03	0.02	0.02	0.02	0.46	0.20
2006 Sadka	Illiq_Trans+Illiq_Perm+MKT	HKMUSBonds	Illiq_Perm	345	20	0.01	0.02	0.04	0.02	0.46	0.19
2011 Berkman Jacobsen Lee	exp_disrisk+unexp_disrisk	HKMCDS	exp_crisis	70	20	0.01	0.00	0.00	0.00	0.52	0.29
2013 Boguth Kuehn	Quart_Con+D_mu+D_sig	HKMCDS	Con	36	20	0.47	0.00	0.01	0.01	0.34	0.50
2013 Boguth Kuehn	Quart_Con+D_mu+D_sig_ext	HKMCDS	Con	36	20	0.07	0.00	0.03	0.04	0.34	0.20
2015 Da Engelberg Gao	FEARS30_ext	HKMCDS	VIX	90	20	0.00	0.00	0.05	0.04	0.05	0.02
2016 Bali Brown Tang	Macro_uncert+MKT	HXZ_24_anoms_porf_48	Macro_uncert	660	48	0.00	0.03	0.04	0.02	0.43	0.07
2016 Bali Brown Tang	Macro_uncert	HXZ_24_anoms_porf_48	Macro_uncert	660	48	0.10	0.03	0.04	0.01	0.43	
2017 Kroencke	Annual_PJ-N&S	HKMCommodity	PJ_N_S	26	23	0.35	0.01	0.02	0.00	0.58	
2017 Kroencke	Annual_UNFIL-N&S	HXZ_24_anoms_porf_48	UNFIL_N_S	52	48	0.06	0.00	0.04	0.03	0.84	
2018 Chen Eaton Paye	adj_PC+logvol	HKMCDS	logvol	143	20	0.00	0.00	0.00	0.03	0.08	0.01
2018 Chen Eaton Paye	adj_logcorwin+logvol	HKMCDS	logvol	143	20	0.16	0.00	0.01	0.01	0.08	0.18
2018 Chen Eaton Paye	adj_PC+logvol	HKMOptions	adj_PC	310	18	0.00	0.00	0.00	0.01	0.65	0.00
2018 Chen Eaton Paye	adj_logamito+logvol	HKMOptions	adj_logamito	310	18	0.00	0.00	0.00	0.00	0.04	0.00
2018 Chen Eaton Paye	adj_loghm+logvol	HKMOptions	adj_loghm	310	18	0.00	0.01	0.03	0.04	0.86	0.00
2018 Chen Eaton Paye	adj_logtick+logvol	HKMOptions	adj_logtick	310	18	0.00	0.00	0.00	0.05	0.54	0.00
2020 Boonsa et al.	NRC+FF3	FF25_ME_MOM	NRC	611	25	0.00	0.00	0.00	0.00	0.07	0.00
2020 Boonsa et al.	NRC+Inf	FF25_ME_MOM	NRC	611	25	0.00	0.00	0.00	0.00	0.07	0.04
2020 Boonsa et al.	monthly_NRC	FF25_ME_MOM	NRC	611	25	0.00	0.00	0.00	0.00	0.07	
2022 Ardia et al.	innov_MCCI+MKT	HKMUSBonds	inn_MCCI	108	20	0.01	0.02	0.05	0.03	0.44	0.19
2022 Chen Tang Yao Zhou	Attention_PCA	FF25_ME_INV	Attention_PCA	456	25	0.13	0.04	0.04	0.02	0.15	
2022 Chen Tang Yao Zhou	Attention_PCA	HKMCDS	Attention_PCA	143	20	0.00	0.01	0.02	0.02	0.16	
2022 Chen Tang Yao Zhou	Attention_PLS	US_anoms_porf_24	Attention_PLS	456	24	0.00	0.01	0.01	0.01	0.14	

Table 14: Correlation between PCs of non-traded factors and popular pricing factors
This table presents the correlation between the PCs of non-traded factors ($eqPC1-5$) and some popular pricing factor models: Fama and French (2015)'s 5-factors (MKT, SMB, HML, RMW, CMA), Giglio and Xiu (2021)'s 7 factors (GX1-7) and Lettau and Pelger (2020)'s 5 factors (RP-PC1-5). The rows stand for the first group of factors while the columns stand for the others.

	MKT	SMB	HML	RMW	CMA
$eqPC1$	-0.08	-0.24	-0.52	-0.26	-0.45
$eqPC2$	-0.89	0.08	0.32	0.14	0.44
$eqPC3$	0.28	0.48	0.45	-0.28	0.19
$eqPC4$	0.23	0.02	-0.23	0.18	-0.15
$eqPC5$	-0.22	-0.81	0.52	0.49	0.36

	GX1	GX2	GX3	GX4	GX5	GX6	GX7
	-0.08	0.03	-0.06	-0.03	0.02	0.04	0.06
	-0.01	-0.03	0.1	0.02	0.01	-0.14	0.12
	0.06	-0.05	0.06	0.14	-0.1	0.01	-0.12
	0.04	0.05	0.05	0.1	-0.09	-0.04	-0.03
	-0.05	-0.06	-0.02	0	-0.03	-0.03	0.12

	RP-PC1	RP-PC2	RP-PC3	RP-PC4	RP-PC5
	-0.03	-0.77	-0.16	0.11	0.52
	-0.84	0.15	0.37	-0.2	0.29
	0.36	-0.35	0.85	0.06	0
	0.17	-0.17	-0.08	0.45	0.05
	-0.34	0.33	0.12	0.72	-0.27

Table 15: Identification and misspecification of PCs of non-traded factors
This table presents the p values for model misspecification tests (HJ distance and J test) and identification tests (KP test, F test, CD test and CF test), as well as the associated number of time series observations (T), the number of test portfolios (K) and the number of PCs (K). *** significant at 1% signficiance level. ** significant at 5% level. * significant at 1% level.

TestAssets	HJ distance	J test	KP test	F test		CD test	CF test	T	N	K
				Incl. intercept	Excl. intercept					
FF25_BEME_OP	0.96	0.99	0.05*	0.01**	0.00***	0.01**	0.00***	552	25	5
FF25_ME_BM	0.04**	0.29	0.15	0.00***	0.00***	0.01**	0.00***	552	25	5
FF25_ME_INV	0.00***	0.50	0.51	0.04**	0.00***	0.11	0.00***	552	25	5
FF25_ME_MOM	0.00***	0.46	0.13	0.01**	0.00***	0.04**	0.00***	552	25	5
HKMCDS	0.00***	0.82	0.75	0.66	0.67	0.55	0.43	143	20	5
HKMCommodity	0.26	0.93	0.91	0.90	0.92	0.82	0.00***	316	23	5
HKMFX	0.15	0.91	0.82	0.86	0.89	0.89	0.33	407	12	5
HKMOptions	0.00***	0.80	0.79	0.78	0.83	0.75	0.53	310	18	5
HKMSovBonds	0.00***	0.00***	0.64	0.70	0.89	0.66	0.55	196	6	5
HKMUSBonds	0.00***	0.96	0.97	0.96	0.97	0.97	0.87	444	20	5
HXZ_24_anoms_porf_48	0.01**	0.13	0.00***	0.00***	0.00***	0.00***	0.00***	552	48	5
US_anoms_porf_24	0.01**	0.10	0.00***	0.00***	0.00***	0.00***	0.00***	552	24	5
Weber10_duration	0.11	0.86	0.92	0.85	0.74	0.91	0.55	534	10	5

Table 16: Pricing results of PCs of non-traded factors for equity portfolios

This table presents the the risk premium (gamma) estimates in percentage from the standard Fama-MacBeth regression and Giglio and Xiu (2021)'s three-pass procedure as well as associated p values for the first five PCs extracted from monthly non-traded factors relative to the sets of test assets used in our paper. Burnside (2016) bootstrap risk premia confidence intervals is also reported. Moreover, it also reports the SDF parameter (lambda) and the associated misspecification-robust CU-GMM procedure (see eq.2 and section 3). *** significant at 1% signficiance level. ** significant at 5% level. * significant at 1% level.

	FF25_BEME_OP					FF25_ME_BM					FF25_ME_INV				
	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5
gamma(FM)	-0.32	-0.09	-0.14	-0.18	0.01	-0.66	-0.22	-0.31	-0.20	-0.10	-0.43	-0.13	-0.07	-0.21	0.02
pval (FM)	0.03**	0.52	0.13	0.05	0.92	0.00***	0.03**	0.00***	0.04**	0.06*	0.01***	0.23	0.51	0.08	0.78
gamma(GX)	-0.08	0.06	0.07	-0.07	0.07	-0.10	0.12	0.05	-0.01	0.01	-0.04	0.18	0.04	-0.05	0.01
pval(GX)	0.06*	0.39	0.11	0.16	0.02**	0.02**	0.10	0.25	0.83	0.77	0.54	0.11	0.43	0.36	0.80
Burnside CI	[-0.49,0.10]	[-0.31,0.19]	[-0.26,0.12]	[-0.37,0.01]	[-0.12,0.27]	[-0.79,-0.11]	[-0.31,0.19]	[-0.45,0.07]	[-0.35,0.17]	[-0.18,0.07]	[-0.64,0.17]	[-0.30,0.28]	[-0.23,0.36]	[-0.39,0.14]	[-0.08,0.19]
lambda(cugmm)	58.44	33.42	19.00	8.13	4.57	176.01	75.32	115.67	69.08	50.54	-576.17	-364.43	-371.64	-224.32	-159.78
pval(cugmm)	0.01**	0.06*	0.16	0.49	0.75	0.26	0.47	0.31	0.31	0.52	0.63	0.62	0.61	0.62	0.56

	FF25_ME_MOM					HXZ_24_anoms_porf_48					US_anoms_porf_24				
	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5
gamma(FM)	2.40E-03	0.18	-0.26	0.47	-0.25	-0.33	-0.11	-0.02	0.02	-0.01	-0.30	-0.09	0.00	0.07	0.05
pval (FM)	0.98	0.10	0.00***	0.00***	0.00***	0.00***	0.09	0.73	0.75	0.91	0.00***	0.08*	0.97	0.38	0.36
gamma(GX)	-0.05	0.20	-0.12	-0.05	-0.10	-0.16	-0.10	0.09	0.03	0.08	-0.18	-0.09	0.06	0.02	0.11
pval(GX)	0.51	0.09*	0.02**	0.36	0.04**	0.00***	0.14	0.03**	0.42	0.04**	0.00***	0.09*	0.13	0.47	0.01
Burnside CI	[-0.36,0.20]	[-0.17,0.38]	[-0.39,0]	[0.03,0.73]	[-0.38,-0.01]	[-0.46,-0.13]	[-0.24,0.08]	[-0.12,0.11]	[-0.16,0.18]	[-0.11,0.11]	[-0.45,-0.13]	[-0.2,0.02]	[-0.12,0.13]	[-0.12,0.22]	[-0.08,0.16]
lambda(cugmm)	-108.90	-105.99	115.70	-156.00	128.09	62.33	27.71	11.55	-17.85	2.64	73.63	26.17	33.99	15.37	7.34
pval(cugmm)	0.13	0.11	0.10	0.08*	0.11	0.00***	0.07*	0.29	0.25	0.91	0.00***	0.01**	0.03**	0.27	0.39

	Weber10_duration				
	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5
gamma(FM)	-0.76	-1.09	0.78	-0.26	0.35
pval (FM)	0.04**	0.12	0.02**	0.54	0.33
gamma(GX)	-0.49	-0.44	0.61	0.19	-0.19
pval(GX)	0.01**	0.01**	0.01**	0.04**	0.04**
Burnside CI	[-1.48,0.74]	[-2.22,1.73]	[-0.43,1.58]	[-1.48,1.43]	[-0.99,1.38]
lambda(cugmm)	221.28	827.58	-238.31	510.19	-418.47
pval(cugmm)	0.74	0.75	0.70	0.77	0.77

Table 17: Pricing results of PCs of non-traded factors for non-equity portfolios

This table presents the the risk premium (gamma) estimates in percentage from the standard Fama-MacBeth regression and Giglio and Xiu (2021)'s procedure as well as associated p values for the first five PCs extracted from monthly non-traded factors relative to the sets of test assets used in our paper. Burnside (2016) bootstrap risk premia confidence intervals is also reported. Moreover, it also reports the SDF parameter (lambda) and the associated misspecification-robust CU-GMM procedure (see eq.2 and section 3). *** significant at 1% significance level. ** significant at 5% level. * significant at 1% level.

