Misspecification and Weak Identification in the Nontraded Factor Zoo*

Amedeo Andriollo[a], Cesare Robotti[b,†], Giulio Rossetti[c], and Xinyi Zhang[d]

[a] *Université Paris Dauphine-PSL, DRM, 75775 Paris, France*
[b,c,d] *Warwick Business School, Coventry, CV4 7AL, United Kingdom*

October 2025

ABSTRACT
To explain the cross-section of asset returns, a "*zoo*" of nontraded factors has been proposed. In contrast to traded factors, nontraded factors exhibit lower correlations with asset returns. Standard inference on risk premium therefore 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 by many empirical studies, while limited efforts have been devoted to "*domesticating*" such factors. After re-evaluating the nontraded factor zoo, we find that the vast majority of the original model specifications published in top academic journals suffer from the aforementioned fragilities. Robust inference indicates that most of the proposed nontraded factors are unpriced in the commonly used portfolios. The findings are more drastic when considering multiple hypothesis testing adjustments, or when incorporating the market factor as an additional control. Complementing these tests, a comprehensive beta-sorted portfolio analysis shows that few nontraded factors translate into economically meaningful investment premiums. However, when summarizing the nontraded factors via PCA, we find that the zoo does carry some non-zero pricing information.

Keywords: Factor Zoo; Nontradable; Misspecification; Weak Identification; Risk Premium; Portfolio Sorting

JEL classification: C58, G12.

*For the helpful comments, we thank the seminar participants at the SoFiE 2024. For their insightful suggestions, Andriollo is grateful to Bertille Antoine and Ao Wang, and Xinyi Zhang is grateful to Arie Gozluklu, Elise Gourier, Christian Julliard, and Ilaria Piatti.

†Corresponding author.

Email addresses: `amedeo.andriollo@dauphine.psl.eu` (A. Andriollo), `cesare.robotti@wbs.ac.uk` (C. Robotti), `giulio.rossetti.1@wbs.ac.uk` (G. Rossetti), and `xinyi.zhang.9@wbs.ac.uk` (X. Zhang).

1 Introduction

The search of factors explaining cross-sectional variations in expected asset returns has produced a vast "factor zoo", encompassing hundreds of candidates (Cochrane, 2011). Despite recent efforts to tame factor proliferation, a definite conclusion remains elusive due to the heterogeneity in methodologies, factor selection, and explored causes. The empirical evidence remains mixed, and questions persist about which factors capture genuine sources of systematic risk rather than statistical artifacts.

While the recent debate in the literature has focused primarily on traded factors, the universe of nontraded factors —those not constructed as portfolios or return spreads—has received comparatively little attention. These factors offer a theoretically appealing link between asset returns and the real economy, yet they pose substantial empirical challenges. Only few studies have underscored the econometric challenges that arise from the inherently weak correlations between nontraded factors and asset returns, an intrinsic feature of such variables that can lead to fragile identification, unreliable risk premium estimates, and ultimately misleading conclusions about their cross-sectional pricing ability[1]. Nevertheless, their proliferation has continued unabated over the past four decades, as reflected in the steady rise in newly proposed nontraded factors (Figure A.10). This persistent growth underscores that the nontraded factor zoo remains an active and influential component of empirical asset pricing research.

This paper provides a comprehensive and systematic re-evaluation of the "nontraded factor zoo" documented in top academic journals. Our approach expands existing work by applying robust methodologies designed to address issues of weak identification, model misspecification, and fragile inference, and complements recent literature that replicates empirical asset pricing studies to tame factor proliferation.

Empirically, we find that weak identification is pervasive across published nontraded factor models: roughly 40% of original specifications show clear signs of identification problems, and this share increases to more than 50% when the market factor is added to the model specification.

In the same spirit of Harvey et al. (2016), we collect 102 distinct nontraded factors published in selected academic journals from July 1986 to February 2023.[2] For practical classification purposes, we define nontraded factors as those not directly constructed as portfolios or as weighted combinations of asset returns.

[1] Gospodinov et al. (2014, 2019)

[2] For publicly sharing their data, we thank D. Ardia, O. Boguth, Y. Chen, Z. Da, F. Duarte, S. Giglio, A. Goyal, J. Grammig, B. Herskovic, D. Huang, M. Kargar, T. Kroencke, Y. Liu, M. Lettau, S. Ludvigson, A. Manela, T. Muir, J. Pan, C. Polk, R. Sadka, R. Stambaugh, P. Schuster, S. Van Nieuwerburgh, M. Vassalou, J. Wang, J. Wurgler, G. Zhou. We have consolidated the nontraded factor zoo into a single comprehensive source (detailed in Section 3.2).

Adopting a taxonomic approach similar to Jensen et al. (2022), we categorize these nontraded factors into six themes: *News*, *Sentiment*, *Consumption*, *Macroeconomics*, and *Intermediary and Aggregate Firm-level Risk*. A classification overview is presented in Section 3.2, while in the appendix we provide the full list of the nontraded factors, with their descriptions and references.

We assess the pricing performance of nontraded factors across a variety of widely used test portfolios by scientifically replicating published model specifications and separately conducting diagnostic tests using robust statistics. Specifically, we investigate: i) whether the original Right-Hand-Side (RHS) model specifications are prone to problems of weak identification and/or model misspecification, and ii) the fragility of inference on risk premia with respect to these issues. Our framework serves as a natural progression from the financial economics literature that has developed robust testing methodologies for reliable pricing inference in the context of nontraded factors.

Our replication study extends to three additional investigations. First, inspired by Gospodinov and Robotti (2021), we evaluate whether nontraded factors have different exposures relative to the market factor or merely serve as its noisy proxies, by examining potential identification failures from collinearity. Second, we examine the robustness of pricing results to the presence or absence of the cross-sectional intercept (zero-beta rate).[3] Finally, we determine how the control of multiple hypothesis testing affects our assessment of these issues from a frequentist perspective.

We find that pricing models that include nontraded factors generally share similar statistical properties, as they are frequently subject to weak identification and potential misspecification. Our results show weak identification is widespread, with a conservative average estimate indicating 40% of original specifications are likely to suffer from this problem. In general, nontraded factors also appear to capture exposures similar to the market factor, as the proportion of weakly identified models increases by over 10% when the market factor is included, relative to single-factor models. The issue becomes even more pronounced when applying Kleibergen and Zhan (2020)'s GRS-FAR test and the Kleibergen and Zhan (2018)'s identification-robust mimicking-portfolio test, which suggest over 75% of the original models are weakly identified.[4] Comparing identification-testing statistics suggests that Kleibergen and Paap (2006)'s test and finite-sample F-test display lower rejection rates than Gospodinov et al. (2017)'s CD-test and Chen and Fang (2019)'s test, possibly

[3]Our benchmark results use models including the cross-sectional intercept.

[4]For roughly 75% of original models, the confidence sets associated to those statistics are not bounded, which means that the models are more susceptible to weak identification rather than misspecification problems (or possibly both).

due to low power or heteroskedasticity issues. About model misspecification, results are less definitive: the Hansen–Jagannathan test indicates about 66% of original models may be misspecified, while Hansen's J-test suggests only 13%. This discrepancy occurs likely because of two reasons. First, we observe small differences between Anderson-Rubin rank statistic and J-statistic in over 80% of models, pointing to potential problems of weak identification. Second, the correlation between the J-test statistic and Chen and Fang (2019)'s statistic suggests that models that are dubbed as correctly specified by the J-test are more prone to suffer from weak identification.

These findings strongly advocate for robust statistical methods when assessing the pricing implications of nontraded factors. As pointed out by the previous studies, conventional approaches vulnerable to weak identification and misspecification tend to overstate factors' pricing performance, leading scholars and practitioners to erroneously conclude these factors possess cross-sectional explanatory power. Importantly, this mechanism of factor proliferation is distinct from the ones associated to data snooping or false discovery rate controls. Standard inference methodologies (Fama and MacBeth, 1973; Shanken, 1985) suggest up to 40% of proposed nontraded factors are priced, while robust procedures indicate that only 5–10% are statistically significant —a striking fourfold difference— at the 5% significance level. This discrepancy intensifies with larger time series samples ($T > 400$), where conventional inference indicates an increasing proportions of priced factors while robust inference maintain stable results. Our findings further highlight that, while excluding the cross-sectional intercept can theoretically improve risk premia identification by allowing constant non-zero exposures to identify the risk premia (Kleibergen and Zhan, 2020), this practice can contribute to factor proliferation by incorrectly attributing to nontraded factors with minimal exposure variation explanatory power that would be c aptured by the zero-beta rate.

For a constructive appraisal, we report the survived models whose pricing abilities are robust to the aforementioned issues, which include Campbell and Vuolteenaho (2004)'s cash flow news factor for US24 anomalies portfolios, Baker and Wurgler (2006)'s sentiment factors for the US anomalies and Fama-French 25 ME/BM portfolios, Parker and Julliard (2005)'s consumption factor for HXZ48 anomalies portfolios, Sadka (2006)'s liquidity factors for US bonds, Bali et al. (2017)'s uncertainty factors for the CDS portfolios, and Chen et al. (2018)'s liquidity factors for option and FF25 ME/MOM portfolios.[5]

After conducting the identification and specification tests, we turn to the economic relevance of nontraded

[5]We dubbed as "survived" 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 to Section 2).

factors by evaluating their pricing performance in out-of-sample terms using beta-sorted portfolios. When sorting individual stocks on estimated betas, only about 10% of the resulting long–short strategies deliver statistically significant returns, and even among these, the absolute average magnitude across all factors is modest. Consistent with previous findings, these results reinforce the conclusion that nontraded factors rarely translate into economically meaningful or robust investment spreads.

Finally, we extend our analysis to a global assessment of the nontraded factor zoo by constructing a five-factor model using Principal Component Analysis, inspired by Giglio and Xiu (2021). These principal components exhibit significant correlations with Fama and French (2015)'s and Lettau and Pelger (2020)'s but not with Giglio and Xiu (2021)'s factors. Our findings support their pricing ability across most of the widely used equity portfolios, lending further support to Bryzgalova et al. (2023a)'s conclusion that they capture some relevant dimensions of risk.

We contribute to two key strands of the asset pricing literature: robust analyses of nontraded risk factors and the reevaluation of empirical findings in the effort to tame the "factor zoo."

Nontraded factors have been integral to empirical asset pricing research. The methodological assessment of their statistical properties, however, only emerges with Kan and Zhang (1999) seminal work. Kleibergen (2009) subsequently demonstrates that inference in two-pass risk premia estimators is problematic under weak identification, proposing robust statistics and reexamining the pricing of macroeconomic factors from Jagannathan and Wang (1996) and Lettau and Ludvigson (2001). Kan et al. (2013) develop methods for evaluating pricing models through cross-sectional R^2, documenting substantially higher misspecification-robust standard errors for nontraded factors. Gospodinov et al. (2014, 2019) and Gospodinov and Robotti (2021) show that several consumption and macroeconomic factors lose their pricing significance once accounting for misspecification and weak identification. Complementing Kan et al. (2013), Kleibergen and Zhan (2015) derive asymptotic distributions for cross-sectional R^2 with potentially weak factors, revealing inflated pricing performance in various studies, with similar evidence emerging for consumption-based asset pricing models in Kleibergen and Zhan (2020). Kleibergen and Zhan (2025) further extend these concerns to nonlinear asset pricing frameworks. For mimicking portfolio approaches (e.g., Vassalou, 2003; Adrian et al., 2014), Kleibergen and Zhan (2018) introduce robust test statistics, and Bryzgalova et al. (2023b) propose risk premia estimation techniques that are robust to weak or irrelevant factors. Our paper extends rather than innovates this literature by providing a comprehensive and unifying analysis across a substantially broader spectrum of factors than previously examined.

In response to Cochrane (2011), asset pricing research has increasingly focused on disciplining the factor *"zoo"*. As an early attempt to *"navigate the zoo"*, Harvey et al. (2016) catalog 313 published papers and offer guidance on the appropriate t-statistic threshold for risk premia inference accounting for false discoveries. Several studies subsequently examine whether this factor proliferation results from data mining, investigating out-of-sample performance (McLean and Pontiff, 2016; Yan and Zheng, 2017; Linnainmaa and Roberts, 2018), and applying multiple hypothesis testing corrections (Chordia et al., 2020; Harvey and Liu, 2020; Giglio et al., 2021), generally suggesting that many anomalies are likely data snooping artifacts. Parallel skepticism has sparked a potential replication crisis. Hou et al. (2020) find that most of 452 replicated anomalies fail current empirical finance standards, while Jensen et al. (2022) and Chen and Zimmermann (2021) argue most pricing factors can be replicated successfully. Sharing Harvey et al. (2016)'s motivation, we pursue scientific replication rather than pure reproduction of results. We match nontraded factors and model specifications with original papers and evaluate them across diverse commonly used portfolios spanning equity and non-equity markets.[6] Unlike Feng et al. (2020)'s LASSO-based approach testing incremental explanatory power of the existing factors, our primary evaluation focuses on individual models.

Cross-sectional factor studies predominantly examine traded factors (e.g., Avramov et al., 2023, Dong et al., 2022 and Engelberg et al., 2023) with few conducting robust inference on both traded and nontraded factors. Recent exceptions include Bryzgalova et al. (2023a), who employed Bayesian techniques to evaluate 2.25 quadrillion models derived from 51 factors, and Zhang et al. (2021)'s comparison of tradable and nontradable pricing ability using Hansen-Jagannathan distance. Our work is distinctive in its specific focus on the nontraded factor zoo, its comprehensive scale, and its robust methodological approach that assesses identification failure and misspecification thoroughly before examining risk premia, all while scientifically replicating the original model specifications, rather than exploring all possible factor combinations.[7]

The rest of the paper is structured as follows. Section 2 summarizes the robust inference and test procedures used in this paper. Section 3 describes the data including factors, test portfolios and model specifications. At individual level, Section 4 discusses the main empirical findings and Section 5 studies the economic relevance of the factors. At the aggregate level, Section 6 studies the pricing ability of the

[6]Testing with respect to multiple asset classes is a necessary since nontraded factors are not ex-ante associated with specific asset classes.

[7]For instance, Kleibergen and Zhan (2022) consider all possible model specifications with factors drawn from a traded and nontraded zoo of 150 factors (15 billion models). Similar approach is used in Bryzgalova et al. (2023a) and Bryzgalova et al. (2023b). This approach departs significantly from the economically and financially motivated model specifications found in published literature.

nontraded factor zoo. Finally, Section 7 concludes.

2 Diagnostic Testing and Robust Inference Procedures

Consider the panel data $\{r_t, f_t\}_{t=1,..,T}$ with r_t being a $N \times 1$ vector of excess asset returns, and f_t being a $K \times 1$ vector of proposed risk factors. Given the vector of asset returns, we will refer to a set of K risk factors as a model specification. A linear factor pricing model is formulated as follows:

$$\mathbb{E}[r_t] = \gamma_0 \mathbf{1}_N + \alpha + \gamma\beta \tag{1}$$

where the main object of interest, γ, is the $K \times 1$ vector of risk premia, $\beta = \text{Cov}(r_t, f_t)(\text{Var}[f_t])^{-1}$ is the $N \times K$ matrix of factor exposures, γ_0 is the cross-sectional intercept (zero-beta rate), and $\mathbf{1}_N$ is a $N \times 1$ vector of ones, following standard notations. The linear asset pricing model is said to be correctly specified if it holds with $\alpha = 0_N$.

Alternatively to the beta representation, the model can be expressed using its Stochastic Discount Factor (SDF) representation, and we write equivalent to Eq.(1):

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

where we have $\lambda_0 = \gamma_0$ and $\lambda_1 = (\text{Var}[f_t])^{-1}\gamma_1$.

For the beta representation, the standard approach of estimating and inferring risk premia is the two-pass methodology (Fama and MacBeth, 1973; Shanken, 1992), while in the SDF framework, asset pricing restrictions are naturally framed as moment conditions, making GMM and maximum likelihood the preferred estimation strategies.

As previously discussed, another common approach for handling nontraded factors is through the mimicking portfolio approach, where the focus shifts to pricing the portfolios maximally correlated with the nontraded factors rather than the nontraded factors themselves. That is, rather than incorporating nontradable factors directly into the asset pricing model, one may use their time-series linear projection onto a set of base assets that span the asset space as risk factors. The analysis then usually proceeds without any substantive changes.

2.1 Overview of the Test Statistics

When a new risk factor is proposed, researchers typically assess its cross-sectional explanatory power by conducting inference on the associated risk premium. For standard inference methods to yield reliable results, two key assumptions must hold: i) the model must be identified, which essentially requires that the matrix of factor exposures has a full rank (i.e., $rk(\beta) = K$), and ii) the asset pricing model must be correctly specified, meaning that pricing errors are zero (i.e., $\alpha = 0_N$). Only after assessing these conditions, a researcher may draw reliable conclusions about the non-zero risk premia (i.e., $\gamma \neq 0$). Therefore, our replication study aims:

- To test whether each of the model specifications is weakly identified, $\mathcal{H}_0^{rk}$: rank$(\beta) < K$, we employ four test statistics proposed by Kleibergen and Paap (2006), Gospodinov et al. (2017), Chen and Fang (2019), and Kleibergen and Zhan (2020).
- To test whether the model is correctly specified, $\mathcal{H}_0^{spec}$: $\alpha = 0_N$, we use the conventional Sargan-Hansen J-test (Hansen, 1982) and the HJ-distance statistic (Hansen and Jagannathan, 1997).
- To conduct robust inference on non-zero risk premia, $\mathcal{H}_0^{rp}$: $\gamma = 0$, we use the robust statistics proposed by Kan et al. (2013), Gospodinov et al. (2014), Burnside (2011), Giglio et al. (2022), Kleibergen (2009), Kleibergen and Zhan (2020), Kleibergen and Zhan (2018).