	HKMCDS					HKMCommodity					HKMFX				
	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5
gamma(FM)	0.18	0.69	1.64	0.01	0.38	0.03	0.24	0.51	-0.35	-0.33	-1.50	-0.34	0.30	1.35	-1.36
pval (FM)	0.62	0.00***	0.00***	0.97	0.19	0.93	0.43	0.30	0.22	0.21	0.03**	0.32	0.42	0.05*	0.01**
gamma(GX)	0.01	-0.09	0.25	0.03	0.00	-0.01	-0.02	0.01	-0.02	-0.02	-0.02	-0.01	0.03	0.02	-0.03
pval(GX)	0.93	0.12	0.02**	0.64	0.87	0.48	0.51	0.71	0.14	0.20	0.57	0.70	0.29	0.43	0.24
Burnside CI	[-1.42,1.92]	[-0.32,0.98]	[0.15,2.01]	[-1.42,2.18]	[-0.91,1.04]	[-0.66,0.56]	[-0.44,0.58]	[-0.67,0.86]	[-0.76,0.35]	[-0.78,0.25]	[-2.55,2.04]	[-1.91,1.57]	[-1.25,2.11]	[-2.6,2.99]	[-3.06,1.15]
lambda(cugmm)	516.68	-164.40	-295.43	842.44	-1582.43	-69.99	-155.44	-222.86	122.19	60.77	703.16	248.73	-64.62	-650.51	108.97
pval(cugmm)	0.79	0.85	0.79	0.81	0.81	0.68	0.45	0.49	0.46	0.57	0.60	0.60	0.86	0.60	0.85

	HKMOptions					HKMSovBonds					HKMUSBonds				
	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5
gamma(FM)	0.11	0.46	-0.45	1.20	-1.53	1.57	0.30	0.90	-2.11	-2.67	-0.68	0.34	0.11	1.09	-0.48
pval (FM)	0.84	0.04**	0.17	0.04**	0.00***						0.04**	0.11	0.56	0.05*	0.36
gamma(GX)	-0.14	0.52	-0.18	-0.08	-0.16	0.02	-0.11	0.08	0.01	-0.03	-0.04	-0.08	0.11	0.04	-0.01
pval(GX)	0.37	0.46	0.49	0.57	0.18						0.08	0.01**	0.01**	0.03**	0.64
Burnside CI	[-3.15,2.31]	[-0.48,1.67]	[-1.53,1.85]	[-1.30,2.99]	[-3.43,1.30]	[-12.50,13.44]	[-8.46,8.52]	[-16.06,22.8]	[-19.16,20.8]	[-25.67,15.28]	[-1.44,1.02]	[-0.88,0.81]	[-0.52,1.01]	[-1.80,2.76]	[-1.59,1.84]
lambda(cugmm)	2865.19	-961.07	-2206.77	92.98	-1103.54						1683.34	695.92	-10.94	3599.95	-918.35
pval(cugmm)	0.90	0.90	0.90	0.97	0.90						0.92	0.91	0.99	0.92	0.92

C Figures

C.1 Identification of CAPM and FF3

Figure 2: The figures report the p values of rank tests for CAPM relative to different equity portfolios at monthly, quarterly and yearly frequencies. Four rank tests are KP test, F test, CD test and CF rank test. The null hypothesis of KP, F and CD tests is $\mathcal{H}_0^{rk}: rk(\beta) = K-1$, and the CF statistic is tested on $\mathcal{H}_0^{rk}: rk(\beta) \leqslant K-1$.

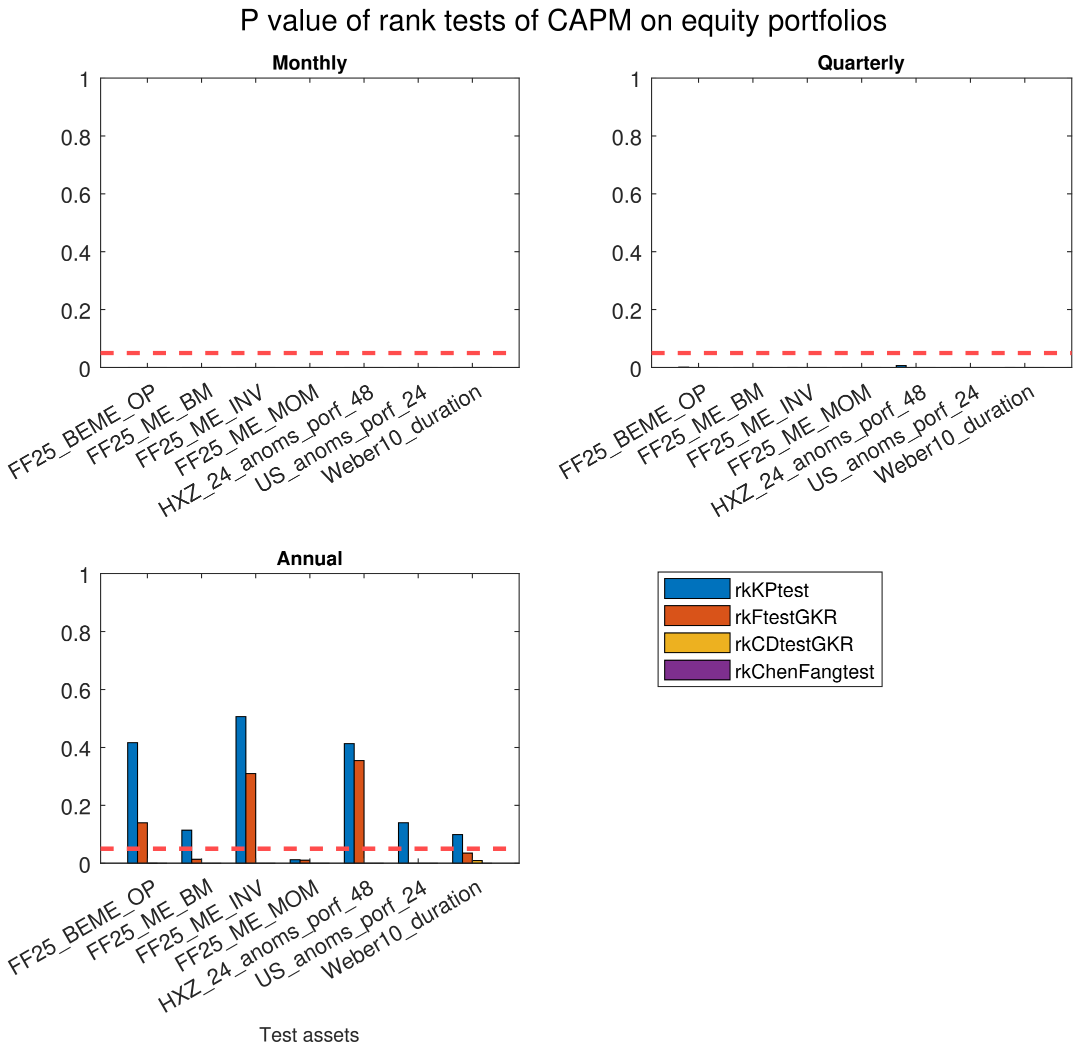


Figure 3: The figures report the p values of rank tests for CAPM relative to different non-equity portfolios at monthly, quarterly and yearly frequencies. Four rank tests are KP test, F test, CD test and CF rank test. The null hypothesis of KP, F and CD tests is $\mathcal{H}_0^{rk} : rk(\beta) = K - 1$, and the CF statistic is tested on $\mathcal{H}_0^{rk} : rk(\beta) \leqslant K - 1$.

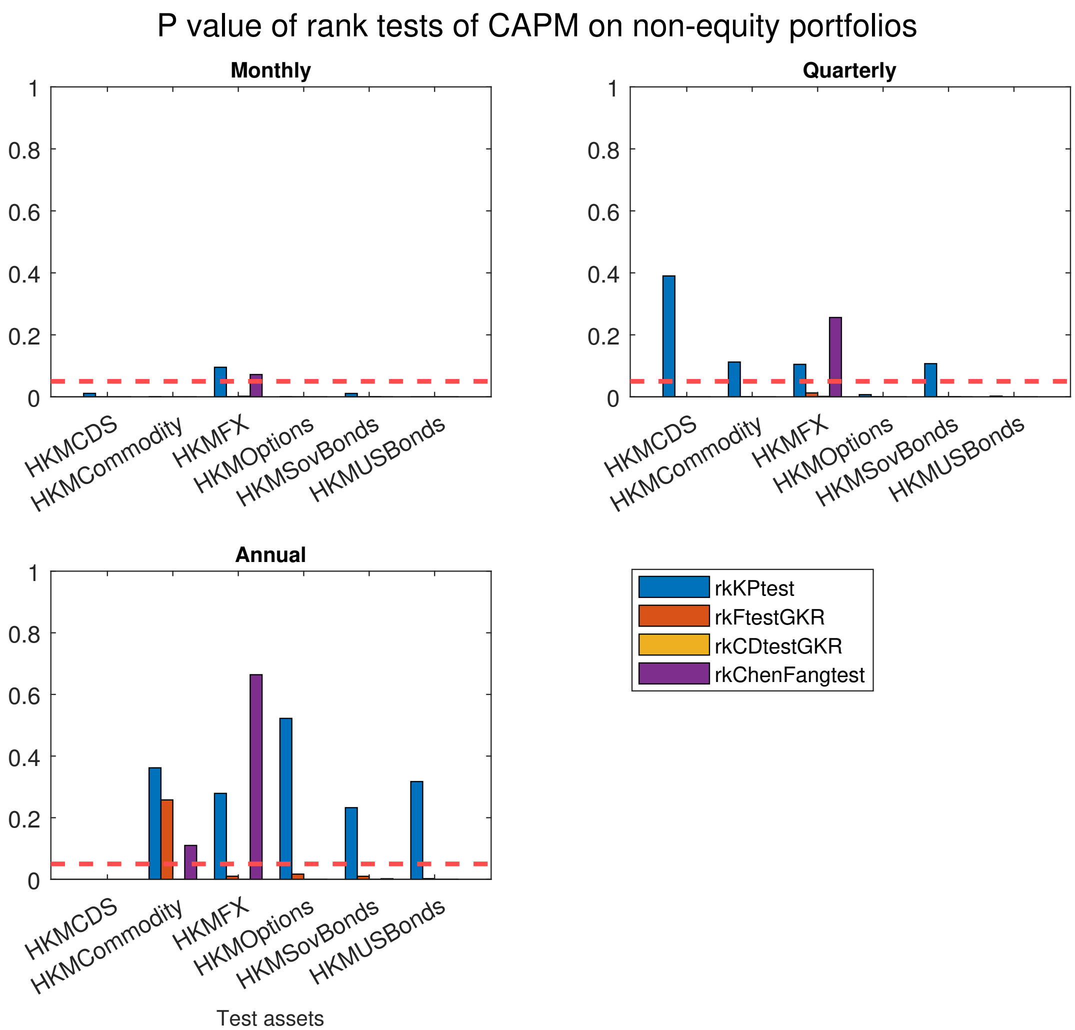


Figure 4: The figures report the p values of rank tests for Fama-French three-factor model relative to equity portfolios at monthly, quarterly and yearly frequencies. Four rank tests are KP test, F test, CD test and CF rank test. The null hypothesis of KP, F and CD tests is $\mathcal{H}_0^{rk} : rk(\beta) = K - 1$, and the CF statistic is tested on $\mathcal{H}_0^{rk} : rk(\beta) \leqslant K - 1$.

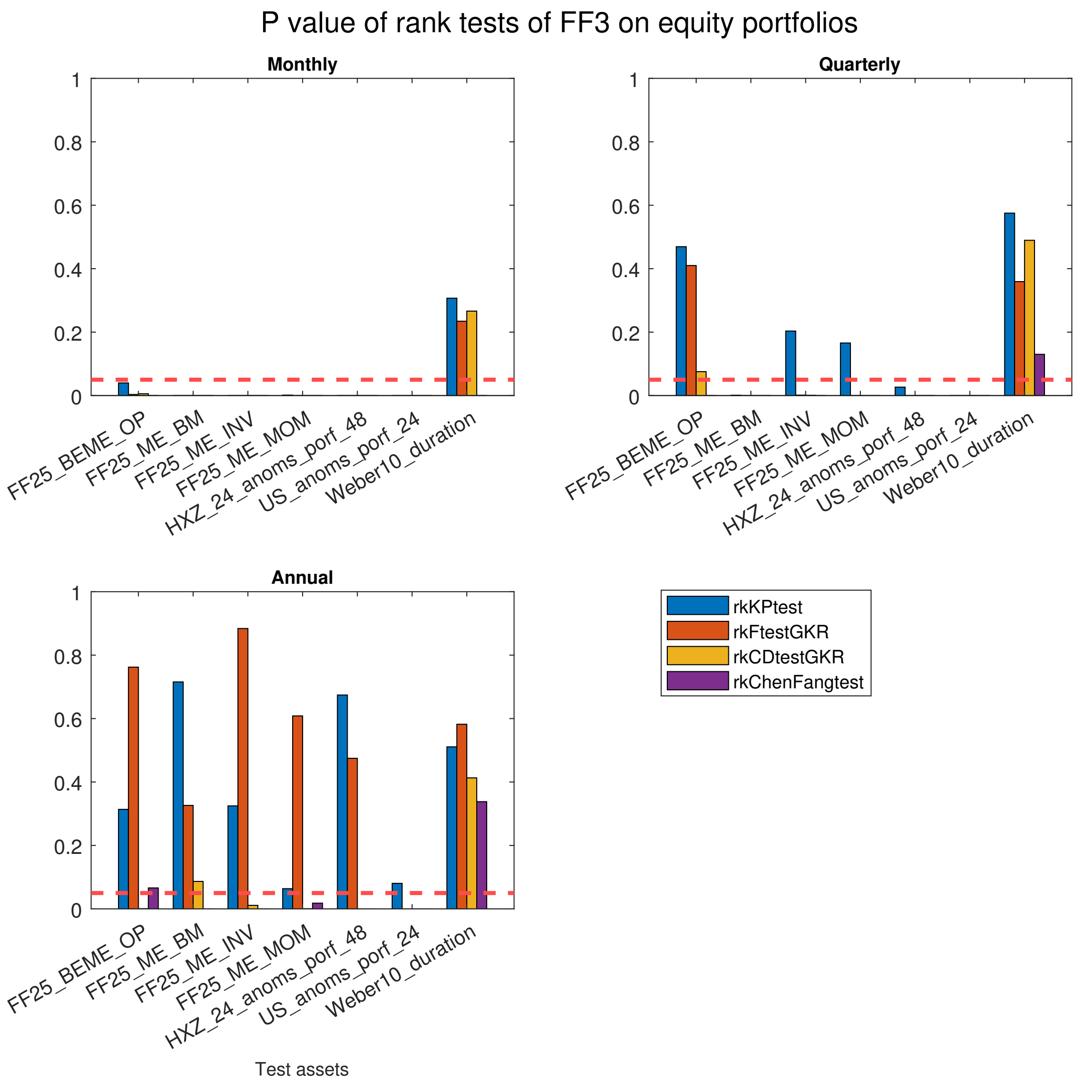
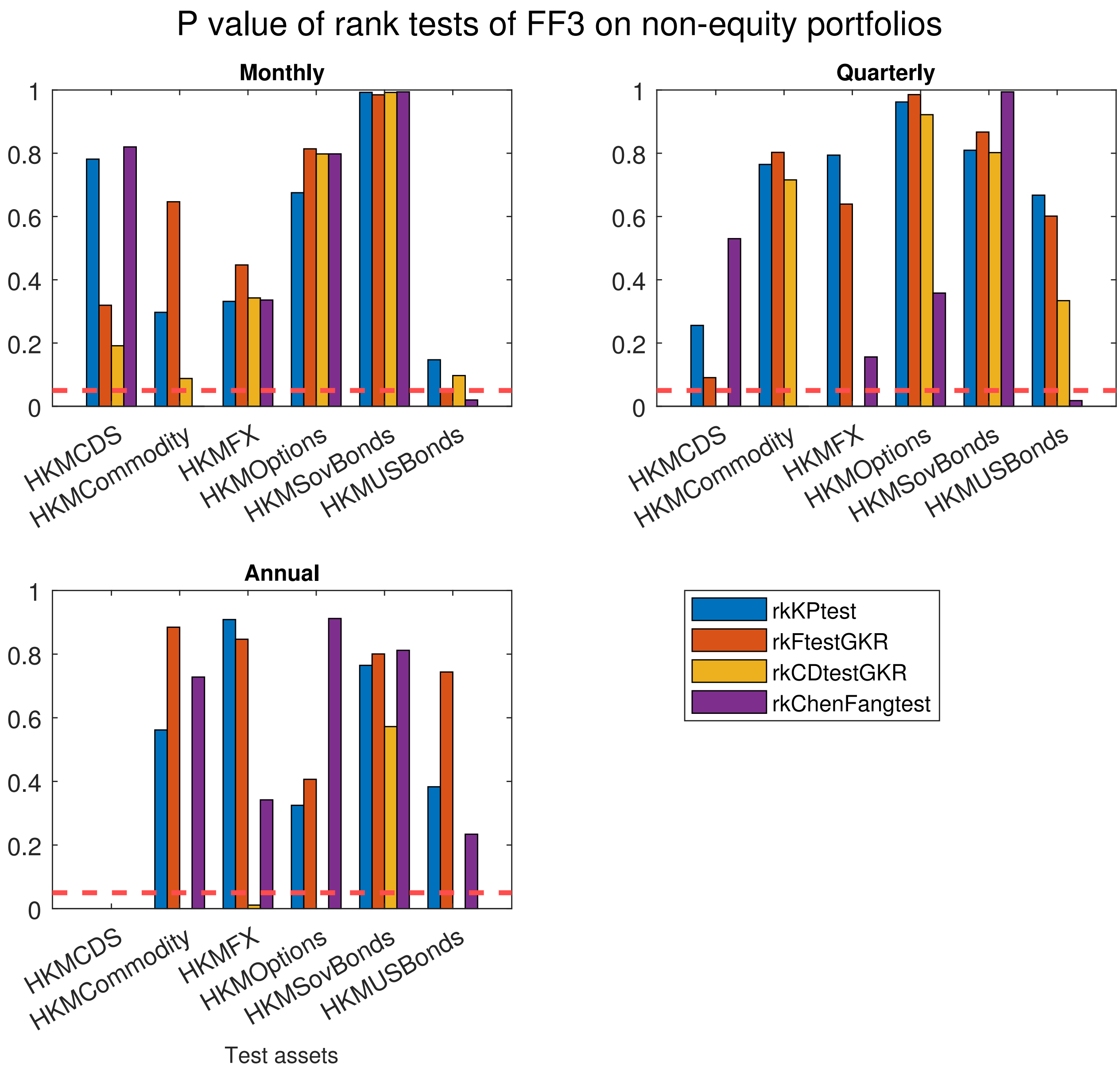


Figure 5: The figures report the p values of rank tests for Fama-French three-factor model relative to non-equity portfolios at monthly, quarterly and yearly frequencies. Four rank tests are KP test, F test, CD test and CF rank test. The null hypothesis of KP, F and CD tests is $\mathcal{H}_0^{rk}: rk(\beta) = K-1$, and the CF statistic is tested on $\mathcal{H}_0^{rk}: rk(\beta) \leqslant K-1$.



C.2 Identification

Figure 6: Identification failure of original models

This figure shows the percentage of identification failure in model specifications that are proposed by the original papers. We use four rank tests: Kleibergen and Paap (2006) rank test (rkKPtest), Gospodinov and Robotti (2021b) finite-sample rank test(rkFtestGKR), Gospodinov et al. (2017) Cragg-Donald rank test (rkCDtestGKR), and Chen and Fang (2019) rank test (rkChenFangtest). We also consider the most conservative situation where we can conclude a model is not under-identified only when all of four test statistics can reject the null of rank deficiency, and the favorable situation where any of test statistic can reject. We also show the percentages calculated with respect to $T \leq 100$, $100 < T \leq 300$ and $T > 300$.