  Regarding standard methodologies, we refer to "FM" (and "FM-GLS") as the inference based on Fama and MacBeth (1973)'s t-statistic, and "Shank" (and "Shank-GLS") as the inference based on the t-statistic associated to Shanken (1992)'s correction term.

Table 1 about here

Table 1 provides the list of the employed main test statistics and their short names. This section continues with a brief overview of the statistics used for model identification diagnostics and robust inference on risk premia, outlining their rationale, key assumptions, and asymptotics. We refer readers to the original sources for a more complete description, and for further details regarding the specification testing procedures.

2.1.1 Testing Weak Identification

With broad applicability, Kleibergen and Paap (2006) propose a generalized rank statistic constructed from the quadratic form of an orthogonal decomposition of the unrestricted matrix whose rank is tested, assuming a $\sqrt{T}-$asymptotically normal estimator of such matrix. Gospodinov et al. (2017) examine how spurious factors affect linear asset pricing models by analyzing the limiting behavior of the Continuously Updated (CU-)GMM estimator's specification test when the moment conditions' derivative matrix lacks full rank. Therefore, they develop a rank test similar to Cragg and Donald (1997), establishing its asymptotic distribution and conservative boundaries.[8] The finite-sample test statistics align with Kleibergen and Zhan (2020) and Kleibergen et al. (2023). The underlying set of assumptions is typical of the GMM literature.[9] The aforementioned statistics test the null hypothesis: $\mathcal{H}_0^{rk} : rk(\beta) = K - 1$.

Chen and Fang (2019) develop a novel technique for rank testing that shifts attention from the integer-valued rank to null hypotheses about eigenvalues. The null hypothesis is recasted as: $\mathcal{H}_0^{rk} : \phi_r(\beta) = 0$, thus examining $\phi_r(\cdot)$, which is the sum of $k - r$ the smallest squared singular values of β. It is relevant to remark that their methodology specifically addresses scenarios where Kleibergen and Paap (2006)'s generalized tests face challenges -when null limiting distributions fail to stochastically dominate those with ranks below the hypothesized level (e.g., heteroskedasticity). Their approach utilizes bootstrap methods to address the non-standard asymptotic distribution under the null hypothesis.

2.1.2 Misspecification-robust Standard Errors

Kan et al. (2013) and Gospodinov et al. (2014) relax the assumption of correct specification to allow for potential model misspecification (i.e., $\alpha \neq 0_N$). Under the conventioanl GMM framework, Kan et al. (2013) derive the asymptotically normal distribution of the cross-sectional sample R^2, and of the t-statistic associated with 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 an additional term related to misspecification.[10] Gospodinov et al. (2014) generalize the previous results to the SDF framework, allowing for the presence of at least one useless factor, and provide a t-statistic to test the risk premia that is robust to both identification failure and potential model

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

[9] Their framework assumes the asset returns and the factor time series to be jointly stationary and ergodic with positive definite variance and finite fourth moment.

[10] While it is more explicit the robustness to misspecification of the t-statistic, for the other set of statistic related to the R^2, they claim that: "*Although it is possible that some of the GMM sample moment conditions are not asymptotically 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).

misspecification. Additionally, they derive these results also with respect to the HJ-distance.

2.1.3 Bootstrap Risk Premia Inference

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. A factor then is considered priced if the bootstrapped confidence interval excludes zero. The approach is designed to account for the generated-regressor issue and displays robustness to the problems of spurious factors and potential model misspecification.

2.1.4 Inference via the Three-step Procedure

Giglio and Xiu (2021) take a novel approach by assuming excess returns are priced by unobservable factors. Merging latent factor and mimicking portfolio literature, they develop a three-step method: first estimating latent factor risk premia, then projecting observed factors onto PCA-extracted latent factors, and finally obtaining the risk premia estimates for the observed risk factors. Their approach claims robustness against omitted and/or spurious factors. Their results are derived under the Bai (2003)-style assumptions: i) the latent factors correctly price the assets, ii) the pricing factors are strong and pervasive, and iii) the cross-sectional dimension is proportional to the time series (i.e., $N, T \to \infty,\ N/T \to c > 0$).

2.1.5 Inference via the GRS-FAR Confidence Intervals

Under analogous settings, Kleibergen (2009) and Kleibergen and Zhan (2020) provide a variety of statistics to test the pricing of factors, which are robust to repackaging and to all possible values of the betas.[11] Among the many, Kleibergen (2009) proposes a statistic *à la* Anderson and Rubin (1949), the Factor AR (FAR) statistic. To avoid pre-test bias, Kleibergen and Zhan (2020) introduces an extension of the FAR test, i.e., the GRS-FAR statistic, which is a joint test on the risk premia and on the correct specification of the model. Particularly, confidence sets need to be derived from inverting the test: the methodology constructs the $100 \times (1 - \alpha')\%$ confidence intervals where α' represents the significance level by performing a grid search for the hypothesized value of the risk premium, taking values at which the null cannot be rejected. Such confidence sets might be: i) unbounded, which suggests weak identification; i) empty, implying that

[11] "*[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. Regarding the robustness in terms of the betas, the main intuition is that the pricing errors (i.e., α) and the betas (i.e., β) are independently distributed when the sample size becomes large.

the model is prone to be misspecified; iii) bounded, so the risk premia are identified and model is correctly specified. Only in the latter case, if the (bounded) confidence intervals do not include zero, we can conclude the factors are priced.

2.1.6 Identification-Robust Mimicking Portfolio Approach

Kleibergen and Zhan (2018) develops an identification-robust methodology for conducting inference on the risk premia in the mimicking portfolio framework. Similarly to the GRS-FAR statistic, by inverting the test, the methodology constructs the $100 \times (1-\alpha')\%$ confidence intervals where α' represents the significance level. When adopting this methodology, we avoid eventual biases due to the particular choice of the spanning portfolios by adopting a data-driven approach, similar to Giglio and Xiu (2021)'s intuition: we use PCA-based factors, treating them as observables, rather than manually selected test assets (refer to Section 3.4).

3 Data and Model Specifications

3.1 Nontraded Risk Factors

A traded (or tradable) risk factor represents a source of systematic risk which can be traded in the financial markets. Well-known examples include the renowned Fama and French (1993)'s three factors, Carhart (1997)'s momentum factor, and more broadly, factors linked to firm characteristics or common sources of risk, typically expressed as portfolio excess returns or return spreads (Harvey et al., 2016, and Barillas and Shanken, 2018).

In contrast, a nontraded (or nontradable) risk factor is a factor that cannot be directly traded — commonly referred to as macroeconomic or economic factors (Balduzzi and Robotti, 2010; Kleibergen and Zhan, 2018).[12] It is worth emphasizing that economic factors constitute just one category of nontradable factors, not the entire spectrum. Some researchers have developed both traded and nontraded versions of the same factor. Pastor and Stambaugh (2019) discuss the difference between the two versions and highlight that the traded one is useful in obtaining an interpretable alpha from a multifactor model including liquidity risk, but the nontraded one should be used when estimating the liquidity risk of an asset.

The distinction between traded and nontraded factors becomes more apparent in a defined asset pricing model. Under a correctly specified model, the risk premium of a traded factor is identical to the expected

[12] We maintain an agnostic stance on why such risk is not traded, whether due to market incompleteness, market frictions, or theoretically motivated models.

excess returns on its exact mimicking portfolio. However, for a nontraded factor, the risk premium generally diverges from its expected value, as it does not represent returns from a tradable position. Beyond cross-sectional regression methodologies (see Cochrane, 2009, chap. 12), researchers have developed methodologies that replace the nontradable factors with their maximally correlated mimicking portfolios, derived through linear projections of nontraded factors onto a set of base assets spanning the asset space (e.g., Vassalou (2003), Adrian et al. (2014)).[13] The theoretical properties and equivalence between these approaches are studied in Balduzzi and Robotti (2008) and Balduzzi and Robotti (2010).

3.2 The Nontraded Factor Zoo

We collect 102 distinct nontraded factors from 51 papers published in top economics, finance, and management academic journals in the last 40 years (from July 1986 to February 2023): *Journal of Finance* (19), *Journal of Financial Economics* (18), *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 six themes based on the risk dimensions measured by them or the data type used to represent the risk factor: Consumption (18), Macroeconomics (32), News (11), Sentiment (17), Liquidity (16), Intermediary and Aggregate Firm-level (8).

Tables A.1-A.6 in Appendix 7 tabulate the factors with their short names, brief descriptions, and references. Similarly to Harvey et al. (2016), our focus is exclusively on published papers that propose new factors.[14] Our protocol for classifying a proposed risk factor as nontraded relies on two criteria: either the extant literature has classified it as nontraded, or it has not been empirically constructed as a function of asset returns. In instances of conflicting classifications between the two, we prioritize the former criterion. To the best of our knowledge, our "zoo" of nontraded factors constitutes the most comprehensive systematic classification of nontradables in the field.

While we exclude studies that examine identical nontraded factors across varying contexts (test assets and/or time periods), we include research proposing distinct empirical proxies for an equivalent underlying source of risk, such as alternative measures of consumption growth (Table A.1). Furthermore, we do not

[13] 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). Additionally, see Kleibergen and Zhan (2018).

[14] We avoid duplicating the recordings of the factors. We report the the first reference proposing their use in explaining cross-sectional (or the time-series) variation of asset returns.

restrict our analysis to factors exclusively designed for pricing the cross-section of expected returns: among 51 papers analyzed, we include 11 studies that demonstrate significant time series explanatory power of the nontraded factors. These nontraded factors are thus included in our classification given that such variables, established as "useful" in the time series dimension, could eventually be used for cross-sectional predictability purposes.

Regarding the nontraded factors that are not publicly shared, we replicate them using macroeconomic series from St. Louis Fed's Federal Reserve Economic Data (FRED).[15] We also use the University of Michigan Inflation Expectation, and Real Gross National Product ($RGNP$)/Gross Domestic Product ($RGDP$) from the Survey of Professional Forecasters of Federal Reserve Bank of Philadelphia. From Amit Goyal's website, we use long-term government bond yield, dividends on S&P 500 index, as well as data to compute term spread and default spread.

Overall, we collect 70 factors at monthly, 22 at quarterly, and 11 at annual frequency. Lastly, the market factor (MKT) and risk-free rate are from Kenneth French's data library.

3.3 Test Portfolios

Ex ante, nontraded factors are generally not associated with a specific asset class. Acknowledging that asset pricing model performance can vary with respect to test assets under model misspecification (Kan et al. (2013)), we employ multiple conventionally used test portfolios. Therefore, our scientific replication systematically examines the cross-sectional pricing ability of nontraded factors in thirteen sets of portfolios, covering seven major asset classes. To match the factors at quarterly and annual frequencies, we compound monthly portfolio returns.

We use seven sets of equity portfolios: i) 25 Fama-French size and book-to-market portfolios (FF25 ME/BM) from July 1926 to September 2022; ii) 25 Fama-French book-to-market and operating profitability portfolios (FF25 $BEME/OP$) from July 1963 to September 2022; iii) 25 Fama-French size and investment portfolios (FF25 ME/INV) from July 1963 to September 2022; iv) 25 Fama-French size and momentum portfolios (FF25 ME/MOM) from January 1927 to September 2022; v) 24 portfolios formed with the bottom and top portfolios of 11 US anomalies listed in Kenneth French's data library as well as two long-short

[15] The macroeconomic series are: Consumer Price Index (CPI) for All Urban Consumers, 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.

portfolios of "quality-minus-junk" (Asness et al., 2019) and "betting-against-beta" (Frazzini and Pedersen, 2014) from AQR's data library (US24 anomalies) spanning from July 1963 to August 2022; vi) 48 portfolios formed with the bottom and top portfolios of 24 anomalies studied by Hou et al. (2020) (HXZ48 anomalies) spanning from January 1967 to December 2021; vii) 10 equity portfolios sorted by cash flow duration used by Weber (2018) (Weber10 duration) from July 1963 to June 2014.[16]

Regarding non-equity portfolios, following Lettau et al. (2014), He et al. (2017), and Lettau et al. (2019), we consider: i) 20 US bond portfolios spanning January 1975 to December 2011 (USbonds); ii) 6 sovereign bond portfolios from January 1995 to April 2011 (SovBonds); iii) 18 S&P 500 index option portfolios from April 1986 to January 2012 (Options); iv) 12 foreign exchange portfolios from March 1976 to January 2010 (FX); v) 23 commodity portfolios from September 1986 to December 2012 (Commodity); vi) 20 CDS portfolios from February 2001 to December 2012 (CDS). The data is available on Asaf Manela's website.

3.4 Replicating the Model Specifications

For a systematic assessment on the nontraded factor zoo, our primary objective is to scientifically replicate the proposed model specifications in the original papers. Hence, we refer to "Original Models" as the batch of model specifications that share the Right-Hand-Side (RHS) variables as closest as possible to the respective published papers. For the papers that run cross-sectional analysis at the individual-asset level, we include the corresponding factors that proxy for the anomalies or characteristics used in the paper. Starting from the 102 nontraded factors, we find 154 unique model specifications that were originally proposed in the published papers.

4 Replication study

This section presents the assessment of the nontraded factor zoo by analyzing the statistical properties of original model specifications, focusing on testing identification strength, model misspecification, and

[16] The 11 US 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. Hou et al. (2020)'s 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*). For more details, please refer to Appendix A.1 of Bandi et al. (2021).

on conducting inference robust to these issues to obtain reliable conclusions on the non-zero risk premia estimates.

Standard inference procedures rely on the assumptions that the underlying asset pricing model is fully identified and correctly specified, and several existing statistical tools are available to evaluate the validity of these assumptions. Our core analysis indicates that a substantial proportion of empirical asset pricing research concerning nontradables frequently overlooks these considerations, as our replication study reveals that diagnostic assessments typically contradict these assumptions are holding. We stress that this methodological oversight results in incorrect conclusions regarding the pricing of the proposed factors, and therefore leading to factor proliferation.

The analysis subsequently extends to single-factor models, specifically selecting nontraded factors and implementing two setups: standalone single-nontradable models and two-factor models with single-nontradables augmented with the market factor. This approach serves two purposes: first, to determine whether the identification weakness is a distinctive feature related to the proposed factor, rather than to the original model specification; second, following Gospodinov and Robotti (2021), to assess whether nontraded factors capture the same linear information as the market factor.

We present the results by portfolio classes, factor themes, and sample sizes, allowing us to best capture differences across Data Generating Processes and accommodate the heterogeneity in terms of the asymptotic properties of the statistics used. If not differently specified, the significance level is set at 5%.

4.1 Testing Weak Identification

We individually test the rank deficiency of the factor exposures for 154 original model specifications with respect to 13 sets of test portfolios. In aggregating the results for the four employed statistics, we refer to the "conservative" scenario, where the null hypothesis of rank deficiency is rejected by all tests (unanimity) and to the "favorable" scenario, where the null hypothesis of rank deficiency is rejected at least by one test. In other words, the conservative scenario provides an upper bound about identification failure, while the favorable scenario provides a lower bound.[17]

Figure 1 about here

[17] 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.

Figure 1 displays the percentages of original model specifications for which we fail to reject the null hypothesis of weak identification. Based on Kleibergen and Paap (2006)'s and Kleibergen and Zhan (2020)'s rank tests (KP and F-stat), we find that approximately 90% and 74%, respectively, of the model specifications may be considered as weakly identified. Gospodinov et al. (2017)'s and Chen and Fang (2019)'s rank tests (CD and CF) also suggest that at least 50% of the model specifications suffer from identification failure. We reject the null hypotheses of all four rank tests only for 6% of them.[18]

Overall, the widespread problem of identification failure in the nontraded factor zoo calls for cautious diagnostic checks before drawing conclusions about risk premia.

We also summarize the results with respect to different sample sizes, T: $T \leq 100$ (small), $100 < T \leq 400$ (medium) and $T > 400$ (large). Accordingly, we calculate the proportions of identification failure in each size group to have a comparison in terms of power of the statistics. The results of the identification tests with different sample sizes are also presented in Figure 1. For short samples ($T \leq 100$), a notable difference is observed between CD and the other three tests. In particular, while CD indicates that only about 30% of the specifications may suffer from identification problems, KP and the F-stat (CF test) indicate that more than 80% (70%) of the models should be considered as weakly identified, therefore pointing towards the potential lack of power of the CD test. In this class of models, we find that most of them include macroeconomic factors and consumption factors that are observed with annual frequency. For moderate sample sizes ($100 < T \leq 400$), 91% of the specifications can be considered weakly identified by the KP test, compared to approximately 75% by the F-stat, and 55% by both CD and CF. Most of these models include quarterly factors, and some also incorporate monthly liquidity factors. In the large sample ($T > 400$), the difference in results between the F-stat and CD is reduced. The former indicates that 64% of the considered specifications suffer from rank deficiency, and the latter reports 59%. The null hypothesis is rejected by KP in a higher proportion (86%) in this large-sample group of specifications. The null hypothesis is also rejected more frequently for the CF test, showing that about 51% of the model specifications cannot be rejected to be weakly identified.

Figures 2 and 3 display the proportion of models that are prone to identification failure across portfolio classes. The issue is generally pervasive across portfolios, being remarked in the non-equity portfolio classes. Among equity portfolios, specifications with Weber 10 portfolios sorted by cash flow duration appear to be the least identified. Empirical asset pricing studies often use FF25 ME/BM portfolios as test portfolios,

[18] It comes from subtracting the percentage of the conservative scenario (94%) from the total.

making them the portfolios that are most frequently claimed to be priced by their factors in the empirical literature. However, we still observe identification failure in at least 45% of the models. Among non-equity portfolios, sovereign bonds and FX are the most likely to be weakly identified portfolios of the models.

Figures 2-3-4 about here

Figure 4 displays the pervasiveness of the problem with respect to the different captured risk dimensions or, in other words, by the themes of the nontraded factors (see Section 3.2). At the aggregate level, models associated with proxies of liquidity, sentiment, and volatility are less likely to be dubbed weakly identified. By looking at CD test and CF test, we report that models associated with consumption and liquidity factors show low percentages of identification failure. The low percentage for consumption factor is partly due to the shorter samples, which implies potential finite-sample distortions when conducting inference on the betas' rank.

Given the large number of tested hypotheses, we control for Type I error using multiple hypothesis correction, paralleling the intuition of Harvey et al. (2016). We control the false discovery rate using the Benjamini and Hochberg (1995)'s approach, setting it at 5% for all model specifications in each test. As expected, the issue becomes more pervasive after correction (Figure A.11 in Appendix **??**). The percentages of models deemed weakly identified increase, ranging from 58% (CD) to almost 99% (KP). Most interestingly, in the favorable scenario, we still reject around 40% of the original model specifications.

Lastly, we report that the issues regarding identification are alleviated when imposing the zero-beta return equals zero, but the improvement is limited. Unsurprisingly, the percentage of rank deficiency decreases by 5% when the intercept is removed in the F-test. As noted in Kleibergen and Zhan (2020), the improvement is due to that the restriction on the cross-sectional intercept allows almost-constant betas to still identify the risk premia.

4.1.1 Single-nontradable factor models

Weak identification of an asset pricing model occurs either when the factors are uncorrelated with the returns on test assets, or when two or more factors that are noisy proxies of the same underlying factor (Gospodinov et al., 2014). Inspired by Gospodinov and Robotti (2021), we explore the incremental linear explanatory power provided by the nontraded zoo compared to the market factor. If a nontraded factor is in fact a noisy proxy of the market factor, the two-factor asset pricing model (i.e., the nontradadable plus

MKT) is very likely to suffer from identification failure due to collinearity. Figure 5 shows the change in the percentage of identification failure when the market factor is added to the single-factor specification. Generally, the percentage of identification failure increases: except for the KP test, the proportion of weakly identified models increases by more than 10% after including the market factor. There is also an approximately 5% increase in the KP test compared to the single-factor benchmark analysis. This suggests that the exposures associated with some nontraded factors are not statistically different from that of the market factor.

Figure 5 about here

4.2 Testing Model Misspecification

Classical inferences on risk premium estimates are computed under the implicit assumption of correct model specification. Deviation from this assumption, which may lead to overrejection of asset pricing models, eventually amplify distortions already induced by identification failure, therefore leading to unreliable inference on the risk premia, especially for nontraded factors.

Figure 6 presents the percentages of models that reject the null hypothesis of correct specification. We observe a significant divergence between the results of Hansen (1982)'s and Hansen and Jagannathan (1997)'s tests (J-test and HJ). While the J-test indicates that only 13% of the models are misspecified,testing via the HJ-distance reveals a substantially higher percentage of misspecified models (approximately 66%). In particular, we fail to reject the null of correct specification using the HJ test for more than 90% of the original model specifications, associated to HXZ48 anomalies, and US Bond portfolios (Figure A.15, Appendix 7). Overall, for non-equity portfolios, testing via $J-$test would lead to failing to reject the null of correct specification more frequently (Figure A.16, Appendix 7).

The divergence between the results of the two testing procedures can be related to two points. First, the J-test suffers from severe distortions in the presence of weak identification. In particular, we find that in more than 80% of the original models, the J-test statistic and the Anderson-Rubin test statistic are close in magnitude, thus questioning the reliability of the J-test in the scenario (refer to Kleibergen and Zhan, 2022). Second, 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.

Figure 6 about here

4.3 Robust Inference on the Risk premia

In the presence of weak identification and/or misspecification, the standard methodologies for risk premium inference are unreliable. The previous conclusions about the ubiquity of these issues strongly motivate us to conduct robust inference on the risk premia, and to compare the results between risk premia estimates. While the core analysis places no restrictions on the zero-beta rate, we also analyze the sensitivity of our results with respect to the zero restriction of the cross-sectional intercept.

The evidence indicates that no single factor is robustly priced across the range of tests and test assets considered. Figure 7 shows the percentages of deemingly priced factors (at a 5% significance level) with respect to the battery of statistics we employ. The percentages are calculated based on t-ratios, with standard errors computed using the Fama and MacBeth (1973) approach, the Jagannathan and Wang (1998) generalized Shanken errors-in-variables method accounting for conditional heteroskedasticity, and robust methods as detailed in Section 2.1.

Figure 7 about here

Using the standard approach (FM), one can conclude that a large proportion of factors are priced in the test portfolios, with approximately 40% of factors exhibiting significant risk premia. The proportion decreases by around 10% as we estimate it with the generalized least squares (FM-GLS). After adjusting for the fact that the betas are estimated, the proportion is also reduced by more than half. Taking into account model misspecification with Kan et al. (2013)'s statistics (CSR and CSR-GLS), around $5-8\%$ of factors exhibit risk premia statistically different from zero. Using Gospodinov et al. (2014)'s test with standard errors robust to identification failure and potential misspecification (HJ) reduces the proportion of factors with significant risk premia even further, to just 5%, which is the lower threshold due to the 5% significance level. Gospodinov et al. (2017)'s approach (CU-GMM) and Giglio and Xiu (2021)'s three-pass procedure (GX) both identify roughly 8% of factors with non-zero risk premia. Burnside (2011)'s bootstrap approach (BS) suggests a higher fraction, at 13%. In general, standard inference overrejects the null of zero pricing ability of the nontraded factors, thus favoring factor proliferation.

The pricing results sorted by test portfolios are presented in Figures A.2 for equity portfolios and A.4 for non-equity portfolios (Appendix 7). Overrejections tend to appear more frequent for non-equity portfolios, especially for the options, CDS, FX and US bonds portfolios. Standard inference suggests that more than 65% (47%) of the nontraded factors are priced in CDS and options (FX and US bonds) portfolios. Looking at the robust t-tests, on the contrary, a large number of nontraded factors turns out to be unpriced. Robust statistics suggest only 4% – 22% of priced factors for CDS, 1% – 8% for commodity, 1% – 8% for FX, 5% – 15% for options, and 3% – 18% for US bonds portfolios. It is worth mentioning that, for sovereign bonds, CSR, CSR-GLS, HJ, and GX all indicate that few of the nontraded factors are priced. For the popular Fama-French equity portfolios, less than 20% of factors are priced and only 10% of priced factors if one uses the methods of HJ and CU-GMM. As expected, if one accounts for the multiple hypothesis correction as well, the pricing relevance of the nontraded factor zoo completely disappears as presented in Figure A.1 (Appendix 7).

For a constructive appraisal, we list the priced nontraded factors with significant risk premia, as suggested the consensus of three misspecification-robust t-tests (CSR GLS, HJ and CU-GMM), in Table 2, without correcting for multiple hypothesis testing. We report the model specifications that also survive the testing for weak identification using the CF test. The nontraded factors that are considered as priced in some cross-sections of test assets by the three tests include: Campbell and Vuolteenaho (2004)'s cash flow news factor for US24 anomalies portfolios, Parker and Julliard (2005)'s consumption factor for HXZ48 anomalies portfolios, Sadka (2006)'s liquidity factors for US bonds, Bali et al. (2017)'s uncertainty factors for the CDS portfolios, Chen et al. (2018)'s liquidity factors for option and FF25 ME/MOM portfolios.

Table 2 about here

Additionally, we carry out a similar analysis using the Mimicking Portfolio Anderson-Rubin statistic in Kleibergen and Zhan (2018) (MPAR) and the GRS Factor Anderson-Rubin (GRS-FAR) statistic in Kleibergen and Zhan (2020). Both tests generate 95% confidence intervals for the risk premia with four possible scenarios: unbounded, empty, bounded including zero, and bounded excluding zero. Only a bounded confidence interval away from zero indicates that the underlying factor is priced. Figure A.6 (Appendix 7) summarizes the results of the testing procedures in terms of frequencies of type of confidence interval. For the mimicking portfolio approach we use the following sets of base assets: 1) the principal components extracted from the equity portfolios, 2) the ones extracted from non-equity portfolios, as well as 3) the ones extracted from both equity and non-equity portfolios in the style of Giglio and Xiu (2021). According to the GRS-FAR test, only approximately 4% of the nontraded factors are priced in the test portfolios with bounded confidence intervals excluding zero, and 4% of the nontraded factors are associated with bounded confidence intervals including zero. Regarding the remaining models, more than 90% of the factors are associated with unbounded intervals (weak identification). With the MPAR approach, no more than 1% of the nontraded factors are considered priced, regardless of the selection of the spanning test assets and the classes of assets. The vast majority of original models lead to an unbounded confidence interval for the nontraded factor risk premia.

Lastly, we explore the impact of restricting the zero-beta rate or, in other words, the sensitivity of the risk premia estimates to the removal of the intercept in the cross-sectional regression. As Kleibergen and Zhan (2020) and Kroencke and Thimme (2021) point out, excluding the intercept may improve the identification, when the model is correctly specified. Figure A.7 (Appendix 7) compares the percentages of non-zero risk premia for all test portfolios in the two scenarios, with and without cross-sectional intercept. Removing the intercept leads to a larger proportion of deemingly priced factors. However, it must be understood rather as a cautionary tale: excluding the intercept may wrongly attribute the explanatory power to the factor with nearly constant exposures, which in turn would be captured by the cross-sectional intercept. Therefore, unless there are strong reasons for supporting correct specification of the proposed model, including the cross-sectional intercept may be essential to avoid (nontraded) factor proliferation.

5 Economic Relevance of Nontraded Factors

The replication study outlined above provides statistical *in-sample* validation of nontraded factors as pricing factors. The findings from the various testing procedures examine whether the proposed models satisfy the cross-sectional restrictions implied by asset-pricing theory, conditional on the beta estimates and realized returns. However, these findings do not directly address whether nontraded factors generate economically meaningful returns in actual trading scenarios. Previous inference on risk premia evaluate the model validity using the full sample information, whereas portfolio analysis allows us to assess the economic relevance of these factors from an *out-of-sample* perspective. To comprehensively evaluate the nontraded factor zoo, we therefore propose a complementary out-of-sample analysis using beta-sorted portfolios.

Following the standard practice in the empirical asset pricing literature (e.g., Fama and French (1993), Adrian et al. (2014) and Bali et al. (2016)), we construct beta-sorted portfolios to study the link between estimated factor exposures and expected returns. For each asset, we estimate rolling betas with respect to a given nontraded factor using a three-year rolling window regression of returns on the factor. These rolling betas are *ex-ante*, time-varying measures of risk exposure that rely only on information available at each portfolio formation time.

In each time period t, assets are sorted into portfolios according to their estimated rolling betas. We then compute the monthly return obtained by holding such portfolios. The resulting cross-section of portfolio returns provides a (nonparametric) estimate of $\mathbb{E}[r_{i,t+1}|\beta_{i,t}]$. To evaluate the pricing relevance of each factor, we focus on the return an investor would realize by buying the portfolio with the highest beta and selling the portfolio with the lowest beta. The time average of this spread gives us an estimate of the unconditional compensation for bearing the corresponding nontraded risk.

This approach has distinct advantages. It provides an intuitive and easy-to-implement mapping between risk exposures and realized returns in the form of a feasible investment strategy. Because betas are estimated using only information available at the time of portfolio formation, the resulting spreads reflect returns that could have been realized on an ex-ante basis.[19] Conversely, a key limitation is that sorting portfolios is based on estimated covariances rather than observable characteristics, therefore introducing estimation noise. We expect this problem to become more pronounced when dealing with nontraded factors, as most of them result as noisy proxies for underlying state pricing variables.

Consistent with our previous findings, the results of the portfolio sorting analysis indicate that nontraded factors do not generate statistically relevant risk premia in terms of spread. However, this finding requires careful interpretation. Poor pricing performance could arise from measurement error in the imperfect proxies used, rather than indicating that the underlying risk factors are truly unimportant. Importantly, our intention is not to advocate (in-sample) formal asset pricing tests over (out-of-sample) beta-sorted portfolios, or vice versa. Rather, these two approaches should be seen as complementary tools in our analysis. We recommend that researchers select between these methods based on the specific research question they are addressing.

5.1 Constructing portfolios

Our scientific replication of the original specifications utilizes 13 standard sets of pre-sorted test portfolios across various asset classes, which are extensively employed in asset pricing research. These portfolios

[19] While the portfolio formation itself uses only past information, some of the proposed nontraded factors are constructed using full-sample estimates (e.g., principal components or macroeconomic aggregates). As a result, a limited degree of look-ahead bias is unavoidable in these cases, although this issue is intrinsic to the original factor construction rather than the portfolio-sorting procedure. Since this bias tends to make factor performance appear stronger in-sample, our interpretation of the results is conservative with respect to the true implementable returns.

are chosen for their potential to efficiently capture cross-sectional return variation while minimizing idiosyncratic risk, making them well-suited for statistical inference and specification testing in unconditional asset pricing frameworks. However, pricing with pre-sorted portfolios entails a fundamental trade-off. While the aggregation of assets into portfolios mitigates idiosyncratic noise, it can also mitigate the variability of betas across individual securities, which in turn may dampen the cross-sectional dispersion in nontraded factor risk exposure.

To complement our previous analysis, we assess the economic performance of nontraded factors by studying the pricing problem from the perspective of an investor who builds portfolios from CRSP single stocks based on estimated risk exposures. Since individual stocks are generally more volatile and less correlated with nontraded factors than portfolios, we expect return betas' spreads to exhibit higher variability than those obtained from pre-sorted portfolios: the additional noise due to estimation error could likely generate more variation in spreads than the one due to differences in factor exposure. Along this perspective, we therefore interpret the lack of pricing power as evidence of inadequacy of the factors in terms of pricing rather than issues due individual stock characteristics.

We construct portfolios by sorting individual stocks based on their estimated betas with respect to each proposed nontraded factor. For each proposed factor at month t, we run a time-series regression of individual stock returns on the model specification, including an intercept. Betas are estimated using a rolling window of the preceding 60 months, and requiring at least 36 months of nonmissing returns for each stock. Similarly to the previous sections, we estimate the conditional betas with respect to three specifications: i) by controlling for the factors included in the original model specification; ii) by sorting with respect to the proposed nontraded factor in isolation (i.e., the single-nontradable factor models); iii) by sorting with respect to the single nontraded factor while controlling for the market factor (MKT). When models include multiple proposed factors, we estimate the betas jointly, controlling for all factors in the specification, and then portfolios are constructed sorting separately for each proposed factor. Regarding the sorting at month t, we use only information available up to that point. Portfolio returns are then measured over the subsequent month $t + 1$. This approach ensures that the resulting returns reflect truly ex-ante strategies and can be interpreted as implementable out-of-sample investment strategies.

Once constructed the portfolios associated to a proposed nontraded factor, we compute the return differential between the portfolio with the highest estimated beta and the one with the lowest beta. This long–short spread represents the investment premium associated with exposure to the given factor. We assess the statistical significance of these premiums using Newey and West (1987) adjusted standard errors. We set a lag length of three to account for serial correlation induced by overlapping returns.[20]

To explore the robustness of our results, we consider a variety of design choices that reflect common practices in the empirical asset pricing literature. Specifically, we have six baseline sorting configurations, by varying: (i) the number of portfolios formed in each month, using either five, ten, or twenty-five beta-sorted portfolios; (ii) the weighting scheme employed within each portfolio, considering both equal- and value-weighted returns. In addition to varying these specifics, we also consider different investment horizons. Portfolios are rebalanced monthly and held for periods of 1, 3, or 6 months. The longer holding horizons are implemented through overlapping portfolios to maintain a balanced monthly time series of returns. In total, combining the three portfolio sorts, two weighting schemes, and three holding periods yields 18 distinct portfolio formation strategies for a given model specification. Despite this set of design choices is not exhaustive, it captures the main empirical conventions and provides a representative benchmark.

5.2 Empirical results

Our analysis focuses on monthly nontraded factors proposed in the universe of studies examined in this paper. Among these studies, 33 papers introduce nontraded factors at the monthly frequency, resulting in a total of 83 distinct factors. Considering the three specifications described earlier, we obtain a total of 4,104 unique beta-sorted long–short investment strategies.[21]

To illustrate the influence of research design choices on estimated investment outcomes, Figure 8 presents the distribution of average long–short portfolio returns (% per month) across all specifications for each proposed nontraded factor. Each horizontal box corresponds to a distinct factor, and the width of the box captures the dispersion in average monthly long–short returns obtained under alternative empirical configurations, including differences in portfolio formation (number of portfolios and weighting schemes), and holding periods.

Figure 8 about here

Boxes centered near the zero vertical line indicate factors that typically yield economically negligible

[20]We refrain from using more complex inference methodologies (Cattaneo et al. (2022)). The reported values should be interpreted as an upper bound on statistical significance rather than as exact inference.

[21]The total number of strategies does not correspond exactly to three times the number of proposed factors because, in some cases, the original model specification coincides with either the single-factor specification or the single-factor augmented with the market factor.

premiums, whereas boxes positioned away from zero reveal factors whose estimated returns are persistently positive or negative across most implementations. Most importantly, narrow boxes denote factors whose estimated premiums are less disperse and so relatively stable across the different specifications. Conversely, wide boxes indicate a larger dispersion and therefore high degree of design sensitivity, what is referred in the literature as nonstandard errors (Walter et al., 2024).

Figure 8 reveals striking heterogeneity across factors, in both magnitude and dispersion. Some nontraded factor portfolios exhibit consistently small average returns, suggesting limited economic relevance across estimation frameworks. Others display wide interquartile ranges and numerous outliers, which in turn implies that their performance can vary substantially depending on methodological choices such as: i) the inclusion of controls, ii) the number of portfolios, or iii) the weighting scheme of portfolios. In several cases, the direction of the long–short premium even changes across specifications, indicating that the inferred pricing implication of the same factor may differ qualitatively across designs.