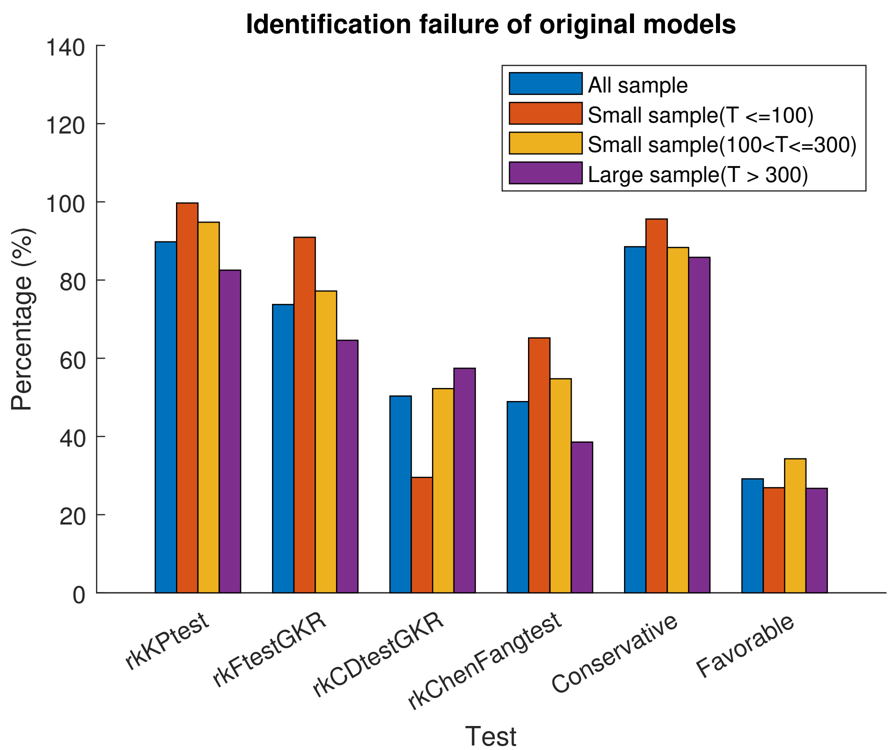


Figure 7: Identification failure of original models corrected for multiple hypothesis testing
This figure shows the percentage of identification failure in model specifications we collect from the original papers. We control false discovery rate at 5% using Benjamini and Hochberg (1995) method. We use four rank tests: Kleibergen and Paap (2006) rank test (rkKPtest), Gospodinov and Robotti (2021b) finite-sample rank test(rkFtestGKR), Gospodinov et al. (2017) Cragg-Donald rank test (rkCDtestGKR), and Chen and Fang (2019) rank test (rkChenFangtest). We also consider the most conservative situation where we can conclude a model is not under-identified only when all of three test statistics can reject the null of rank deficiency, and the favorable situation where any of test statistic can reject. We also present the percentages across length of time series observations $T \leq 100$, $100 < T \leq 300$ and $T > 300$..

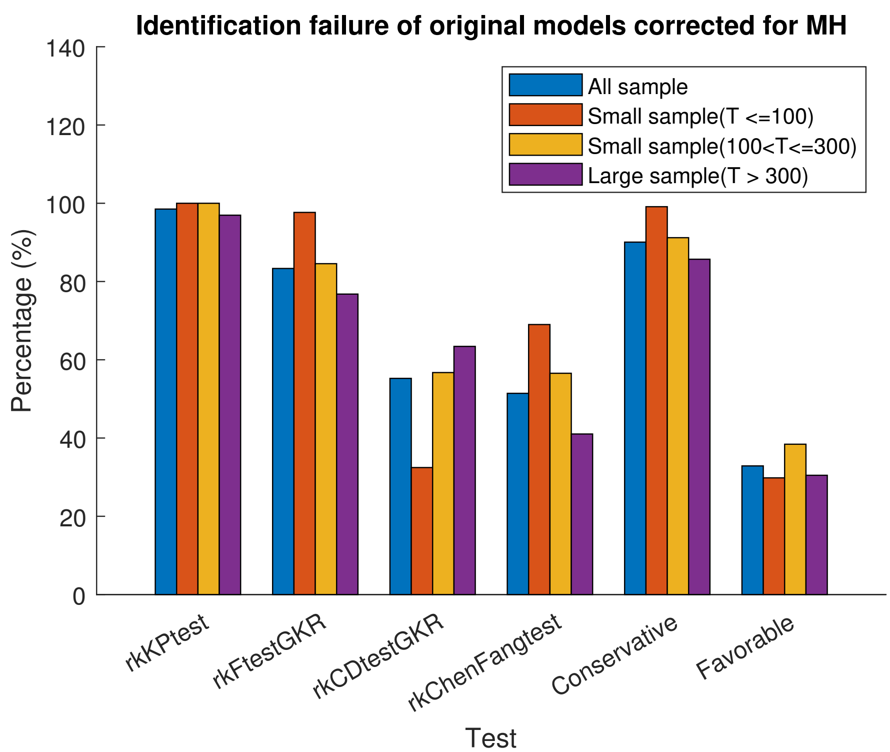


Figure 8: This figure shows the percentage of identification failure for model specifications with and without market factor. We use five rank tests: Kleibergen and Paap (2006) rank test (rkKPtest), Gospodinov and Robotti (2021b) finite-sample rank test(rkFtestGKR and rkFtestGKR_no_int) with and without intercept, Gospodinov et al. (2017) Cragg-Donald rank test (rkCDtestGKR), and Chen and Fang (2019) rank test (rkChenFangtest).

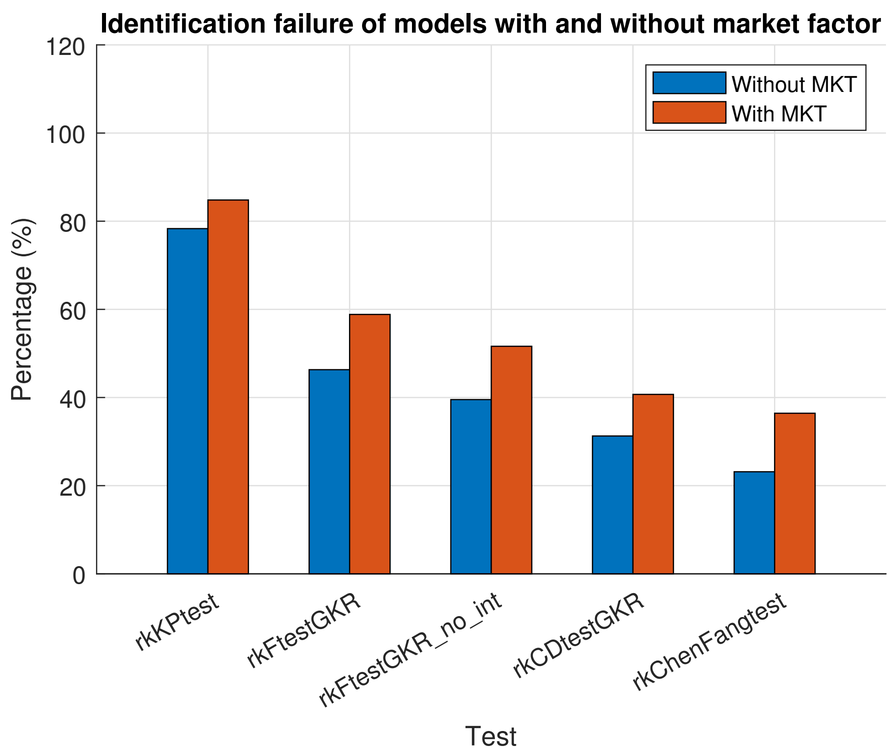


Figure 9: This figure presents the percentage of identification failure for different sets of equity portfolios. We control false discovery rate at 5% using Benjamini and Hochberg (1995) method. We use four rank tests: Kleibergen and Paap (2006) rank test (rkKPtest), Gospodinov and Robotti (2021b) finite-sample rank test(rkFtestGKR), Gospodinov et al. (2017) Cragg-Donald rank test (rkCDtestGKR), and Chen and Fang (2019) rank test (rkChenFangtest).

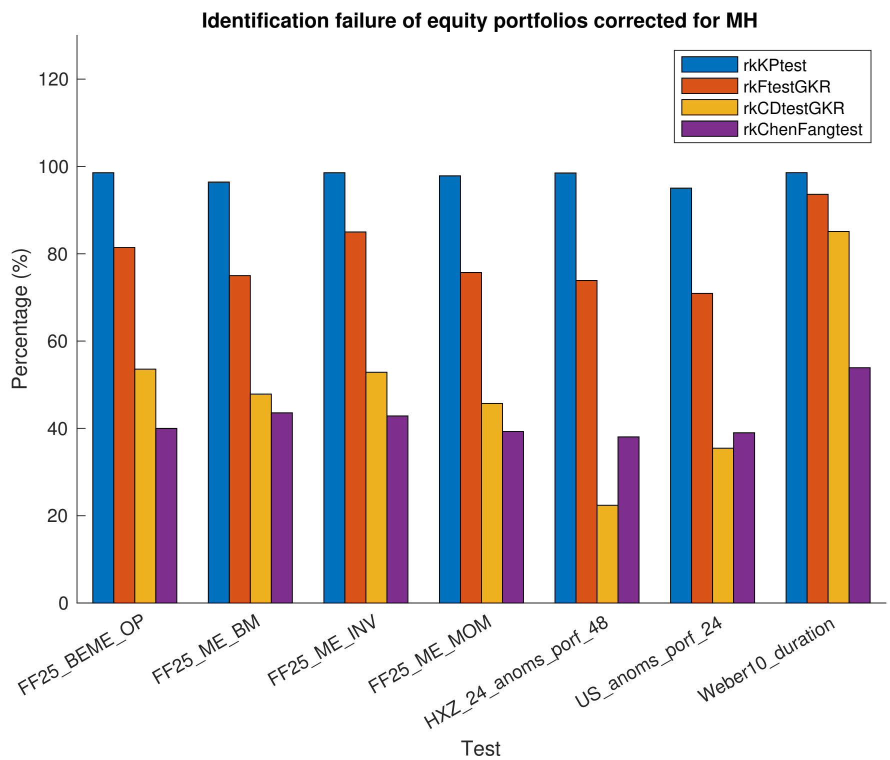


Figure 10: This figure presents the percentage of identification failure for different sets of non-equity portfolios. We control false discovery rate at 5% using Benjamini and Hochberg (1995) method. We use four rank tests: Kleibergen and Paap (2006) rank test (rkKPtest), Gospodinov and Robotti (2021b) finite-sample rank test(rkFtestGKR), Gospodinov et al. (2017) Cragg-Donald rank test (rkCDtestGKR), and Chen and Fang (2019) rank test (rkChenFangtest).