In general, the empirical results reveal that, out of these 4,104 strategies, only 422 (10.3%) produce average returns that are statistically different from zero at the conventional 5% significance level. This relatively small fraction suggests that, despite the large number of proposed nontraded factors in the literature, only a limited subset appears to deliver economically meaningful compensation for risk exposure once realistic portfolio formation and estimation procedures are taken into account.[22]

Table 3 summarizes the results of the beta-sorted portfolio analysis across the 33 studies proposing monthly nontraded factors. For each paper, we report the number of model specifications considered, the percentage of statistically significant long–short investment strategies (based on a 5% significance level), and the absolute mean and median investment premiums among those significant cases. Reporting absolute values allows us to focus on the economic magnitude of the estimated returns, avoiding the offsetting effects of positive and negative premiums.

Table 3 about here

Across studies, the proportion of statistically significant strategies remains generally low, and even when premiums are statistically different from zero, the corresponding premiums are economically small. The

[22] We do not adjust for multiple hypothesis testing. The qualitative conclusions are unlikely to change; if anything, this can be viewed as an upper bound on the number of significant results, since applying such corrections would reduce the set of potential discoveries.

absolute mean and median monthly returns reported in Table 3 rarely exceed 40 basis points, indicating that the magnitude of realized compensation for nontraded risk exposures is limited. [23]

When weighting each study by the number of significant specifications it contributes, the average magnitude of the significant long–short spreads across all nontraded factors is approximately 0.48% per month (about 5.8% annualized). This value should be interpreted as an indication of the small economic magnitude of realized returns among significant strategies rather than as a direct estimate of a risk premium. Overall, these results suggest that the economic impact of nontraded factor exposures, while present in some cases, tends to be moderate in size. The small magnitude of the premiums together with the low share of statistically significant outcomes points to limited economic relevance of most nontraded factors when evaluated through implementable investment strategies.

To illustrate how the estimated factor premiums vary across alternative beta specifications, Table A.7 in the Appendix decomposes the results for each study into the original, single-factor, and market-adjusted models, while Figures A.17-A.19 plot the distribution of p-values across all beta-sorted investment strategies. Each point in the figures represents the p-value associated with the average long–short return for a given model specification and study. The patterns are broadly consistent with the aggregate evidence reported in Table 3.

Taken together, these results provide an aggregate view of the economic relevance of nontraded factors once we move from in-sample cross-sectional tests to out-of-sample portfolio-based evaluation.

6 The Pricing of the Zoo of Nontradables

While the previous sections propose a replication analysis at the level of individual model (or nontraded factor) specifications, this section evaluates the pricing ability of nontradables at an aggregate level using equity portfolios at monthly frequency.

We study the explanatory power of the nontraded factor universe with respect to the cross-section of test portfolios by means of Giglio and Xiu (2021)'s methodology, therefore building tracking portfolios via Principal Components (PCs). In order to have a balanced panel, we start by considering the time series of 37 nontradables, spanning from January 1970 to December 2015, at monthly frequency. We then project

[23] While some studies exhibit a higher share of significant results, this heterogeneity should not be interpreted as contradictory evidence. Differences in sample periods, estimation windows, portfolio construction choices, and data coverage naturally contribute to variation in outcomes across studies.

each nontraded factor to the mimicking portfolios, that is, the first 7 PCs extracted from the pool of equity portfolios described in Section 3. Lastly, in light of Giglio and Xiu (2021), we summarize the linear information entailed by these 259 ($= 37 \times 7$) projected portfolios into five portfolios by extracting the first five principal components, denoted as $eqPC1 - 5$. It is important to note that extracting the principal components without the first projection might be misleading in economic terms, as the nontraded factors are not portfolios and a linear combination of such factors does not have a direct economic connotation.

To understand the linear information summarized in those portfolios, Table 4 displays the correlations with the (traded) factors in three popular pricing models: Fama and French (2015)'s five factors, Giglio and Xiu (2021)'s seven factors (GX-PC), and Lettau and Pelger (2020)'s five factors (RP-PC). We notice that the PCs of nontradables are correlated with the Fama and French (2015)'s and Lettau and Pelger (2020)'s factors, but they exhibit low correlations with Giglio and Xiu (2021)'s. This finding indicates that Giglio and Xiu (2021)'s factors may incompletely capture the risk dimensions present in the nontraded factor zoo, which could account for their empirical results of insignificant risk premia for some nontradables.

Table 4 about here

Generally speaking, we conclude that: i) the first PC ($eqPC1$) exhibits a mild correlation with RP-$PC2$; ii) the second PC ($eqPC2$) is strongly correlated with MKT and RP-$PC1$; iii) the third PC ($eqPC3$) demonstrates a strong correlation with RP-$PC3$ and a mild correlation with SMB and HML; iv) the fifth PC ($eqPC5$) also shows a strong correlation with SMB and RP-$PC4$. Interestingly, these findings align with Lettau and Pelger (2020)'s. The authors show that their proposed latent factors RP-$PC1$, RP-$PC3$, and RP-$PC5$ are highly correlated, respectively, with the CRSP value-weighted market return, the value factor, and the reversal portfolios. RP-$PC2$ is mildly correlated with a number of anomalies, and RP-$PC4$ is mainly driven by the portfolios related to momentum, value interaction, and trading frictions. Lastly, regarding the fourth PC ($eqPC4$), it seems that it is not strongly collinear with any of the aforementioned traded factors. We also find that the loadings seem to remain similar across the 37 nontraded factors when constructing the PCs. The only exceptions are $eqPC4$ and $eqPC5$, which place considerably more weights on sentiment, consumption and macroeconomic factors.

Table 5 (6 in the Appendix) present the risk premium estimates of the model using the Fama and MacBeth (1973) procedure, Giglio and Xiu (2021) methodology, and the misspecification-robust CU-GMM (Gospodinov et al. (2017)), as well as the Burnside (2016)'s bootstrap confidence interval for equity (non-equity) portfolios.[24] We find that, across different equity portfolios, $eqPC1$ tends to be priced (in particular, in the HXZ48 anomalies and US24 anomalies portfolios). We highlight that $eqPC4$ is significantly priced in the FF25 ME/MOM portfolios, despite its mild correlation (≈ -0.23) with both MOM and SMB.

Table 5 about here

7 Conclusion

After collecting 102 distinct nontraded factors from 51 papers published in top economics and finance journals over the past 40 years, we provide a comprehensive evaluation of the pricing ability of the nontraded factor zoo by combining inference robust to weak identification and model misspecification with an analysis based on beta-sorted portfolios that directly map estimated factor exposures into realized returns.

To our knowledge, it is the first study to extensively reexamine, at such a large scope, the statistical and economic properties of nontraded factors, evaluating their risk premia both in cross-sectional asset pricing tests with respect to a wide range of equity and non-equity portfolios and through implementable portfolio strategies.

Remarkably, we find that around half of the originally proposed models tend to be weakly identified. Only a minority (around 7%) of the models are considered identified by all rank tests we used. The problem of misspecification is also widespread. We document that the J-test can deliver a misleading conclusion, as the J-test rejects approximately 13% of the models, in sharp contrast to the 70% rejection rate indicated by the HJ-distance test. Furthermore, although a large proportion of nontraded factors appear priced under the standard FM approach, various robust statistics present a contrasting view: very few factors are actually priced in the test portfolios.

With respect to the original model specifications, we do not find any nontraded factor that exhibits universal pricing ability across different sets of test portfolios. However, to provide a constructive appraisal, we list the nontraded factors that are likely priced in some cross-sections of assets. The list is based on a consensus of both the identification and the misspecification-robust risk premium tests.

These conclusions are consistent with the evidence obtained from our beta-sorted portfolio analysis, which evaluates the realized economic payoffs of the same nontraded factors. Across more than 4,000 long–short strategies constructed from rolling beta estimates, only about 10% yield average returns that

[24] Note that the risk premia are identified up to a sign because of the procedure.

are statistically different from zero at the 5% level. Even among these significant cases, the corresponding long–short premiums are modest in magnitude. This limited performance suggests that most nontraded factors fail to deliver economically meaningful compensation for risk exposure.

The consistency between the weak-identification results from formal in-sample tests and the small realized premiums from out-of-sample portfolio strategies reinforces the conclusion that the majority of proposed nontraded factors have limited pricing relevance.

In an attempt to summarize the linear information contained in the nontraded factor zoo, we extract principal components from a balanced panel of 37 nontradables. We find some positive evidence on the pricing ability of the nontradables with respect to most of the equity portfolios.

References

Adrian, Tobias, Erkko Etula, and Tyler Muir, 2014, Financial intermediaries and the cross-section of asset returns, *Journal of Finance* 69, 2557–2596.

Anderson, Theodore W, and Herman Rubin, 1949, Estimation of the parameters of a single equation in a complete system of stochastic equations, *Annals of Mathematical Statistics* 20, 46–63.

Andrade, Sandro C, Adelphe Ekponon, and Alexandre Jeanneret, 2023, Sovereign risk premia and global macroeconomic conditions, *Journal of Financial Economics* 147, 172–197.

Ardia, David, Keven Bluteau, Kris Boudt, and Koen Inghelbrecht, 2022, Climate change concerns and the performance of green vs. brown stocks, *Management Science* .

Asness, Clifford S, Andrea Frazzini, and Lasse Heje Pedersen, 2019, Quality minus junk, *Review of Accounting Studies* 24, 34–112.

Atanasov, Victoria, Stig V Møller, and Richard Priestley, 2020, Consumption fluctuations and expected returns, *Journal of Finance* 75, 1677–1713.

Avramov, Doron, Si Cheng, Lior Metzker, and Stefan Voigt, 2023, Integrating factor models, *Journal of Finance* 78, 1593–1646.

Bai, Jushan, 2003, Inferential theory for factor models of large dimensions, *Econometrica* 71, 135–171.

Bai, Jushan, and Serena Ng, 2006, Evaluating latent and observed factors in macroeconomics and finance, *Journal of Econometrics* 131, 507–537.

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *Journal of Finance* 61, 1645–1680.

Balduzzi, Pierluigi, and Cesare Robotti, 2008, Mimicking portfolios, economic risk premia, and tests of multi-beta models, *Journal of Business & Economic Statistics* 26, 354–368.

Balduzzi, Pierluigi, and Cesare Robotti, 2010, Asset pricing models and economic risk premia: A decomposition, *Journal of Empirical Finance* 17, 54–80.

Bali, Turan G, Stephen J Brown, and Yi Tang, 2017, Is economic uncertainty priced in the cross-section of stock returns?, *Journal of Financial Economics* 126, 471–489.

Bali, Turan G, Robert F Engle, and Scott Murray, 2016, *Empirical asset pricing: The cross section of stock returns* (John Wiley & Sons).

Bandi, Federico M, Shomesh E Chaudhuri, Andrew W Lo, and Andrea Tamoni, 2021, Spectral factor models, *Journal of Financial Economics* 142, 214–238.

Barillas, Francisco, and Jay Shanken, 2018, Comparing asset pricing models, *Journal of Finance* 73, 715–754.

Benjamini, Yoav, and Yosef 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, Henk, Ben Jacobsen, and John B Lee, 2011, Time-varying rare disaster risk and stock returns, *Journal of Financial Economics* 101, 313–332.

Birru, Justin, and Trevor Young, 2022, Sentiment and uncertainty, *Journal of Financial Economics* 146, 1148–1169.

Boguth, Oliver, and Lars-Alexander Kuehn, 2013, Consumption volatility risk, *Journal of Finance* 68, 2589–2615.

Boguth, Oliver, and Mikhail Simutin, 2018, Leverage constraints and asset prices: Insights from mutual fund risk taking, *Journal of Financial Economics* 127, 325–341.

Boons, Martijn, Fernando Duarte, Frans De Roon, and Marta Szymanowska, 2020, Time-varying inflation risk and stock returns, *Journal of Financial Economics* 136, 444–470.

Breeden, Douglas T, 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, *Journal of Financial Economics* 7, 265–296.

Bryzgalova, Svetlana, Jiantao Huang, and Christian Julliard, 2023a, Bayesian solutions for the factor zoo: We just ran two quadrillion models, *Journal of Finance* 78, 487–557.

Bryzgalova, Svetlana, Alberto Quaini, Fabio Trojani, and Ming Yuan, 2023b, Tradable factor risk premia and oracle tests of asset pricing models, *Swiss Finance Institute Research Paper* .

Burnside, Craig, 2011, The cross section of foreign currency risk premia and consumption growth risk: Comment, *American Economic Review* 101, 3456–3476.

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

Campbell, John Y, 1996, Understanding risk and return, *Journal of Political Economy* 104, 298–345.

Campbell, John Y, Stefano Giglio, Christopher Polk, and Robert Turley, 2018, An intertemporal capm with stochastic volatility, *Journal of Financial Economics* 128, 207–233.

Campbell, John Y, and Tuomo Vuolteenaho, 2004, Bad beta, good beta, *American Economic Review* 94, 1249–1275.

Carhart, Mark M, 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Cattaneo, Matias D, Richard K Crump, and Weining Wang, 2022, Beta-sorted portfolios, *arXiv preprint arXiv:2208.10974* .

Chan, KC, Silverio Foresi, and Larry HP Lang, 1996, Does money explain asset returns? theory and empirical analysis, *Journal of Finance* 51, 345–361.

Chapman, David A, 1997, The cyclical properties of consumption growth and the real term structure, *Journal of Monetary Economics* 39, 145–172.

Chen, Andrew Y, and Tom Zimmermann, 2021, Open source cross-sectional asset pricing, *Critical Finance Review, Forthcoming* .

Chen, Jian, Guohao Tang, Jiaquan Yao, and Guofu Zhou, 2022, Investor attention and stock returns, *Journal of Financial and Quantitative Analysis* 57, 455–484.

Chen, Nai-Fu, Richard Roll, and Stephen A Ross, 1986, Economic forces and the stock market, *Journal of Business* 383–403.

Chen, Qihui, and Zheng Fang, 2019, Improved inference on the rank of a matrix, *Quantitative Economics* 10, 1787–1824.

Chen, Yong, Gregory W Eaton, and Bradley 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, Yong, Bing Han, and Jing Pan, 2021, Sentiment trading and hedge fund returns, *Journal of Finance* 76, 2001–2033.

Chen, Zilin, Zhi Da, Dashan Huang, and Liyao Wang, 2023, Presidential economic approval rating and the cross-section of stock returns, *Journal of Financial Economics* 147, 106–131.

Chordia, Tarun, Amit Goyal, and Alessio Saretto, 2020, Anomalies and false rejections, *Review of Financial Studies* 33, 2134–2179.

Cochrane, John, 2009, *Asset pricing: Revised edition* (Princeton university press).

Cochrane, John H, 2011, Presidential address: Discount rates, *Journal of Finance* 66, 1047–1108.

Cragg, John G, and Stephen G Donald, 1997, Inferring the rank of a matrix, *Journal of Econometrics* 76, 223–250.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2015, The sum of all fears investor sentiment and asset prices, *Review of Financial Studies* 28, 1–32.

Dittmar, Robert F, 2002, Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns, *Journal of Finance* 57, 369–403.

Dong, Xi, Yan Li, David E Rapach, and Guofu Zhou, 2022, Anomalies and the expected market return, *Journal of Finance* 77, 639–681.

Elton, Edwin J, Martin J Gruber, and Christopher R Blake, 1995, Fundamental economic variables, expected returns, and bond fund performance, *Journal of Finance* 50, 1229–1256.

Engelberg, Joseph, R David McLean, Jeffrey Pontiff, and Matthew C Ringgenberg, 2023, Do cross-sectional predictors contain systematic information?, *Journal of Financial and Quantitative Analysis* 58, 1172–1201.

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F, and Kenneth R French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu, 2020, Taming the factor zoo: A test of new factors, *Journal of Finance* 75, 1327–1370.

Frazzini, Andrea, and Lasse Heje Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1–25.

Giglio, Stefano, Yuan Liao, and Dacheng Xiu, 2021, Thousands of alpha tests, *Review of Financial Studies* 34, 3456–3496.

Giglio, Stefano, and Dacheng Xiu, 2021, Asset pricing with omitted factors, *Journal of Political Economy* 129, 1947–1990.

Giglio, Stefano, Dacheng Xiu, and Dake Zhang, 2022, Test assets and weak factors, Technical report, National Bureau of Economic Research.

Goldberg, Jonathan, and Yoshio Nozawa, 2021, Liquidity supply in the corporate bond market, *Journal of Finance* 76, 755–796.

Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2012, Further results on the limiting distribution of gmm sample moment conditions, *Journal of Business & Economic Statistics* 30, 494–504.

Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2014, Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors, *Review of Financial Studies* 27, 2139–2170.

Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2017, Spurious inference in reduced-rank asset-pricing models, *Econometrica* 85, 1613–1628.

Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2019, Too good to be true? fallacies in evaluating risk factor models, *Journal of Financial Economics* 132, 451–471.

Gospodinov, Nikolay, and Cesare Robotti, 2021, Capital share risk in us asset pricing: A reappraisal, *Journal of Finance, Replications and Comments* .

Grammig, Joachim, and Stephan Jank, 2016, Creative destruction and asset prices, *Journal of Financial and Quantitative Analysis* 51, 1739–1768.

Hansen, Lars Peter, 1982, Large sample properties of generalized method of moments estimators, *Econometrica: Journal of the Econometric Society* 1029–1054.

Hansen, Lars Peter, and Ravi Jagannathan, 1997, Assessing specification errors in stochastic discount factor models, *Journal of Finance* 52, 557–590.

Harvey, Campbell R, and Yan Liu, 2020, False (and missed) discoveries in financial economics, *Journal of Finance* 75, 2503–2553.

Harvey, Campbell R, Yan Liu, and Heqing Zhu, 2016, … and the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

He, Zhiguo, Bryan Kelly, and Asaf Manela, 2017, Intermediary asset pricing: New evidence from many asset classes, *Journal of Financial Economics* 126, 1–35.

Herskovic, Bernard, Bryan Kelly, Hanno Lustig, and Stijn Van Nieuwerburgh, 2016, The common factor in idiosyncratic volatility: Quantitative asset pricing implications, *Journal of Financial Economics* 119, 249–283.

Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating anomalies, *Review of Financial Studies* 33, 2019–2133.

Hu, Grace Xing, Jun Pan, and Jiang Wang, 2013, Noise as information for illiquidity, *Journal of Finance* 68, 2341–2382.

Huang, Dashan, Fuwei Jiang, Jun Tu, and Guofu Zhou, 2015, Investor sentiment aligned: A powerful predictor of stock returns, *Review of Financial Studies* 28, 791–837.

Huang, Dashan, Jiangyuan Li, and Liyao Wang, 2021, Are disagreements agreeable? evidence from information aggregation, *Journal of Financial Economics* 141, 83–101.

Huynh, Thanh D, and Ying Xia, 2021, Climate change news risk and corporate bond returns, *Journal of Financial and Quantitative Analysis* 56, 1985–2009.

Jagannathan, Ravi, and Yong Wang, 2007, Lazy investors, discretionary consumption, and the cross-section of stock returns, *Journal of Finance* 62, 1623–1661.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional capm and the cross-section of expected returns, *Journal of Finance* 51, 3–53.

Jagannathan, Ravi, and Zhenyu Wang, 1998, An asymptotic theory for estimating beta-pricing models using cross-sectional regression, *Journal of Finance* 53, 1285–1309.

Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen, 2022, Is there a replication crisis in finance?, *Journal of Finance* .

Kan, Raymond, Cesare Robotti, and Jay Shanken, 2013, Pricing model performance and the two-pass cross-sectional regression methodology, *Journal of Finance* 68, 2617–2649.

Kan, Raymond, and Chu Zhang, 1999, Two-pass tests of asset pricing models with useless factors, *Journal of Finance* 54, 203–235.

Kargar, Mahyar, 2021, Heterogeneous intermediary asset pricing, *Journal of Financial Economics* 141, 505–532.

Kleibergen, Frank, 2009, Tests of risk premia in linear factor models, *Journal of Econometrics* 149, 149–173.

Kleibergen, Frank, Lingwei Kong, and Zhaoguo Zhan, 2023, Identification robust testing of risk premia in finite samples, *Journal of Financial Econometrics* 21, 263–297.

Kleibergen, Frank, and Richard Paap, 2006, Generalized reduced rank tests using the singular value decomposition, *Journal of Econometrics* 133, 97–126.

Kleibergen, Frank, and Zhaoguo Zhan, 2015, Unexplained factors and their effects on second pass r-squared's, *Journal of Econometrics* 189, 101–116.

Kleibergen, Frank, and Zhaoguo Zhan, 2018, Identification-robust inference on risk premia of mimicking portfolios of non-traded factors, *Journal of Financial Econometrics* 16, 155–190.

Kleibergen, Frank, and Zhaoguo Zhan, 2020, Robust inference for consumption-based asset pricing, *Journal of Finance* 75, 507–550.

Kleibergen, Frank, and Zhaoguo Zhan, 2022, Misspecification and weak identification in asset pricing, *arXiv preprint arXiv:2206.13600* .

Kleibergen, Frank, and Zhaoguo Zhan, 2025, Double robust inference for continuous updating gmm, *Quantitative Economics* 16, 295–327.

Kroencke, Tim A, 2017, Asset pricing without garbage, *Journal of Finance* 72, 47–98.

Kroencke, Tim Alexander, and Julian Thimme, 2021, A skeptical appraisal of robust asset pricing tests, in *Proceedings of Paris December 2021 Finance Meeting EUROFIDAI-ESSEC*.

Kuehn, Lars-Alexander, Mikhail Simutin, and Jessie Jiaxu Wang, 2017, A labor capital asset pricing model, *Journal of Finance* 72, 2131–2178.

Lettau, Martin, and Sydney Ludvigson, 2001, Resurrecting the (c) capm: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238–1287.

Lettau, Martin, Sydney C Ludvigson, and Sai Ma, 2019, Capital share risk in us asset pricing, *Journal of Finance* 74, 1753–1792.

Lettau, Martin, Matteo Maggiori, and Michael Weber, 2014, Conditional risk premia in currency markets and other asset classes, *Journal of Financial Economics* 114, 197–225.

Lettau, Martin, and Markus Pelger, 2020, Factors that fit the time series and cross-section of stock returns, *Review of Financial Studies* 33, 2274–2325.

Linnainmaa, Juhani T, and Michael R Roberts, 2018, The history of the cross-section of stock returns, *Review of Financial Studies* 31, 2606–2649.

Liu, Yukun, and Ben Matthies, 2022, Long-run risk: Is it there?, *Journal of Finance* 77, 1587–1633.

Lustig, Hanno N, and Stijn G Van Nieuwerburgh, 2005, Housing collateral, consumption insurance, and risk premia: An empirical perspective, *Journal of Finance* 60, 1167–1219.

Manela, Asaf, and Alan Moreira, 2017, News implied volatility and disaster concerns, *Journal of Financial Economics* 123, 137–162.

Martin, Ian WR, and Christian Wagner, 2019, What is the expected return on a stock?, *Journal of Finance* 74, 1887–1929.

McLean, R David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability?, *Journal of Finance* 71, 5–32.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Papanikolaou, Dimitris, 2011, Investment shocks and asset prices, *Journal of Political Economy* 119, 639–685.

Parker, Jonathan A, and Christian Julliard, 2005, Consumption risk and the cross section of expected returns, *Journal of Political Economy* 113, 185–222.

Pástor, L'uboš, and Robert F Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

Pastor, Lubos, and Robert F Stambaugh, 2019, Liquidity risk after 20 years, Technical report, National Bureau of Economic Research.

Sadka, Ronnie, 2006, Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk, *Journal of Financial Economics* 80, 309–349.

Savov, Alexi, 2011, Asset pricing with garbage, *Journal of Finance* 66, 177–201.

Shanken, Jay, 1985, Multivariate tests of the zero-beta capm, *Journal of Financial Economics* 14, 327–348.

Shanken, Jay, 1992, On the estimation of beta-pricing models, *Review of Financial Studies* 5, 1–33.

Sweeney, Richard J, and Arthur D Warga, 1986, The pricing of interest-rate risk: evidence from the stock market, *Journal of Finance* 41, 393–410.

Vassalou, Maria, 2003, News related to future gdp growth as a risk factor in equity returns, *Journal of Financial Economics* 68, 47–73.

Vassalou, Maria, and Yuhang Xing, 2004, Default risk in equity returns, *Journal of Finance* 59, 831–868.

Walter, Dominik, Rüdiger Weber, and Patrick Weiss, 2024, Methodological uncertainty in portfolio sorts, *Available at SSRN 4164117* .

Weber, Michael, 2018, Cash flow duration and the term structure of equity returns, *Journal of Financial Economics* 128, 486–503.

Yan, Xuemin Sterling, and Lingling Zheng, 2017, Fundamental analysis and the cross-section of stock returns: A data-mining approach, *Review of Financial Studies* 30, 1382–1423.

Yogo, Motohiro, 2006, A consumption-based explanation of expected stock returns, *Journal of Finance* 61, 539–580.

Zhang, Xiang, Yangyi Liu, Kun Wu, and Bertrand Maillet, 2021, Tradable or nontradable factors—what does the hansen–jagannathan distance tell us?, *International Review of Economics & Finance* 71, 853–879.

Table 1: **Main Test Statistics.** The table presents the list of test statistics used in the main paper. For each test statistic, we report their mnemonic, a description, and their reference to the previous literature.

Testing	Mnemonic	Description	Reference
Identification	KP	Rank test	Kleibergen and Paap (2006)
	CD	Cragg-Donald rank test	Gospodinov et al. (2017)
	F-stat	Finite-sample Rank test	Gospodinov et al. (2017)
			Kleibergen and Zhan (2020)
	CF	Rank test	Chen and Fang (2019)
Specification	J-test	Hansen J-test	Hansen (1982)
	HJ	Hansen-Jagannathan distance test	Hansen and Jagannathan (1997)
Risk premium	FM	Fama-Macbeth Confidence Intervals (CIs)	Fama and MacBeth (1973)
	Shank	Shanken-corrected CIs	Jagannathan and Wang (1998)
	FM-GLS	Fama-Macbeth GLS CIs	Fama and MacBeth (1973)
	Shank-GLS	Shanken-corrected GLS CIs	Jagannathan and Wang (1998)
	CSR	Misspecification-robust CIs	Kan et al. (2013)
	CSR-GLS	Misspecification-robust GLS CIs	Kan et al. (2013)
	HJ	HJ-distance CIs	Gospodinov et al. (2014)
	CU-GMM	Continously-Updated GMM CIs	Gospodinov et al. (2014)
	GX	Giglio Xiu CIs	Giglio and Xiu (2021)
	BS	Burnside Bootstrap-based CIs	Burnside (2011)
	GRS-FAR	GRS-FAR CIs	Kleibergen and Zhan (2020)
	MPAR	Mimicking Portfolio AR CIs	Kleibergen and Zhan (2018)

Table 2: **List of priced factors in original models.** This table lists all factors whose risk premia are uniformly significant under three misspecification-robust tests (CSR-GLS, HJ, and CU-GMM), together with the associated model specifications, test assets, time-series sample size (T), number of test portfolios N, the number of pricing factors K. The reported model specifications are deemed to be identified when testing for weak identification using the CF rank test. In addition to the tests, it also reports the p-values from the FM, GX, and BS approaches.

								Misspecification-robust					
Reference	Model	Test Portfolios	Factors	T	N	K	FM	CSR–GLS	HJ	CU–GMM	GX	BS	CF
Chan et al. (1996)	$M2C_{adj}$	Options	$M2C_{adj}$	310	18	1	0.00	0.01	0.03	0.03	0.26	0.07	0.01
Chan et al. (1996)	$M2C_{unadj}$	Options	$M2C_{unadj}$	310	18	1	0.00	0.03	0.05	0.03	0.31	0.12	0.02
Chan et al. (1996)	$M3C_{adj}$	Options	$M3C_{adj}$	239	18	1	0.00	0.01	0.02	0.00	0.02	0.00	0.01
Chan et al. (1996)	$M3C_{unadj}$	Options	$M3C_{unadj}$	239	18	1	0.00	0.02	0.02	0.00	0.03	0.02	0.00
Campbell and Vuolteenaho (2004)	$NEWS_{CF}+NEWS_{DR}$	US24 anomalies	$NEWS_{CF}$	462	24	2	0.00	0.04	0.04	0.01	0.00	0.00	0.00
Parker and Julliard (2005)	Quart_PJconsumption	HXZ48 anomalies	PJconsumption	208	48	1	0.00	0.00	0.01	0.00	0.06	0.06	0.00
Baker and Wurgler (2006)	$SENT$+FF3+MOM	HXZ48 anomalies	$SENT$	660	48	5	0.00	0.02	0.04	0.03	0.13	0.00	0.00
Baker and Wurgler (2006)	$SENT$	FF25 ME/BM	$SENT$	681	25	1	0.32	0.00	0.00	0.01	0.30	0.21	0.03
Baker and Wurgler (2006)	$SENT$	HXZ48 anomalies	$SENT$	660	48	1	0.07	0.03	0.04	0.01	0.13	0.05	0.02
Sadka (2006)	$ILLIQ\_perm+MKT$	USBonds	$ILLIQ\_perm$	345	20	2	0.03	0.02	0.02	0.02	0.46	0.20	0.00
Sadka (2006)	$ILLIQ\_trans+ILLIQ\_perm+MKT$	USBonds	$ILLIQ\_perm$	345	20	3	0.01	0.02	0.04	0.02	0.46	0.16	0.00
Savov (2011)	$garbage_{ann.}$+FF3	US24 anomalies	$garbage$	43	24	4	0.00	0.00	0.02	0.01	0.09	0.00	0.00
Huang et al. (2015)	PLS_SENT_{lag}	FF25 ME/INV	PLS_SENT_{lag}	681	25	1	0.11	0.01	0.01	0.02	0.08	0.06	0.00
Huang et al. (2015)	$PLS\_SENT_{ORTH,lag}$	FF25 ME/BM	$PLS\_SENT_{ORTH,lag}$	681	25	1	0.09	0.00	0.00	0.04	0.11	0.05	0.02
Huang et al. (2015)	$PLS\_SENT_{ORTH,lag}$	FF25 ME/INV	$PLS\_SENT_{ORTH,lag}$	681	25	1	0.15	0.01	0.01	0.02	0.13	0.10	0.00
Bali et al. (2016)	UNC_1+MKT+controls	CDS	UNC_1	143	20	8	0.00	0.03	0.05	0.04	0.04	0.00	0.00
He et al. (2017)	$HKM+MKT+MOM$	US24 anomalies	HKM	587	24	3	0.00	0.00	0.00	0.01	0.02	0.00	0.00
Kroencke (2017)	$\Delta^{Q4}\hat{y}_t^{N\&S}$	HXZ48 anomalies	$\Delta^{Q4}\hat{y}_t^{N\&S}$	52	48	1	0.06	0.00	0.04	0.03	0.84	0.23	0.01
Chen et al. (2018)	$ILLIQ_{PC,adj}$	Options	$ILLIQ_{PC,adj}$	310	18	1	0.00	0.00	0.00	0.00	0.65	0.00	0.01
Chen et al. (2018)	$ILLIQ_{AMI,adj}$	Options	$ILLIQ_{AMI,adj}$	310	18	1	0.00	0.00	0.00	0.00	0.53	0.00	0.01
Chen et al. (2018)	$ILLIQ_{TO,adj}$	Options	$ILLIQ_{TO,adj}$	310	18	1	0.00	0.00	0.00	0.00	0.04	0.00	0.01
Chen et al. (2018)	$ILLIQ_{HM,adj}$	FF25 ME/MOM	$ILLIQ_{HM,adj}$	1068	25	1	0.00	0.00	0.01	0.03	0.00	0.01	0.00
Chen et al. (2018)	$ILLIQ_{HM,adj}$	Options	$ILLIQ_{HM,adj}$	310	18	1	0.00	0.00	0.02	0.03	0.86	0.00	0.00
Chen et al. (2022)	$Attention_{PCA}$	FF25 ME/INV	$Attention_{PCA}$	456	25	1	0.13	0.04	0.04	0.02	0.15	0.10	0.00
Chen et al. (2022)	$Attention_{PCA}$	CDS	$Attention_{PCA}$	143	20	1	0.00	0.01	0.02	0.02	0.16	0.06	0.04

Table 3: **Significance of nontraded beta portfolios.** The table reports summary statistics for beta-sorted portfolios of nontraded factors proposed in the literature. For each study, we construct high-minus-low portfolios based on rolling betas estiamates and compute the time-series average of their long-short (LS) returns. The columns report the total number of specifications, the percentage of statistically significant specifications, and the absolute mean and median of the significant long-short return premiums across specifications. All tests are based on monthly returns.

Paper	# Specs	% Sign.	Abs. Avg. (%)	Abs. Med. (%)
Chen et al. (1986)	360	8.61	0.296	0.282
Sweeney and Warga (1986)	36	0.00		
Campbell (1996)	54	0.00		
Chan et al. (1996)	144	11.11	0.386	0.342
Jagannathan and Wang (1996)	72	25.00	0.606	0.631
Dittmar (2002)	504	5.56	0.380	0.402
Pástor and Stambaugh (2003)	72	0.00		
Campbell and Vuolteenaho (2004)	36	2.78	0.539	0.539
Vassalou and Xing (2004)	54	5.56	0.670	0.671
Baker and Wurgler (2006)	180	54.44	0.475	0.429
Sadka (2006)	180	3.89	0.447	0.443
Berkman et al. (2011)	54	0.00		
Hu et al. (2013)	36	5.56	0.656	0.656
Da et al. (2015)	162	0.00		
Huang et al. (2015)	72	6.94	0.384	0.356
Bali et al. (2017)	360	18.33	0.580	0.600
Herskovic et al. (2016)	54	50.00	0.430	0.410
He et al. (2017)	108	35.19	0.358	0.346
Kuehn et al. (2017)	54	3.70	0.414	0.414
Manela and Moreira (2017)	36	33.33	0.478	0.464
Boguth and Simutin (2018)	54	11.11	0.285	0.276
Chen et al. (2018)	360	1.67	0.308	0.290
Martin and Wagner (2019)	72	0.00		
Boons et al. (2020)	54	0.00		
Chen et al. (2021)	54	1.85	1.012	1.012
Goldberg and Nozawa (2021)	144	13.89	0.804	0.740
Huang et al. (2021)	108	12.04	0.685	0.701
Huynh and Xia (2021)	108	0.00		
Ardia et al. (2022)	72	0.00		
Birru and Young (2022)	144	7.64	0.533	0.436
Chen et al. (2023)	54	12.96	0.441	0.439
Chen et al. (2022)	108	0.93	0.574	0.574
Liu and Matthies (2022)	144	2.08	0.505	0.457

Table 4: **Correlation Between Principal Components of Nontraded Factors and Popular Pricing Factors.** This table presents the correlations between the principal components (PCs) of nontraded factors ($eqPC1-5$) and some popular pricing factors: Fama and French (2015)'s 5 factors (MKT, SMB, HML, RMW, CMA), Giglio and Xiu (2021)'s 7 factors (GX-PC) and Lettau and Pelger (2020)'s 5 factors (RP-PC).