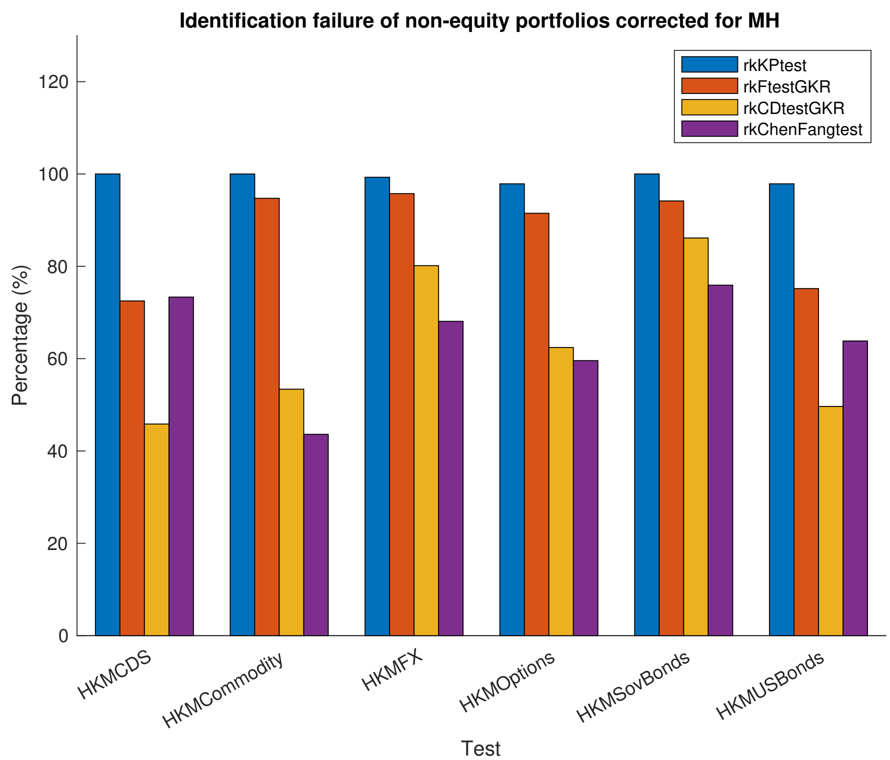
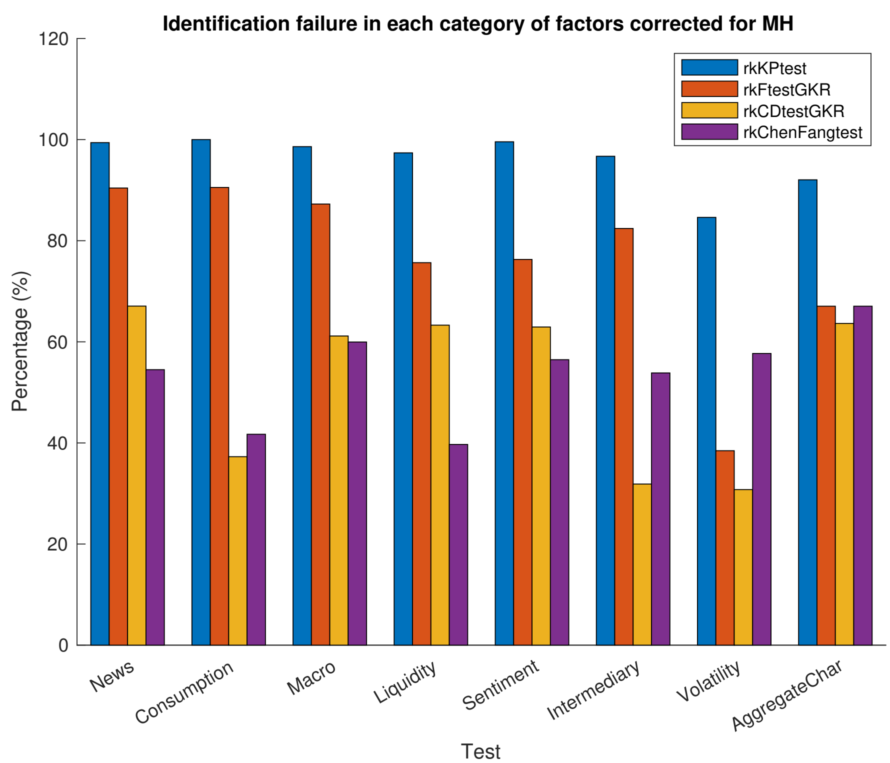


Figure 11: This figure presents the percentage of identification failure for different categories of factors. We control false discovery rate at 5% using Benjamini and Hochberg (1995) method. We use four rank tests: Kleibergen and Paap (2006) rank test (rkKPtest), Gospodinov and Robotti (2021b) finite-sample rank test(rkFtestGKR), Gospodinov et al. (2017) Cragg-Donald rank test (rkCDtestGKR), and Chen and Fang (2019) rank test (rkChenFangtest).



C.3 Misspecification

Figure 12: This figure reports the fraction of model misspecification in over-identifying restriction J test and Hansen-Jagannathan distance test. We also compare the percentage across number of time series observations $T \leq 100$, $100 < T \leq 300$ and $T > 300$..

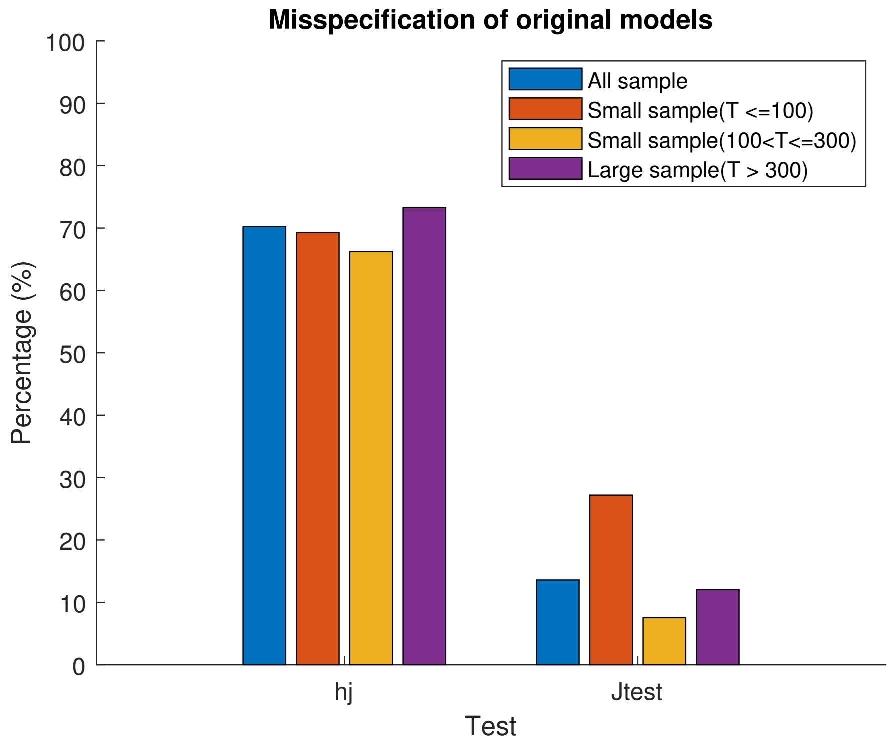


Figure 13: This figure reports the fraction of model misspecification for equity portfolios in over-identifying restriction J test and Hansen-Jagannathan distance test.

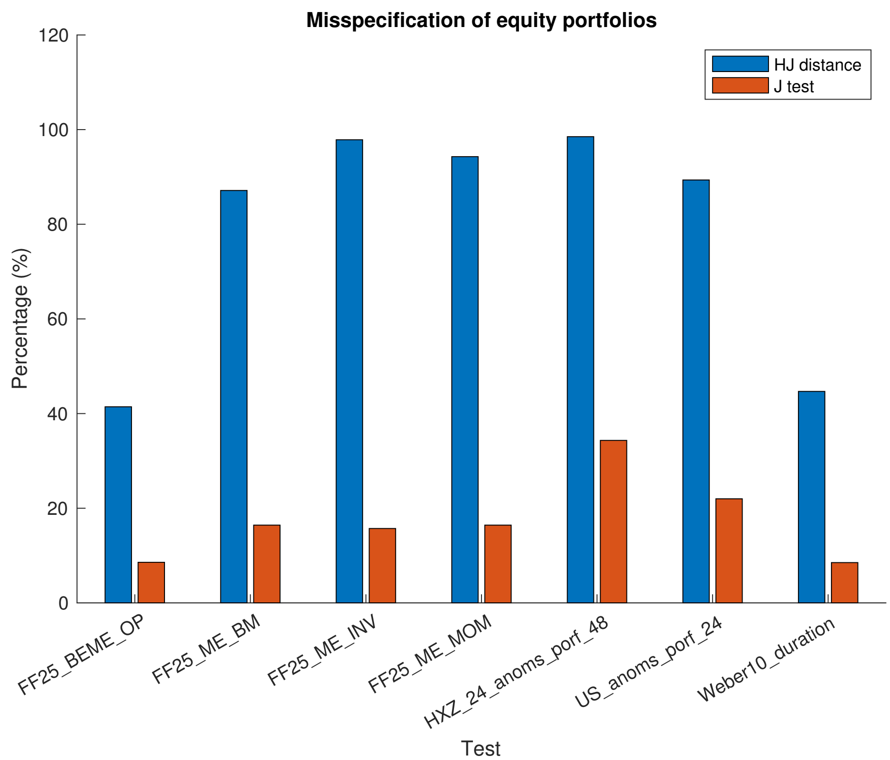
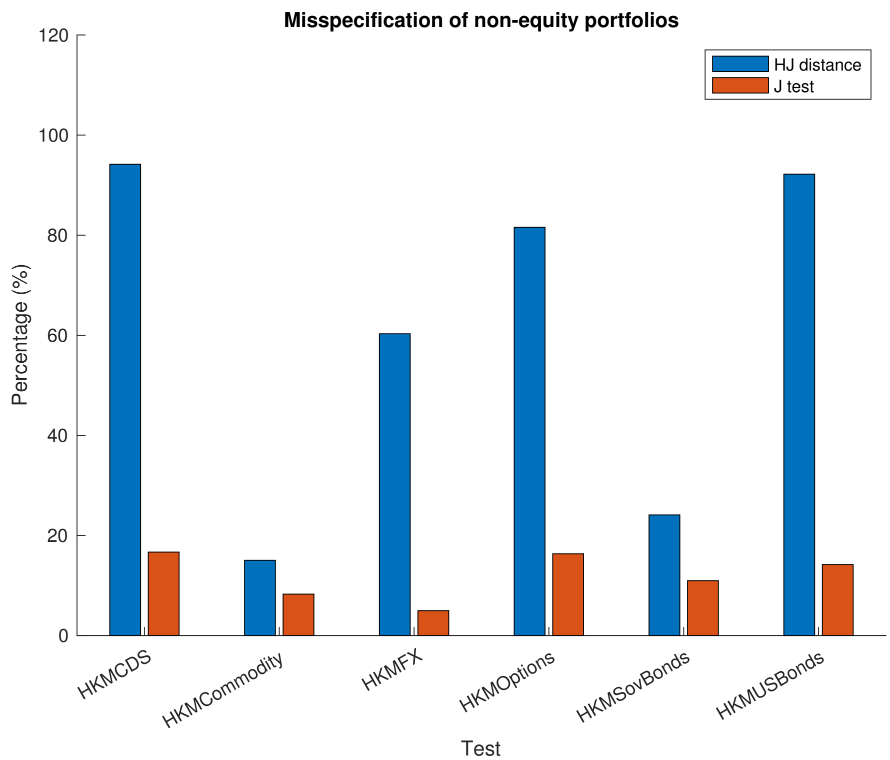
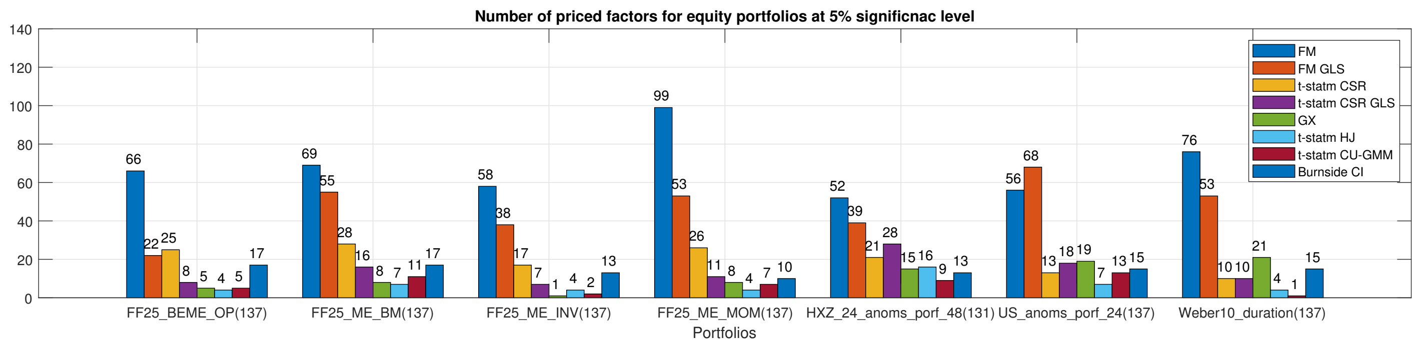


Figure 14: This figure reports the fraction of model misspecification for non-equity portfolios in over-identifying restriction J test and Hansen-Jagannathan distance test.