Panel A: Correlations with Fama and French (2015)'s five factors							
	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		
Panel B: Correlations with Giglio and Xiu (2021)'s seven factors							
	GX-$PC1$	GX-$PC2$	GX-$PC3$	GX-$PC4$	GX-$PC5$	GX-$PC6$	GX-$PC7$
$eqPC1$	-0.08	0.03	-0.06	-0.03	0.02	0.04	0.06
$eqPC2$	-0.01	-0.03	0.1	0.02	0.01	-0.14	0.12
$eqPC3$	0.06	-0.05	0.06	0.14	-0.1	0.01	-0.12
$eqPC4$	0.04	0.05	0.05	0.1	-0.09	-0.04	-0.03
$eqPC5$	-0.05	-0.06	-0.02	0	-0.03	-0.03	0.12
Panel C: Correlations with Lettau and Pelger (2020)'s five factors							
	RP-$PC1$	RP-$PC2$	RP-$PC3$	RP-$PC4$	RP-$PC5$		
$eqPC1$	-0.03	-0.77	-0.16	0.11	0.52		
$eqPC2$	-0.84	0.15	0.37	-0.2	0.29		
$eqPC3$	0.36	-0.35	0.85	0.06	0		
$eqPC4$	0.17	-0.17	-0.08	0.45	0.05		
$eqPC5$	-0.34	0.33	0.12	0.72	-0.27		

Table 5: **Pricing results of PCs of nontraded factors for equity portfolios** This table reports risk premia (γ) in percentage from the standard Fama–MacBeth regression and Giglio and Xiu (2021)'s three-pass procedure, with associated p-values in parentheses, for the first five PCs extracted from monthly nontraded factors relative to the sets of test assets used in our paper. Burnside (2016) bootstrap confidence intervals for risk premia are also reported. We further report the SDF parameter (λ) estimated via a misspecification-robust CU-GMM procedure. *** significant at the 1% level. ** significant at the 5% level. * significant at the 10% 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
γ(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
	(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)
γ(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
	(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)
λ(CU-GMM)	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
	(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)
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]

	FF25 ME/MOM					HXZ48 anomalies					US24 anomalies				
	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5
γ(FM)	0.00	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
	(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)
γ(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
	(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)
λ(CU-GMM)	-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
	(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)
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]	[-0.20,0.02]	[-0.12,0.13]	[-0.12,0.22]	[-0.08,0.16]

	Weber10 duration				
	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5
γ(FM)	-0.76**	-1.09	0.78**	-0.26	0.35
	(0.04)	(0.12)	(0.02)	(0.54)	(0.33)
γ(GX)	-0.49**	-0.44**	0.61**	0.19**	-0.19**
	(0.01)	(0.01)	(0.01)	(0.04)	(0.04)
λ(CU-GMM)	221.28	827.58	-238.31	510.19	-418.47
	(0.74)	(0.75)	(0.70)	(0.77)	(0.77)
Burnside CI	[-1.48,0.74]	[-2.22,1.73]	[-0.43,1.58]	[-1.48,1.43]	[-0.99,1.38]

Table 6: **Pricing results of PCs of nontraded factors for non-equity portfolios**

This table reports risk premia (γ) in percentage from the standard Fama–MacBeth regression and Giglio and Xiu (2021)'s procedure, with associated p-values in parentheses, for the first five PCs extracted from monthly nontraded factors relative to the sets of test assets used in our paper. Burnside (2016) bootstrap confidence intervals for risk premia are also reported. We further report the SDF parameter (λ) estimated via a misspecification-robust CU-GMM procedure. *** significant at the 1% level. ** significant at the 5% level. * significant at the 10% level.

	CDS					Commodity					FX				
	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5
γ(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**
	(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)
γ(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
	(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)
λ(CU-GMM)	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
	(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)
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]	[-2.60,2.99]	[-3.06,1.15]

	Options					SovBonds					USBonds				
	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5	eqPC1	eqPC2	eqPC3	eqPC4	eqPC5
γ(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
	(0.84)	(0.04)	(0.17)	(0.04)	(0.00)						(0.04)	(0.11)	(0.56)	(0.05)	(0.36)
γ(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
	(0.37)	(0.46)	(0.49)	(0.57)	(0.18)						(0.08)	(0.01)	(0.01)	(0.03)	(0.64)
λ(CU-GMM)	2865.19	-961.07	-2206.77	92.98	-1103.54						1683.34	695.92	-10.94	3599.95	-918.35
	(0.90)	(0.90)	(0.90)	(0.97)	(0.90)						(0.92)	(0.91)	(0.99)	(0.92)	(0.92)
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]	[-16.06,22.80]	[-19.16,20.80]	[-25.67,15.28]	[-1.44,1.02]	[-0.88,0.81]	[-0.52,1.01]	[-1.80,2.76]	[-1.59,1.84]

Figure 1: **Identification failure of original models.** This figure shows the percentages of identification failure in model specifications that are proposed by the original papers. The rank tests used are Kleibergen and Paap (2006) rank test (KP), finite-sample rank test (F-stat) used in Kleibergen and Zhan (2020) as well as the test excluding intercept (F-stat (excl. intercept)), Cragg-Donald rank test (CD) used in Gospodinov et al. (2017), and Chen and Fang (2019) rank test (CF). We also consider the most conservative situation where we can conclude a model is identified only when all of four test statistics can reject the null of rank deficiency, and the favorable situation where any of the test statistics can reject. We also show the percentages calculated with respect to time series sample size $T \leq 100$ (small), $100 < T \leq 400$ (medium) and $T > 400$ (large).

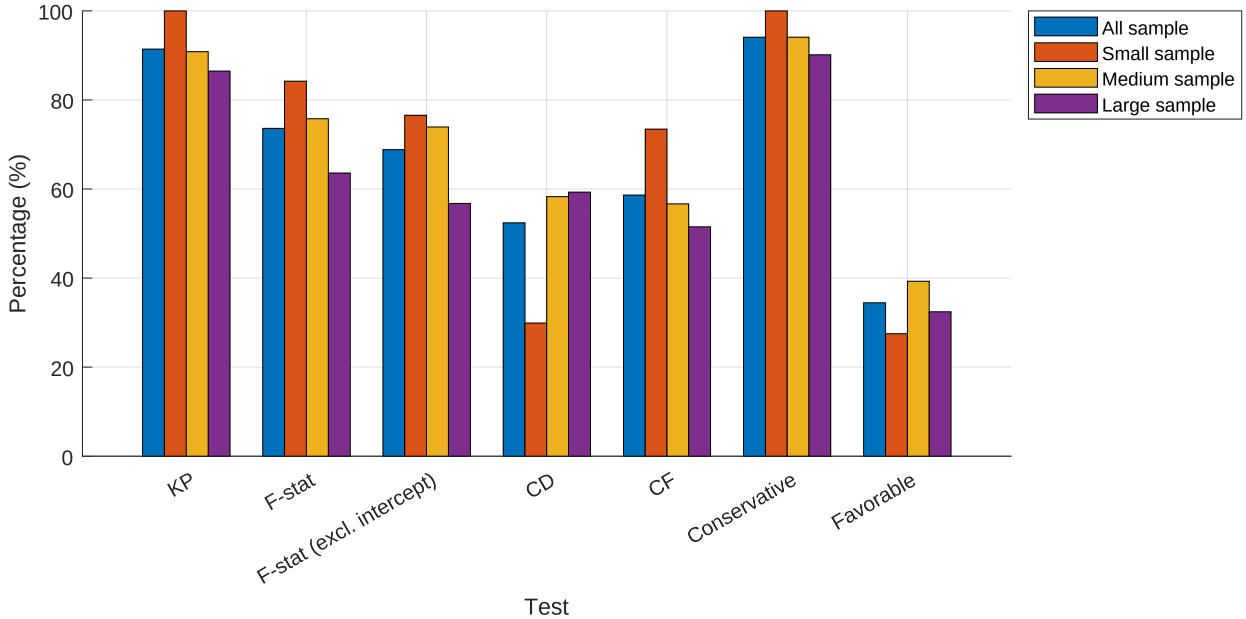


Figure 2: **Identification failure of model specifications with equity portfolios.** This figure presents The percentages of identification failure for different sets of equity portfolios. We use four rank tests: Kleibergen and Paap (2006) rank test (KP), Kleibergen and Zhan (2020) finite-sample rank test(F-stat) and the test excluding intercept (F-stat (excl. intercept)), Gospodinov et al. (2017) Cragg-Donald rank test (CD), as well as Chen and Fang (2019) rank test (CF).

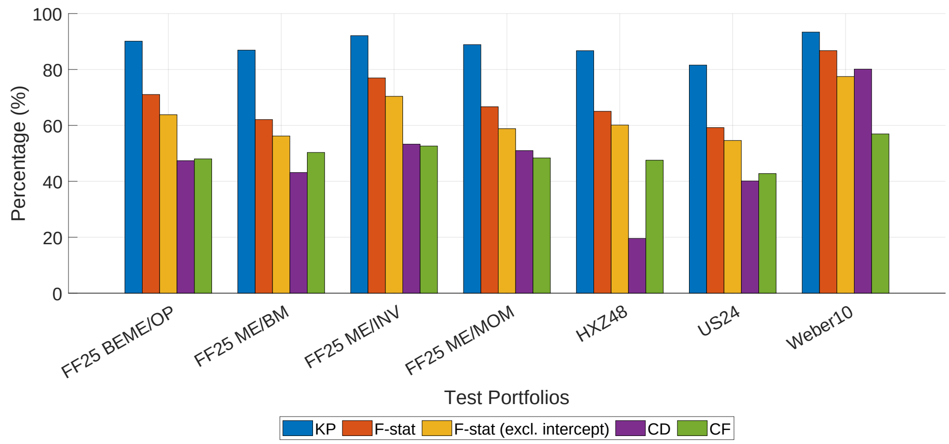


Figure 3: **Identification failure of model specifications with non-equity portfolios.** This figure presents The percentages of identification failure for different sets of non-equity portfolios. We use four rank tests: Kleibergen and Paap (2006) rank test (KP), Kleibergen and Zhan (2020) finite-sample rank test(F-stat) and the test excluding intercept (F-stat(excl. intercept)), Gospodinov et al. (2017) Cragg-Donald rank test (CD), as well as Chen and Fang (2019) rank test (CF).

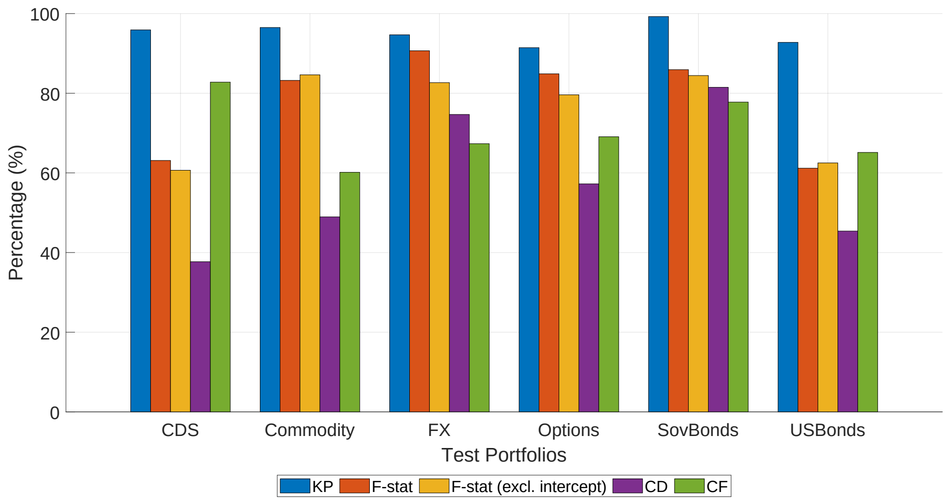


Figure 4: **Identification failure of model specifications in each category of factors.** This figure presents the percentages of identification failure for different categories of factors. We use four rank tests: Kleibergen and Paap (2006) rank test (KP), Kleibergen and Zhan (2020) finite-sample rank test(F-stat) and the test excluding intercept (F-stat (excl. intercept)), Gospodinov et al. (2017) Cragg-Donald rank test (CD), as well as Chen and Fang (2019) rank test (CF).

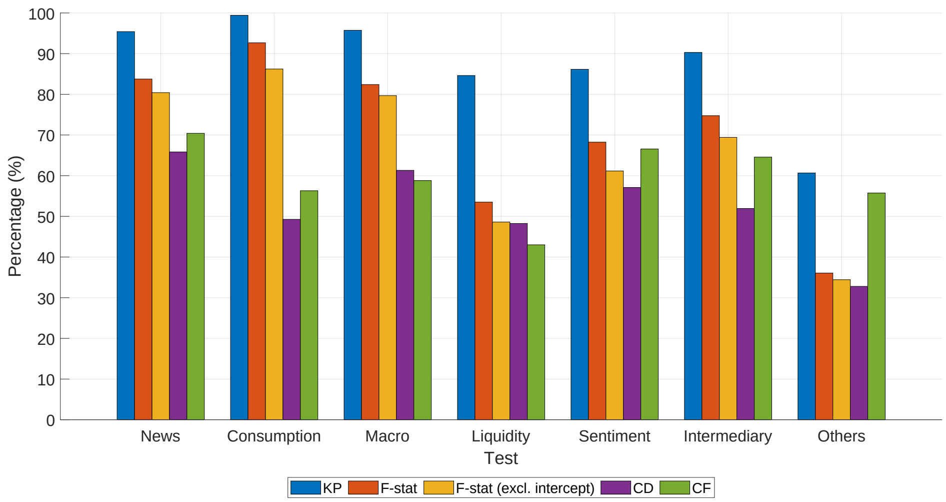


Figure 5: **Identification failure of models with and without controlling market factor.** This figure shows the percentages of identification failure for model specifications with and without market factor. We use five rank tests: Kleibergen and Paap (2006) rank test (KP), Kleibergen and Zhan (2020) finite-sample rank test(F-stat and F-stat (excl. intercept)) with and without intercept, Gospodinov et al. (2017) Cragg-Donald rank test (CD), and Chen and Fang (2019) rank test (CF).

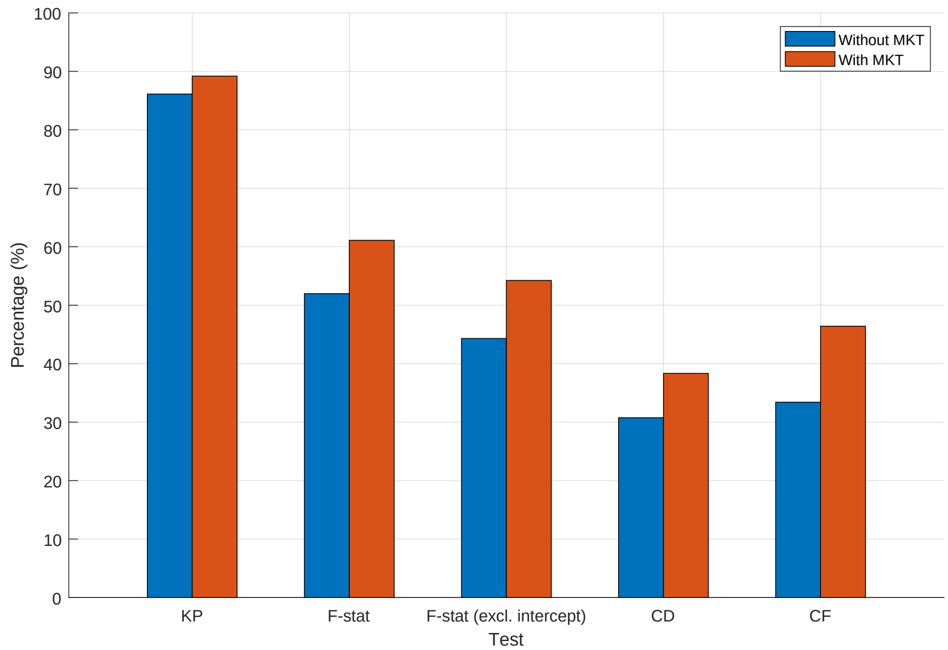
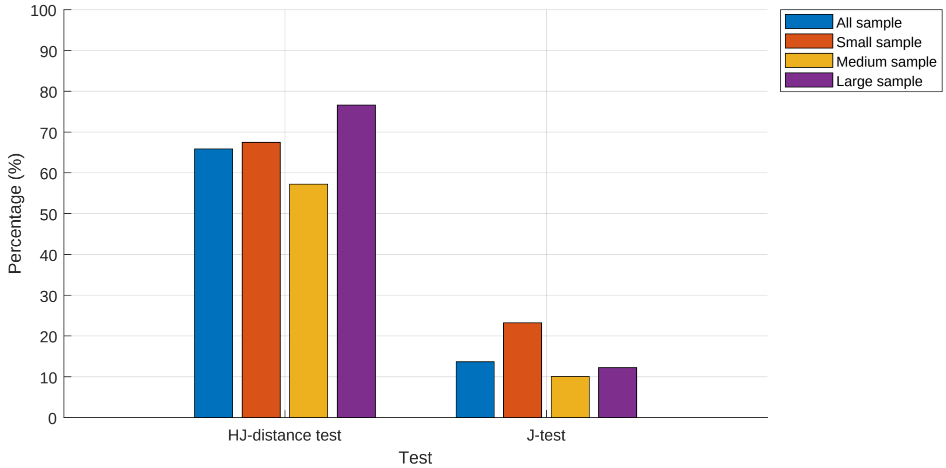


Figure 6: **The percentages of model misspecification.** This figure reports the proportion of model misspecification identified by the over-identifying restriction J-test and the Hansen-Jagannathan (HJ) distance test. The percentages represent the proportion of models that reject the null hypothesis of correct model specification. We also compare The percentages across number of time series observations $T \leq 100$ (small), $100 < T \leq 400$ (medium) and $T > 400$ (large).



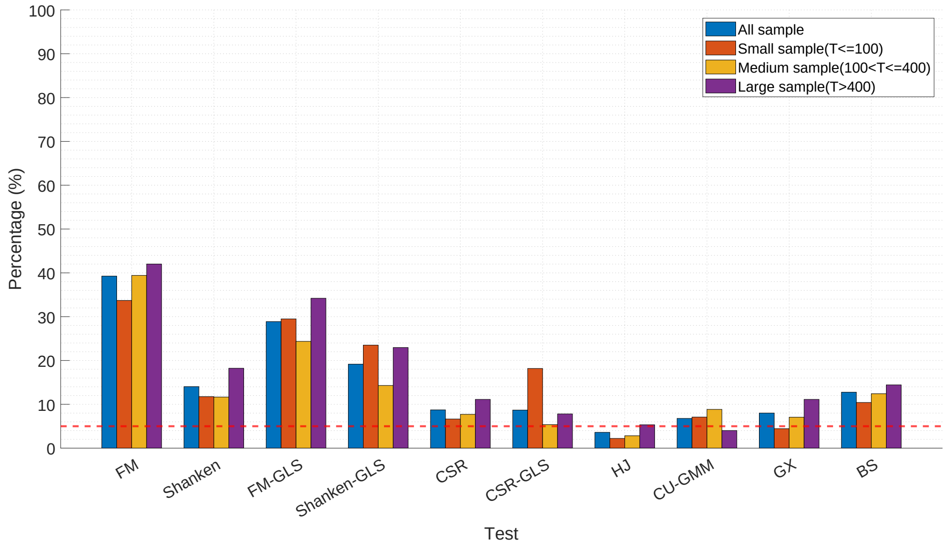


Figure 7: **The percentages of priced factors at the 5% significance level**
The figure shows the percentages of priced factors using the 5% significance level pooling across all model specifications and test assets. The percentages are based on the rejection of the null hypothesis that the risk premium is equal to zero. The test statistics are calculated using the standard errors from Fama-MacBeth approach (FM and FM-GLS), Jagannathan and Wang (1998) generalized Shanken's errors-in-variables allowing for conditional heteroskedasticity (Shanken and Shanken-GLS), as well as misspecification-robust standard errors in Kan et al. (2013) (CSR and CSR-GLS), in Gospodinov et al. (2014) based on Hansen-Jagannathan distance (HJ) and in Gospodinov et al. (2017) in linear SDF estimated using continuously updated GMM (CU-GMM). We also report the percentages of rejection with Giglio and Xiu (2021) three-pass approach and Burnside (2011) bootstrap approach. For each test, we also present the results with different time series sample sizes $T \leq 100$ (small), $100 < T \leq 400$ (medium) and $T > 400$ (large).

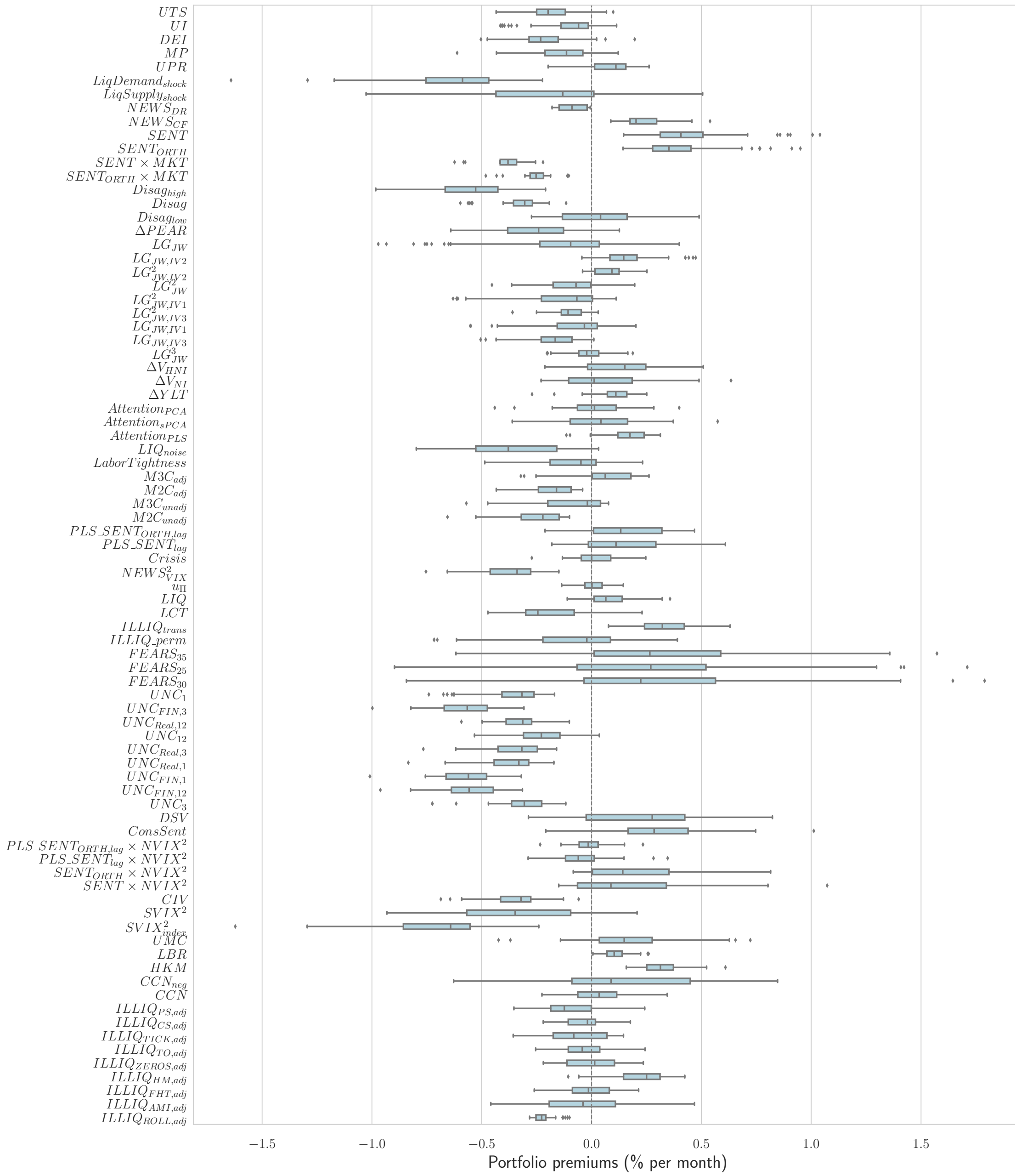


Figure 8: **Distribution of average long–short returns across empirical designs.**
The figure presents the cross-specification distribution of average long–short portfolio returns for each proposed nontraded factor. Each box summarizes the dispersion of mean monthly returns (in percent) across all portfolio construction choices and beta estimation frameworks, including variations in the number of portfolios, weighting schemes, holding horizons, and control factors. Boxes that are narrow indicate factors whose economic performance is stable across empirical implementations, while wide boxes reveal strong sensitivity to design choices.

Appendix

List of Nontraded Factors

The following tables list the nontraded factors that we include in our analysis. For the sake of scientific replication, we highlight that we did not replicate the factors that are already available on the authors' websites. For further details, we invite the readers to refer to their papers or to the personal websites 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 nontraded factors are grouped into six broad categories based on the risk dimensions the factors are measuring: *Consumption* (Table A.1), *Macroeconomics* (Table A.2), *News* (Table A.3), *Sentiment* (Table A.4), *Liquidity* (Table A.5), *Intermediary and aggregate firm-level risk* (Table A.6). For each factor, we list its short name, brief description, and reference. The short name is the label used in our dataset and in our replication package.

Table A.1: **Consumption**

Short name	**Factor description**	**Reference**	**Label**
$\hat{g}_t$	Polynomials of the growth of personal consumption expenditures	Chapman (1997)	C1997
$\hat{g}_t^2$	Polynomials of the growth of personal consumption expenditures	Chapman (1997)	C1997
$\hat{\tau}$	Polynomials of the growth of temporary technology shock	Chapman (1997)	C1997
$\hat{\tau}^2$	Polynomials of the growth of temporary technology shock	Chapman (1997)	C1997
$PersC_t$	(Quarterly/Annual) real personal consumption expenditures on non-durable goods per capita from NIPA	Parker and Julliard (2005)	JP2005
$\Delta c_{t,DUR}$	Durable goods consumption	Yogo (2006)	Y2006
$\Delta c_{t,Q4}$	Yearly consumption growth based on the fourth quarter (Q4)	Jagannathan and Wang (2007)	JW2007
$garbage$	Per capita garbage growth	Savov (2011)	S2011
$\Delta\hat{\mu}_t$	Changes in beliefs about the conditional mean of consumption growth	Boguth and Kuehn (2013)	BK2013
$\Delta\hat{\sigma}_t$	Changes in beliefs about consumption growth volatility	Boguth and Kuehn (2013)	BK2013
$\Delta\hat{y}_t^{N\&S}$	Annual growth of unfiltered NIPA N&S consumption	Kroencke (2017)	K2017
$\Delta\hat{y}_t^{N}$	Annual growth of unfiltered NIPA nondurables consumption	Kroencke (2017)	K2017
$\Delta^{Q4}\hat{y}_t^{N\&S}$	Annual growth of unfiltered Q4-to-Q4 NIPA N&S consumption	Kroencke (2017)	K2017
$\Delta^{Q4}\hat{y}_t^{N}$	Annual growth of unfiltered Q4-to-Q4 NIPA nondurables consumption	Kroencke (2017)	K2017
cc_t	Cyclical consumption	Atanasov et al. (2020)	AMP2020
$\Delta\mathbb{E}(Cons_{SPF})$	Changes in mean consumption growth forecast (SPF)	Andrade et al. (2023)	AEJ2023
$\Delta\sigma(Cons_{SPF})$	Changes in interquartile range of consumption growth forecast (SPF)	Andrade et al. (2023)	AEJ2023
PC_{SPF}	First principal component of $\mathbb{E}(Cons_{SPF})$ and $\sigma(Cons_{SPF})$	Andrade et al. (2023)	AEJ2023

Notes: N&S = nondurables and services; NIPA = National Income and Product Accounts; SPF = Survey of Professional Forecasters; Q4 = fourth quarter.

Table A.2: **Macroeconomics**

Short name	**Factor description**	**Reference**	**Label**
ΔYLT	Change in long-term government bond yield	Sweeney and Warga (1986)	SW1986
MP	Monthly growth of industrial production	Chen et al. (1986)	CRR1986
DEI	Change in expected inflation	Chen et al. (1986)	CRR1986
UI	Unexpected inflation	Chen et al. (1986)	CRR1986
UPR	Difference between BAA-rated corporate and long-term government bond yields	Chen et al. (1986)	CRR1986
UTS	Term structure spread	Chen et al. (1986)	CRR1986
$\Delta\pi_{MICH}$	Change in inflation expectation (Michigan survey)	Elton et al. (1995)	EGB1995
ΔGDP_{surv}	Change in expectation in GNP/GDP survey	Elton et al. (1995)	EGB1995
GNP	Real GNP growth rate	Campbell (1996)	C1996
LBR	Real labor income growth	Campbell (1996)	C1996
$M2C_{unadj}$, $M2C_{adj}$	Seasonally unadjusted/adjusted per-capita inside money (growth)	Chan et al. (1996)	CFL1996
$M3C_{unadj}$, $M3C_{adj}$	Seasonally unadjusted/adjusted per-capita inside money (growth)	Chan et al. (1996)	CFL1996
LG_{JW}	Per-capita labor income growth lagged one month (to match data release lag)	Jagannathan and Wang (1996)	JW1996
cay	Consumption–aggregate wealth ratio	Lettau and Ludvigson (2001)	LL2001
$cay \times MKT$	Market factor conditional on *cay*	Lettau and Ludvigson (2001)	LL2001
$cay \times LaborGrowth$	Labor income growth conditional on *cay*	Lettau and Ludvigson (2001)	LL2001
$LG_{JW,IV1}$, $LG_{JW,IV2}$, $LG_{JW,IV3}$, $LG^2_{JW,IV1}$, $LG^2_{JW,IV2}$, $LG^2_{JW,IV3}$	Polynomials of labor income growth	Dittmar (2002)	D2002
$myfar$	Fixed-assets-based ratio of housing to human wealth	Lustig and Van Nieuwerburgh (2005)	LN2005
$mymor$	Mortgage-based ratio of housing to human wealth	Lustig and Van Nieuwerburgh (2005)	LN2005
$myrwr$	Residential-wealth-based ratio of housing to human wealth	Lustig and Van Nieuwerburgh (2005)	LN2005
$Crisis$	Unexpected disaster risk	Berkman et al. (2011)	BJL2011
IMC	Investment-minus-consumption producers' innovations	Papanikolaou (2011)	–
IS_{PI}	Relative price of new equipment	Papanikolaou (2011)	–
PAG	Net growth rate of patenting activity (Quarterly/Annual)	Grammig and Jank (2016)	GJ2016
UNC	Macro uncertainty	Bali et al. (2017)	BBT2016
UNC_{FIN}	Financial uncertainty	Bali et al. (2017)	BBT2016
UNC_{Real}	Real uncertainty	Bali et al. (2017)	BBT2016

Continued on next page

Table A.2 — *continued*

Short name	**Factor description**	**Reference**	**Label**
$LaborTightness$	Labor tightness	Kuehn et al. (2017)	KSW2017
ΔKS	Non-overlapping, one-period growth in capital share	Lettau et al. (2019)	LLM2019
u_{Π}	Inflation innovation from ARMA(1,1)	Boons et al. (2020)	BDRS2020
$PC1_4$	First principal component of: changes in mean and IQR of consumption growth forecast, change in macroeconomic index, and lagged yield-curve slope	Andrade et al. (2023)	AEJ2023

Table A.3: **News**

Short name	**Factor description**	**Reference**	**Label**
$NEWS_{DR}$	Discount-rate news	Campbell and Vuolteenaho (2004)	CV2004
$NEWS_{CF}$	Cash-flow news	Campbell and Vuolteenaho (2004)	CV2004
$NEWS_{DR2}$	Discount-rate news	Campbell et al. (2018)	CGPT2017
$NEWS_{CF2}$	Cash-flow news	Campbell et al. (2018)	CGPT2017
$NEWS_{\sigma(r)}$	News about market return variance	Campbell et al. (2018)	CGPT2017
$NEWS^2_{VIX}$	News-implied volatility	Manela and Moreira (2017)	MM2017
CCN	Residuals from an AR(1) model of WSJ Climate Change News Index	Huynh and Xia (2021)	HX2021
CCN_{neg}	Residuals from an AR(1) model of CH Negative Climate Change News Index	Huynh and Xia (2021)	HX2021
ΔV_{NI}	AR(1) innovation of Wall Street Journal news index about economic growth prospects (N-index)	Liu and Matthies (2022)	LM2022
ΔV_{HNI}	AR(1) innovation of filtered N-index (HN-index)	Liu and Matthies (2022)	LM2022
UMC	Unexpected media climate change concerns	Ardia et al. (2022)	ABBI2022

Table A.4: **Sentiment**

Short name	Factor description	Reference	Label
$SENT$	Investor sentiment	Baker and Wurgler (2006)	BW2006
$SENT_{ORTH}$	Investor sentiment orthogonal to macro variables	Baker and Wurgler (2006)	BW2006
$SENT \times MKT$	Interaction between investor sentiment and market return	Baker and Wurgler (2006)	BW2006
$SENT_{ORTH} \times MKT$	Interaction between orthogonal investor sentiment and market return	Baker and Wurgler (2006)	BW2006
PLS_SENT_{lag}	Investor sentiment index lagged one month	Huang et al. (2015)	HJTZ2015
$PLS\_SENT_{ORTH,lag}$	Investor sentiment index orthogonal to macro variables, lagged one month	Huang et al. (2015)	HJTZ2015
$FEARS_{30}$	Avg. change in web-search volume from the 30 terms with the strongest negative association to market returns	Da et al. (2015)	DEG2015
$FEARS_{35}$	Avg. change in web-search volume from the 35 terms with the strongest negative association to market returns	Da et al. (2015)	DEG2015
$FEARS_{25}$	Avg. change in web-search volume from the 25 terms with the strongest negative association to market returns	Da et al. (2015)	DEG2015
$Disag_{high}$	Disagreement in high-sentiment periods	Huang et al. (2021)	HLW2021
$Disag_{low}$	Disagreement in low-sentiment periods	Huang et al. (2021)	HLW2021
$Disag$	Disagreement index	Huang et al. (2021)	HLW2021
$ConsSent$	Monthly change in Michigan US consumer sentiment index	Chen et al. (2021)	CHP2021
$SENT \times NVIX^2$	Interaction between Baker and Wurgler (2006) investor sentiment and Manela and Moreira (2017) news-implied volatility (squared)	Birru and Young (2022)	BY2022
$SENT_{ORTH} \times NVIX^2$	Interaction between Baker and Wurgler (2006) orthogonal sentiment and Manela and Moreira (2017) news-implied volatility (squared)	Birru and Young (2022)	BY2022
$PLS\_SENT_{lag} \times NVIX^2$	Interaction between Huang et al. (2015) investor sentiment and Manela and Moreira (2017) news-implied volatility (squared)	Birru and Young (2022)	BY2022
$PLS\_SENT_{ORTH,lag} \times NVIX^2$	Interaction between Huang et al. (2015) orthogonal sentiment and Manela and Moreira (2017) news-implied volatility (squared)	Birru and Young (2022)	BY2022
$Attention_{PCA}$	First principal component from 12 attention proxies using PCA	Chen et al. (2022)	CTYZ2023
$Attention_{PLS}$	First principal component from 12 attention proxies using PLS	Chen et al. (2022)	CTYZ2023
$Attention_{sPCA}$	First principal component from 12 attention proxies using sPCA	Chen et al. (2022)	CTYZ2023
$\Delta PEAR$	Monthly change in presidential economic approval rating index	Chen et al. (2023)	CDHW2023

Notes: NVIX = News-Implied Volatility; PCA = principal component analysis; PLS = partial least squares; sPCA = sparse PCA.