C.4 Cross-sectional risk premium

Figure 15: The top figure shows the number of priced factors for equity portfolios at 5% significance level using the Fama-MacBeth approach, generalized Shanken's EIV allowing for conditional heteroskedasticity (Jagannathan and Wang (1998)), misspecification-robust t-statistic (and their GLS estimators), three-pass approach by Giglio and Xiu (2021), misspecification-robust t-statistic based on Hansen-Jagannathan distance (Gospodinov et al. (2014)), misspecification-robust t-statistic in linear SDF estimated using CU-GMM. (Gospodinov et al. (2017)), and Burnside's bootstrap confidence interval (Burnside (2011)). For some papers, we include multiple model specifications with different control factors that are used in the original paper. We consider a factor as priced if it has a non-zero price of risk in at least one of the specifications. Total number of factors that can be tested for each set of portfolios is indicated in a parenthesis associated with each set of portfolios. The bottom figure shows the number of priced factors when we control the false discovery rate at 5% for each test statistic in each type of test portfolios.



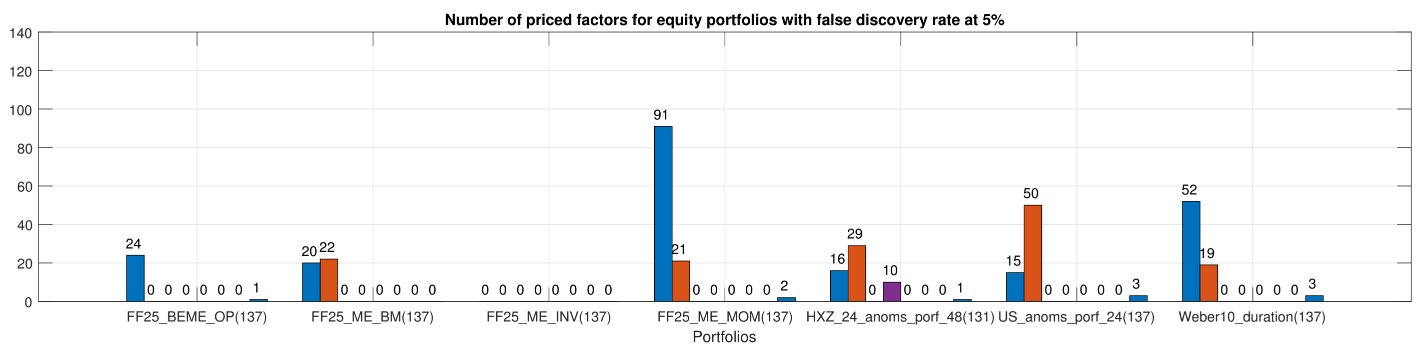


Figure 16: The top figure shows the number of priced factors for non-equity portfolios at 5% significance level using the Fama-MacBeth approach, generalized Shanken's EIV allowing for conditional heteroskedasticity (Jagannathan and Wang (1998)), misspecification-robust t-statistic (and their GLS estimators), three-pass approach by Giglio and Xiu (2021), misspecification-robust t-statistic based on Hansen-Jagannathan distance (Gospodinov et al. (2014)), misspecification-robust t-statistic in linear SDF estimated using CU-GMM. (Gospodinov et al. (2017)), and Burnside's bootstrap confidence interval (Burnside (2011)). For some papers, we include multiple model specifications with different control factors that are used in the original paper. We consider a factor as priced if it has a non-zero price of risk in at least one of the specifications. Total number of factors that can be tested for each set of portfolios is indicated in a parenthesis associated with each set of portfolios. The bottom figure shows the number of priced factors when we control the false discovery rate at 5% for each test statistic in each type of test portfolios.

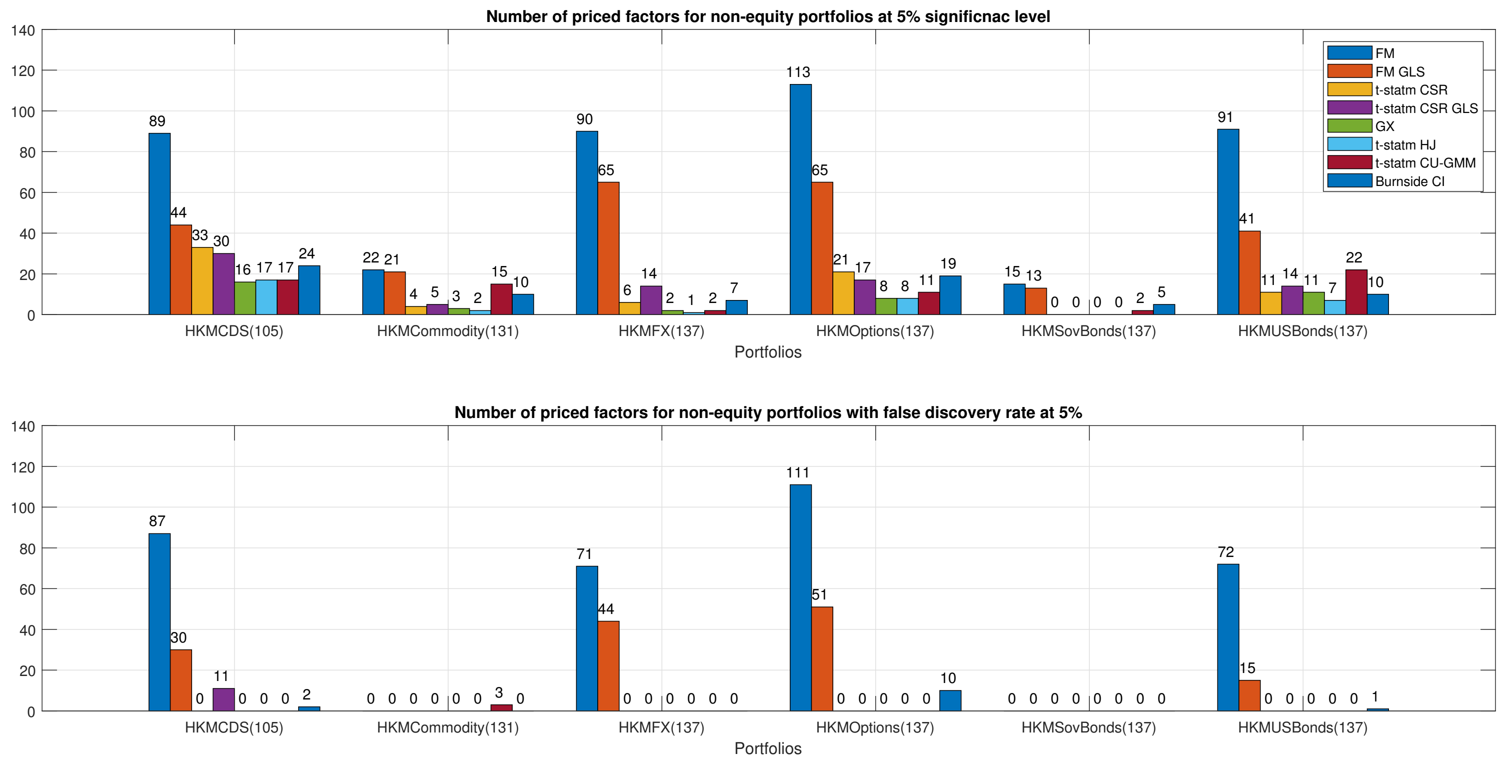


Figure 17: This figure compares the number of priced factors in equity portfolios suggested by each of the tests at 5% significance level across length of time series observations T <= 100, 100< T <= 300 and T > 300. Total number of factors that can be tested for each set of portfolios is indicated in a parenthesis associated with each set of portfolios.

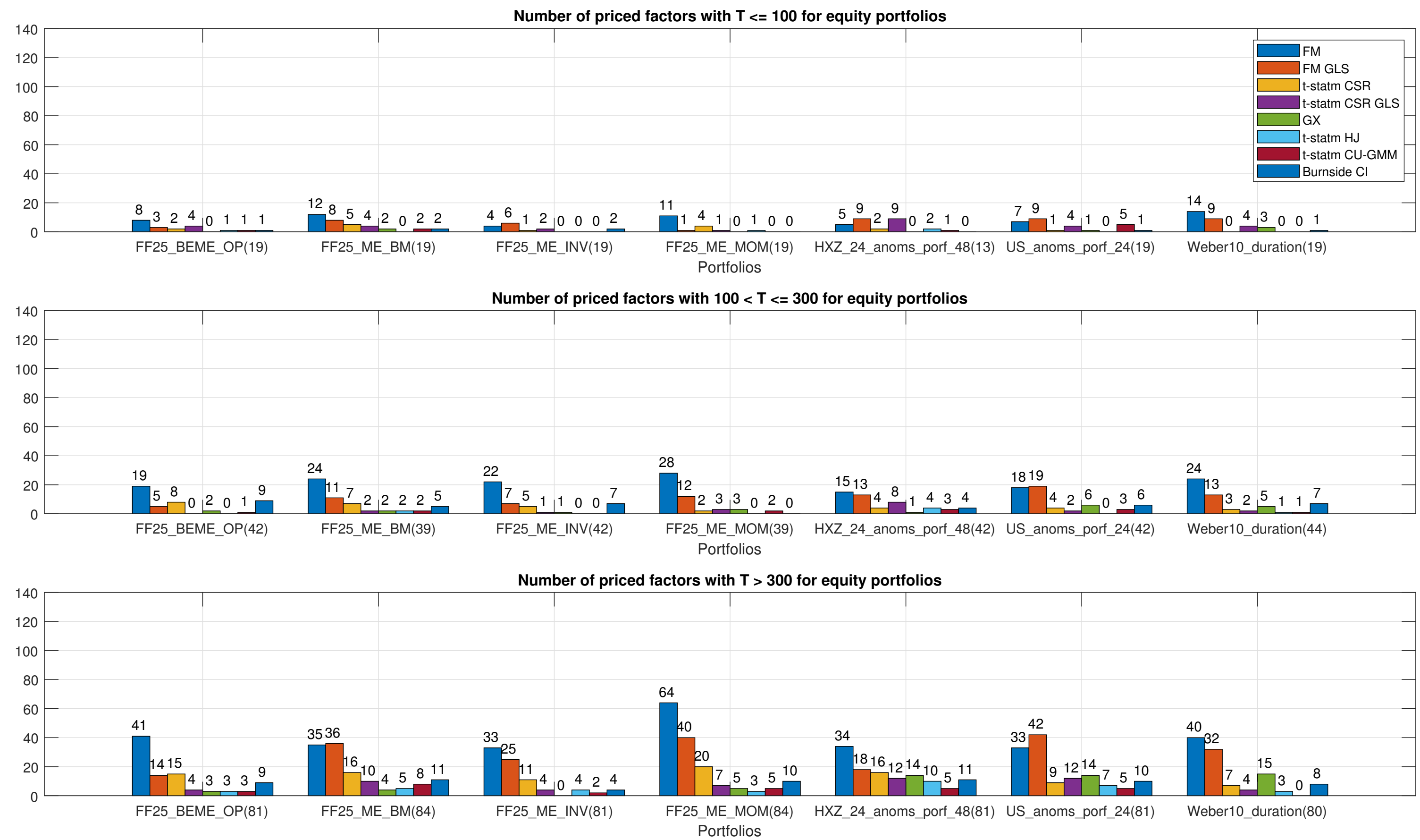


Figure 18: This figure compares the number of priced factors in non-equity portfolios suggested by each of the tests at 5% significance level across length of time series observations T <= 100, 100< T <= 300 and T > 300. Total number of factors that can be tested for each set of portfolios is indicated in a parenthesis associated with each set of portfolios. The length of CDS and sovereign bond portfolios does not exceed 300, so they do not have any observations with T>300.

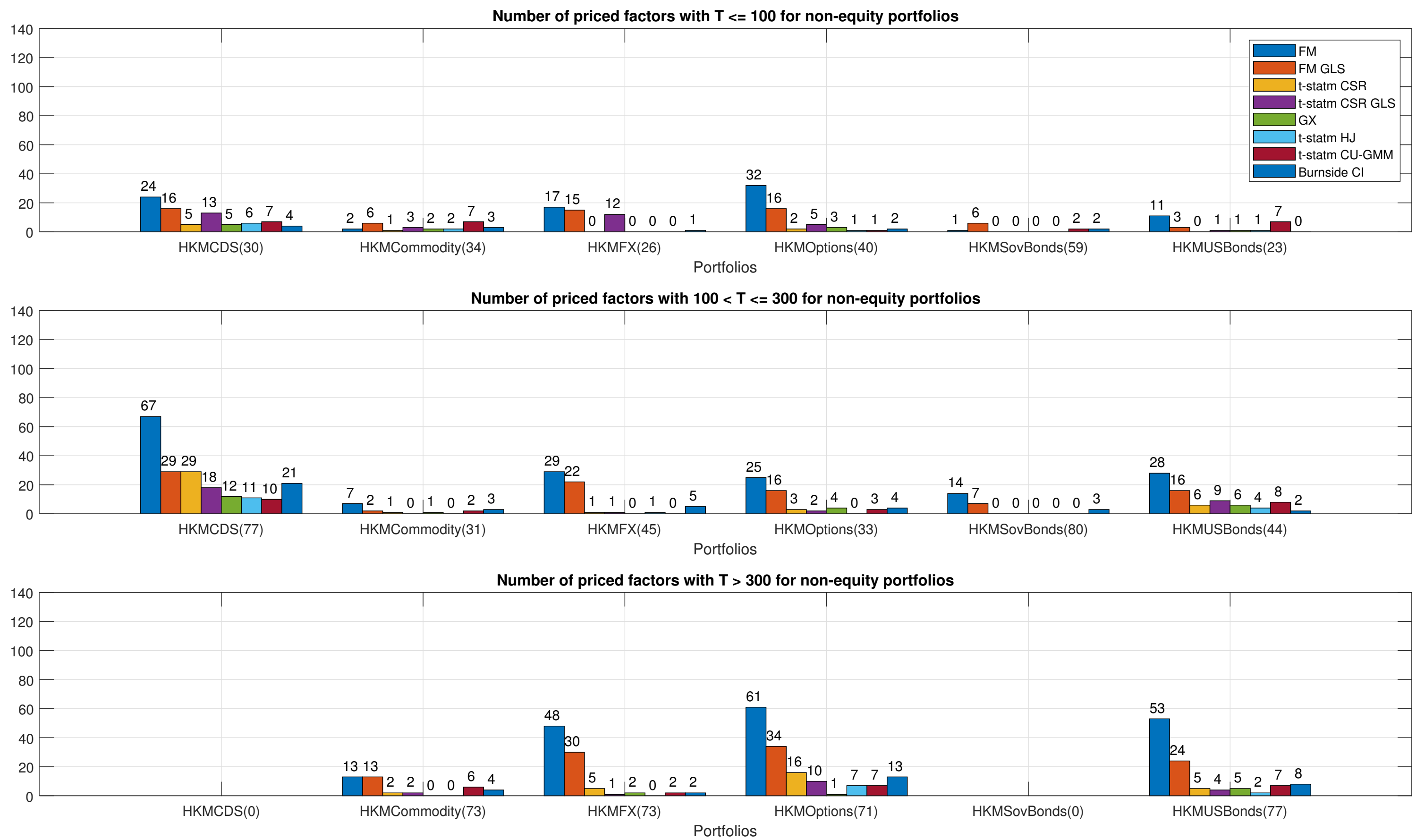


Figure 19: This figure shows the number of bounded 95% of confidence intervals of non-zero risk premium for equity portfolios. We use the techniques advocated by Kleibergen and Zhan (2020) to obtain confidence interval by inverting GRS-FAR test and mimicking portfolio approach (MPAR) robust to weak identification proposed in Kleibergen and Zhan (2018). We employ 5 types of base assets for the MPAR: GX, RP-PCs, SMV PCs, and PCs from equity portfolios and non-equity portfolios used in our paper.

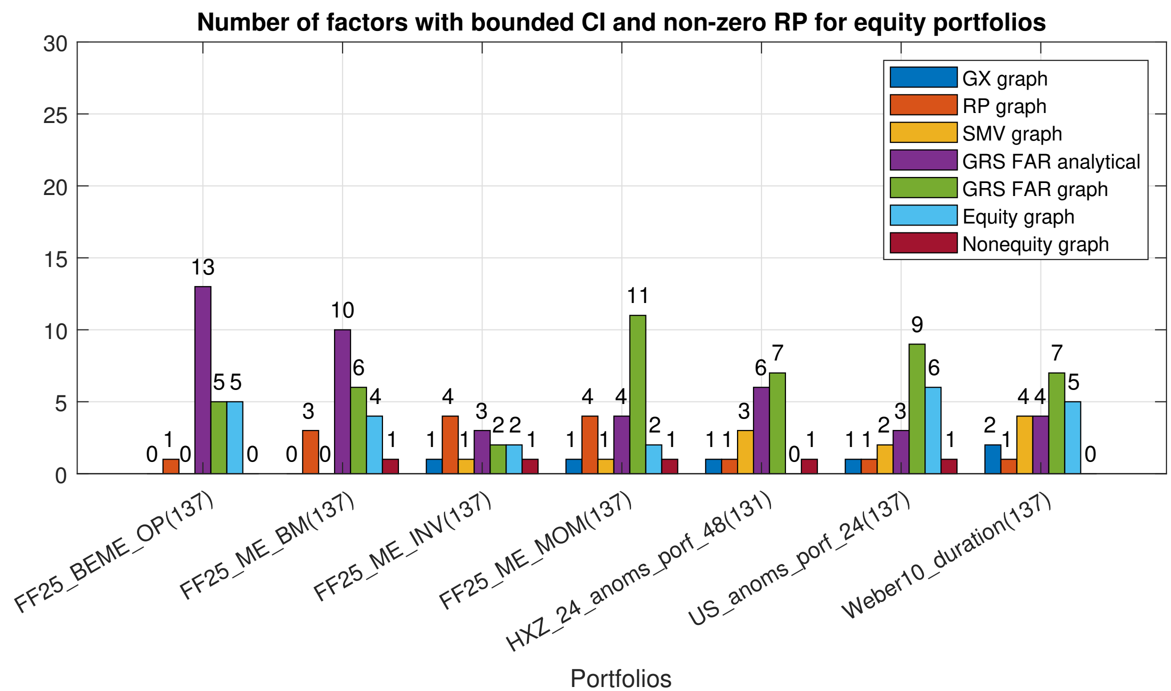


Figure 20: This figure compares the number of bounded 95% of confidence intervals of non-zero risk premium for equity portfolios across length of time series observations T <= 100, 100< T <= 300 and T > 300. We use the techniques advocated by Kleibergen and Zhan (2020) to obtain confidence interval by inverting GRS-FAR test and mimicking portfolio approach (MPAR) robust to weak identification proposed in Kleibergen and Zhan (2018). We employ 5 types of base assets for the MPAR: GX, RP-PCs, SMV PCs, and PCs from equity portfolios and non-equity portfolios used in our paper. Total number of factors that can be tested for each set of portfolios is indicated in a parenthesis associated with each set of portfolios.

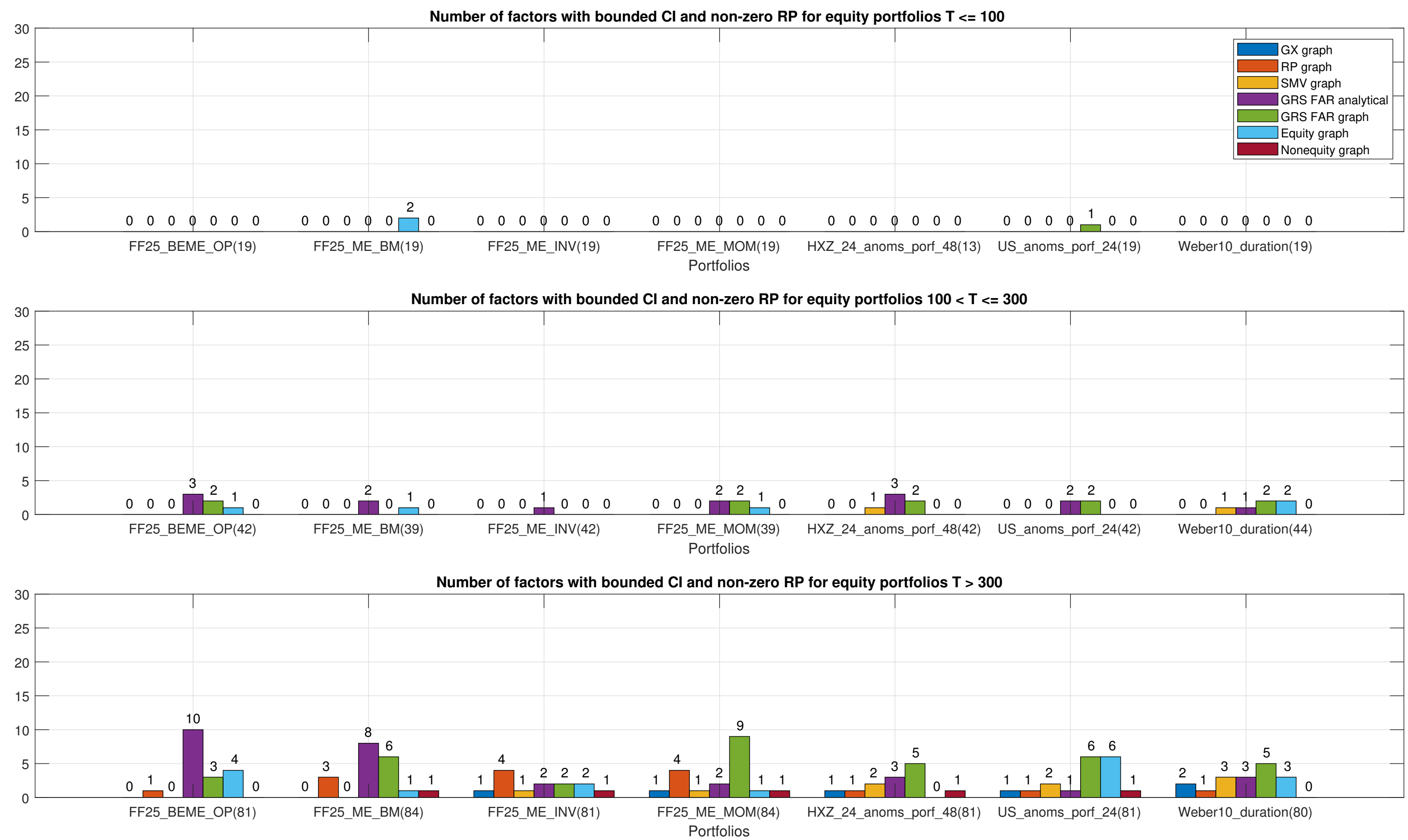


Figure 21: This figure shows the number of bounded 95% of confidence intervals of non-zero risk premium for non-equity portfolios. We use the techniques advocated by Kleibergen and Zhan (2020) to obtain confidence interval by inverting GRS-FAR test and mimicking portfolio approach (MPAR) robust to weak identification proposed in Kleibergen and Zhan (2018). We employ 5 types of base assets for the MPAR: GX, RP-PCs, SMV PCs, and PCs from equity portfolios and non-equity portfolios used in our paper.

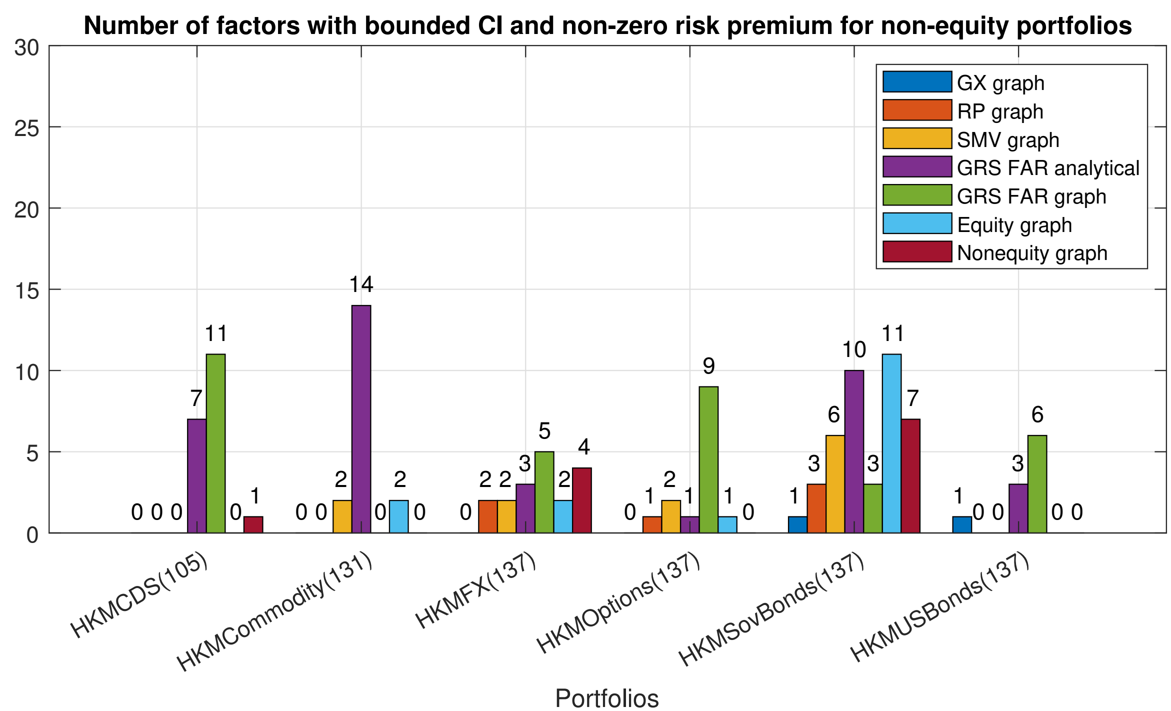


Figure 22: This figure compares the number of bounded 95% of confidence intervals of non-zero risk premium for non-equity portfolios across length of time series observations T <= 100, 100< T <= 300 and T > 300. We use the techniques advocated by Kleibergen and Zhan (2020) to obtain confidence interval by inverting GRS-FAR test and mimicking portfolio approach (MPAR) robust to weak identification proposed in Kleibergen and Zhan (2018). We employ 5 types of base assets for the MPAR: GX, RP-PCs, SMV PCs, and PCs from equity portfolios and non-equity portfolios used in our paper. Total number of factors that can be tested for each set of portfolios is indicated in a parenthesis associated with each set of portfolios.

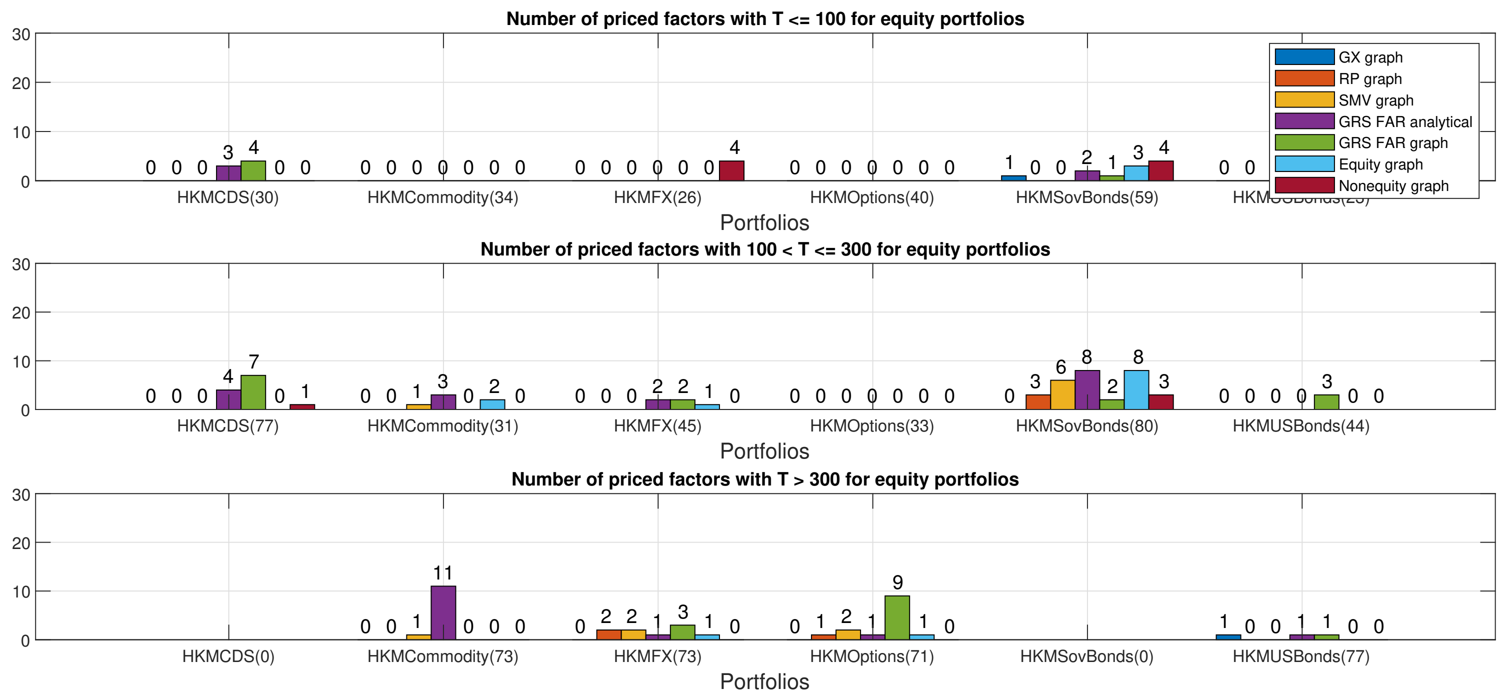


Figure 23: This figure compares the percentage of priced factors in model specifications with and without intercept in the cross-sectional analysis with test statistics including the Fama-MacBeth approach, generalized Shanken's EIV allowing for conditional heteroskedasticity (Jagannathan and Wang (1998)), misspecification-robust t-statistic (and their GLS estimators), three-pass approach by Giglio and Xiu (2021), misspecification-robust t-statistic based on Hansen-Jagannathan distance (Gospodinov et al. (2014)), misspecification-robust t-statistic in linear SDF estimated using CU-GMM. (Gospodinov et al. (2017)), and Burnside's bootstrap approach (Burnside (2011)).

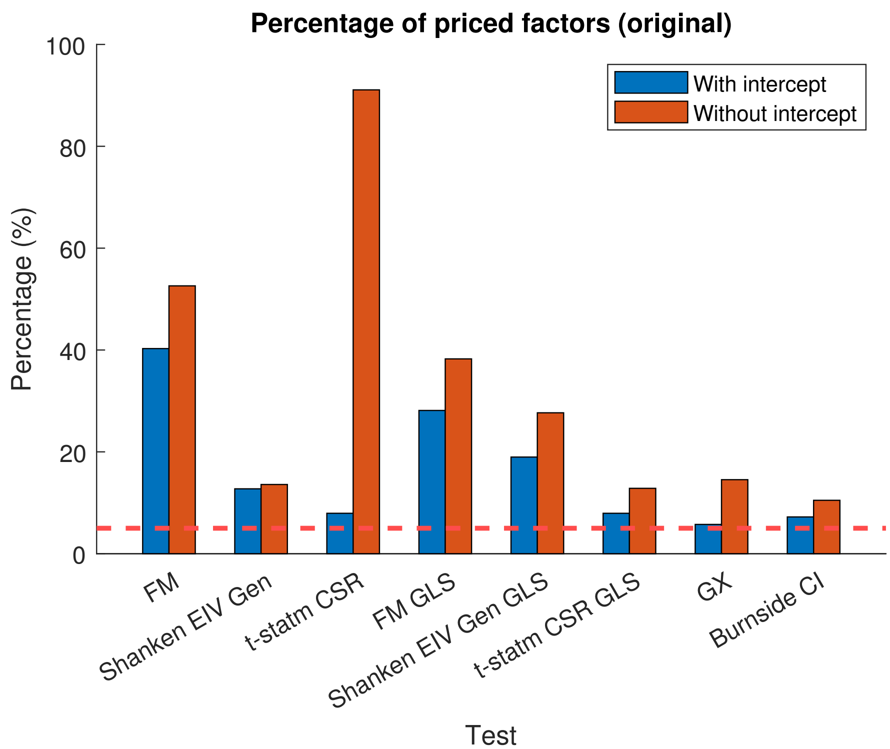


Figure 24: This figure compares the percentage of priced factors in model specifications with and without intercept in the cross-sectional analysis with test statistics including the Fama-MacBeth approach, generalized Shanken's EIV allowing for conditional heteroskedasticity (Jagannathan and Wang (1998)), misspecification-robust t-statistic (and their GLS estimators), three-pass approach by Giglio and Xiu (2021), misspecification-robust t-statistic based on Hansen-Jagannathan distance (Gospodinov et al. (2014)), misspecification-robust t-statistic in linear SDF estimated using CU-GMM. (Gospodinov et al. (2017)), and Burnside's bootstrap approach (Burnside (2011)). We control false discovery rate at 5% using Benjamini and Hochberg (1995) method.