Table A.5: **Liquidity**

Short name	**Factor description**	**Reference**	**Label**
LIQ	Aggregate stock market liquidity	Pástor and Stambaugh (2003)	PS2003
$ILLIQ_{trans}$	Transitory illiquidity	Sadka (2006)	S2006
$ILLIQ_{perm}$	Permanent illiquidity	Sadka (2006)	S2006
LIQ_{noise}	Monthly change of a market-wide liquidity measure derived from the U.S. Treasury market	Hu et al. (2013)	HPW2013
$ILLIQ_{ROLL,adj}$	Aggregate stock market illiquidity: structural-break and volatility-adjusted Roll measure (ROLL)	Chen et al. (2018)	CEP2018
$ILLIQ_{CS,adj}$	Aggregate stock market illiquidity: structural-break and volatility-adjusted Corwin–Schultz measure (CS)	Chen et al. (2018)	CEP2018
$ILLIQ_{FHT,adj}$	Aggregate stock market illiquidity: structural-break and volatility-adjusted Fong–Holden–Trzcinka measure (FHT)	Chen et al. (2018)	CEP2018
$ILLIQ_{TICK,adj}$	Aggregate stock market illiquidity: structural-break and volatility-adjusted effective tick measure (TICK)	Chen et al. (2018)	CEP2018
$ILLIQ_{ZEROS,adj}$	Aggregate stock market illiquidity: structural-break and volatility-adjusted zeros measure (ZEROS)	Chen et al. (2018)	CEP2018
$ILLIQ_{AMI,adj}$	Aggregate stock market illiquidity: structural-break and volatility-adjusted Amihud illiquidity measure (AMI)	Chen et al. (2018)	CEP2018
$ILLIQ_{TO,adj}$	Aggregate stock market illiquidity: structural-break and volatility-adjusted Amihud measure based on the ratio of absolute daily returns to daily turnover	Chen et al. (2018)	CEP2018
$ILLIQ_{PS,adj}$	Aggregate stock market illiquidity: structural-break and volatility-adjusted Pastor–Stambaugh measure (PS)	Chen et al. (2018)	CEP2018
$ILLIQ_{HM,adj}$	Aggregate stock market illiquidity: structural-break and volatility-adjusted Hou–Moskowitz measure (HM)	Chen et al. (2018)	CEP2018
$ILLIQ_{PC,adj}$	First principal component of the break- and volatility-adjusted illiquidity measures above	Chen et al. (2018)	CEP2018
$LiqSupply_{shock}$	Liquidity supply shock in the corporate bond market	Goldberg and Nozawa (2021)	GN2021
$LiqDemand_{shock}$	Liquidity demand shock in the corporate bond market	Goldberg and Nozawa (2021)	GN2021

Notes: ROLL = Roll (1984); CS = Corwin–Schultz (2012); FHT = Fong–Holden–Trzcinka (2017); TICK = effective tick; ZEROS = zero-return days; AMI = Amihud (2002); PS = Pastor–Stambaugh (2003); HM = Hou–Moskowitz (2005).

Table A.6: **Intermediary, Volatility, and Aggregate Firm-level Risk**

Short name	**Factor description**	**Reference**	**Label**
Leverage	(Seasonally adjusted) log changes in the level of broker–dealer leverage	Adrian et al. (2014)	AEM2014
HKM	Intermediary capital risk factor (monthly and quarterly)	He et al. (2017)	HKM2017
LCT	AR(1) residuals of the value-weighted average beta of aggregate stock holdings of all actively managed equity funds (leverage-constraint tightness)	Boguth and Simutin (2018)	BS2018
$HIfac$	Heterogeneous intermediary factor	Kargar (2021)	K2021
DSV	Equally weighted average of firm default likelihood	Vassalou and Xing (2004)	VX2004
CIV	Innovation in common idiosyncratic volatility (equally weighted average of firm-level market-model residual return variance)	Herskovic et al. (2016)	HKLV2016
$SVIX^2$	Value-weighted average of stock risk-neutral variance	Martin and Wagner (2019)	MW2019
$SVIX^2_{index}$	Risk-neutral variance of the market index	Martin and Wagner (2019)	MW2019

Table A.7: **Significance of nontraded beta portfolios by specification.** The table reports summary statistics for beta-sorted portfolios of nontraded factors proposed in the literature. For each study, we construct high-minus-low portfolios based on rolling betas estiamates and compute the time-series average of their long-short (LS) returns. The columns report the total number of specifications, the percentage of statistically significant specifications, and the absolute mean and median of the significant long-short return premiums across originals, single-factor, and market specifications. All tests are based on monthly returns.

Reference	Type	# Specs	% Sign.	Avg. of Sign. Premiums (%)	Med. of Sign. Premiums (%)
Chen et al. (1986)	original	180	5.56	0.271	0.291
	single	90	7.78	0.246	0.252
	single + MKT	90	15.56	0.339	0.318
Chan et al. (1996)	original	72	9.72	0.411	0.417
	single + MKT	72	12.50	0.365	0.335
Jagannathan and Wang (1996)	original	54	27.78	0.575	0.620
	single	18	16.67	0.766	0.750
Dittmar (2002)	original	324	7.72	0.377	0.404
Campbell and Vuolteenaho (2004)	original	36	2.78	0.539	0.539
Vassalou and Xing (2004)	original	36	8.33	0.670	0.671
Baker and Wurgler (2006)	original	144	51.39	0.475	0.431
	single + MKT	36	66.67	0.476	0.399
Sadka (2006)	original	144	3.47	0.419	0.426
	single	36	5.56	0.518	0.518
Hu et al. (2013)	original	18	5.56	0.701	0.701
	single	18	5.56	0.610	0.610
Huang et al. (2015)	original	36	8.33	0.431	0.376
	single + MKT	36	5.56	0.314	0.314
Bali et al. (2017)	original	36	44.44	0.461	0.416
	single	162	13.58	0.713	0.665
	single + MKT	162	17.28	0.544	0.530
Herskovic et al. (2016)	original	18	50.00	0.451	0.410
	single	18	55.56	0.464	0.442
	single + MKT	18	44.44	0.364	0.323
He et al. (2017)	original	90	42.22	0.358	0.346
Kuehn et al. (2017)	single	18	11.11	0.414	0.414
Manela and Moreira (2017)	original	18	11.11	0.683	0.683
	single + MKT	18	55.56	0.436	0.457
Boguth and Simutin (2018)	original	54	11.11	0.285	0.276
Chen et al. (2018)	single + MKT	180	3.33	0.308	0.290
Chen et al. (2021)	single	18	5.56	1.012	1.012
Goldberg and Nozawa (2021)	original	108	18.52	0.804	0.740
Huang et al. (2021)	original	54	14.81	0.803	0.805
	single + MKT	18	11.11	0.303	0.303
Birru and Young (2022)	single + MKT	72	15.28	0.533	0.436
Chen et al. (2023)	original	36	16.67	0.434	0.404
	single + MKT	18	5.56	0.488	0.488
Chen et al. (2022)	single + MKT	54	1.85	0.574	0.574
Liu and Matthies (2022)	original	144	2.08	0.505	0.457

Table A.8: **Top Three Strategies by Factor (Absolute Returns)**

Factor	Strategy	Reference	Premium (%)	SR (annual)
$UNC_{FIN,1}$	`25|EW|1m`	Bali et al. (2017)	−1.009 (−2.15)	−0.398
$UNC_{FIN,3}$	`25|EW|1m`	Bali et al. (2017)	−0.998 (−2.13)	−0.394
$UNC_{FIN,12}$	`25|EW|1m`	Bali et al. (2017)	−0.962 (−2.03)	−0.379
LCT	`Dec.|EW|1m`	Boguth and Simutin (2018)	−0.369 (−2.21)	−0.425
LCT	`Dec.|EW|1m`	Boguth and Simutin (2018)	−0.353 (−2.12)	−0.410
LCT	`Dec.|EW|3m`	Boguth and Simutin (2018)	−0.307 (−1.97)	−0.376
$SENT$	`25|VW|1m`	Baker and Wurgler (2006)	1.040 (3.02)	0.446
$SENT$	`25|VW|1m`	Baker and Wurgler (2006)	1.006 (3.06)	0.436
$SENT_{ORTH}$	`25|VW|1m`	Baker and Wurgler (2006)	0.951 (2.86)	0.410
$SENT \times NVIX^2$	`25|VW|1m`	Birru and Young (2022)	1.073 (2.94)	0.457
$SENT_{ORTH} \times NVIX^2$	`25|VW|1m`	Birru and Young (2022)	0.815 (2.37)	0.351
$SENT \times NVIX^2$	`25|VW|3m`	Birru and Young (2022)	0.803 (2.52)	0.382
$\Delta PEAR$	`25|VW|3m`	Chen et al. (2023)	−0.641 (−2.14)	−0.324
$\Delta PEAR$	`Dec.|VW|1m`	Chen et al. (2023)	−0.545 (−2.03)	−0.376
$\Delta PEAR$	`Dec.|VW|1m`	Chen et al. (2023)	−0.488 (−1.98)	−0.345
$ILLIQ_{HM,adj}$	`Quint.|VW|1m`	Chen et al. (2018)	0.425 (2.79)	0.413
$ILLIQ_{HM,adj}$	`Quint.|VW|3m`	Chen et al. (2018)	0.339 (2.24)	0.337
$ILLIQ_{HM,adj}$	`Quint.|VW|6m`	Chen et al. (2018)	0.307 (2.08)	0.313
$M2C_{unadj}$	`Dec.|VW|1m`	Chan et al. (1996)	−0.656 (−3.85)	−0.520
$M3C_{unadj}$	`25|VW|6m`	Chan et al. (1996)	−0.570 (−2.19)	−0.342
$M2C_{unadj}$	`Dec.|VW|1m`	Chan et al. (1996)	−0.527 (−2.88)	−0.395

Continued on next page

Table A.8 — *continued*

Factor	Strategy	Reference	Premium (%)	SR (annual)
$ConsSent$	`25|VW|1m`	Chen et al. (2021)	1.012 (2.09)	0.381
MP	`25|VW|1m`	Chen et al. (1986)	−0.612 (−2.12)	−0.278
DEI	`25|VW|3m`	Chen et al. (1986)	−0.504 (−2.30)	−0.294
DEI	`25|VW|6m`	Chen et al. (1986)	−0.475 (−2.23)	−0.289
$Attention_{sPCA}$	`25|EW|1m`	Chen et al. (2022)	0.574 (2.13)	0.362
$NEWS_{CF}$	`25|VW|1m`	Campbell and Vuolteenaho (2004)	0.539 (2.16)	0.331
$LG^2_{JW,IV1}$	`25|VW|1m`	Dittmar (2002)	−0.631 (−2.40)	−0.337
$LG^2_{JW,IV1}$	`25|VW|3m`	Dittmar (2002)	−0.614 (−2.89)	−0.387
$LG^2_{JW,IV1}$	`25|VW|1m`	Dittmar (2002)	−0.609 (−2.76)	−0.368
$LiqDemand_{shock}$	`25|VW|1m`	Goldberg and Nozawa (2021)	−1.641 (−2.22)	−0.755
$LiqDemand_{shock}$	`25|VW|6m`	Goldberg and Nozawa (2021)	−1.293 (−2.12)	−0.779
$LiqDemand_{shock}$	`25|EW|1m`	Goldberg and Nozawa (2021)	−1.172 (−2.52)	−0.801
PLS_SENT_{lag}	`25|VW|3m`	Huang et al. (2015)	0.562 (1.97)	0.253
$PLS\_SENT_{ORTH,lag}$	`Quint.|VW|1m`	Huang et al. (2015)	0.376 (2.14)	0.300
$PLS\_SENT_{ORTH,lag}$	`Quint.|VW|3m`	Huang et al. (2015)	0.356 (2.08)	0.297
CIV	`25|EW|1m`	Herskovic et al. (2016)	−0.686 (−3.24)	−0.423
CIV	`25|EW|1m`	Herskovic et al. (2016)	−0.644 (−2.95)	−0.402
CIV	`25|EW|3m`	Herskovic et al. (2016)	−0.591 (−2.93)	−0.412
HKM	`25|EW|1m`	He et al. (2017)	0.521 (2.38)	0.346
HKM	`25|EW|6m`	He et al. (2017)	0.518 (2.39)	0.371
HKM	`25|EW|6m`	He et al. (2017)	0.516 (2.41)	0.356
$Disag_{high}$	`25|VW|3m`	Huang et al. (2021)	−0.982	−0.376

Continued on next page

Table A.8 — *continued*

Factor	Strategy	Reference	Premium (%)	SR (annual)
			(−2.32)	
$Disag_{high}$	`Dec.|EW|3m`	Huang et al. (2021)	−0.934 (−2.76)	−0.439
$Disag_{high}$	`25|EW|3m`	Huang et al. (2021)	−0.874 (−2.32)	−0.376
LIQ_{noise}	`25|VW|6m`	Hu et al. (2013)	−0.701 (−2.10)	−0.352
LIQ_{noise}	`Dec.|EW|1m`	Hu et al. (2013)	−0.610 (−1.99)	−0.392
LG_{JW}	`25|VW|1m`	Jagannathan and Wang (1996)	−0.971 (−3.52)	−0.445
LG_{JW}	`25|VW|1m`	Jagannathan and Wang (1996)	−0.934 (−3.10)	−0.447
LG_{JW}	`25|VW|1m`	Jagannathan and Wang (1996)	−0.810 (−2.77)	−0.383
$LaborTightness$	`25|VW|3m`	Kuehn et al. (2017)	−0.486 (−2.12)	−0.267
$LaborTightness$	`Dec.|VW|1m`	Kuehn et al. (2017)	−0.342 (−2.01)	−0.251
ΔV_{NI}	`25|VW|1m`	Liu and Matthies (2022)	0.635 (2.39)	0.350
ΔV_{HNI}	`Dec.|VW|3m`	Liu and Matthies (2022)	0.457 (2.28)	0.344
ΔV_{HNI}	`Dec.|VW|1m`	Liu and Matthies (2022)	0.422 (2.02)	0.307
$NEWS^2_{VIX}$	`25|VW|1m`	Manela and Moreira (2017)	−0.754 (−2.22)	−0.330
$NEWS^2_{VIX}$	`25|VW|1m`	Manela and Moreira (2017)	−0.657 (−2.09)	−0.318
$NEWS^2_{VIX}$	`25|VW|3m`	Manela and Moreira (2017)	−0.612 (−2.03)	−0.290
$ILLIQ_{trans}$	`Dec.|VW|6m`	Sadka (2006)	0.587 (1.99)	0.432
$ILLIQ_{trans}$	`Dec.|VW|6m`	Sadka (2006)	0.486 (2.08)	0.398
$ILLIQ_{trans}$	`Quint.|VW|6m`	Sadka (2006)	0.449 (2.21)	0.437
DSV	`25|EW|1m`	Vassalou and Xing (2004)	0.823 (2.08)	0.393
DSV	`Dec.|EW|1m`	Vassalou and Xing (2004)	0.671 (2.01)	0.394
DSV	`Quint.|EW|1m`	Vassalou and Xing (2004)	0.516 (1.96)	0.397

Figures

Risk Premia Inference

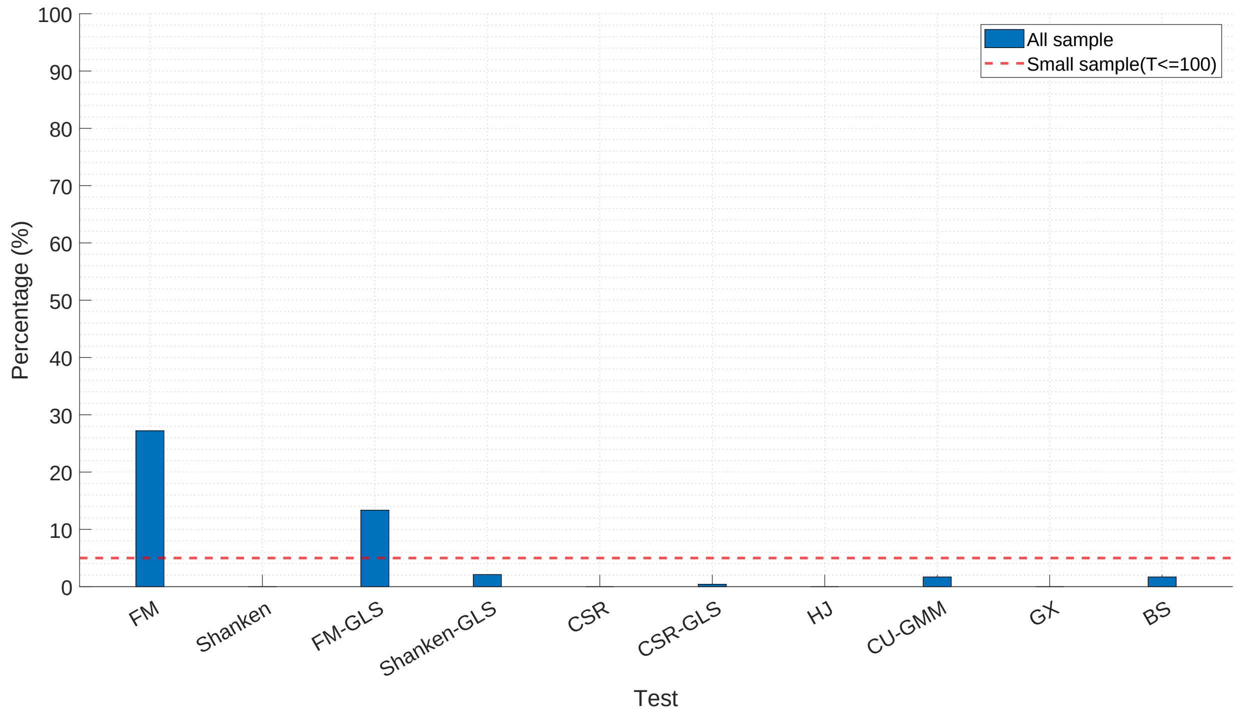


Figure A.1: **The percentages of priced factors at the 5% significance level controlling for multiple hypothesis testing** The figure shows the percentages of priced factors, pooling across all model specifications and test assets, and controlling the false discovery rate at a 5% in each test with Benjamini and Hochberg (1995) technique. The percentages are based on the rejection of the null hypothesis that the risk premium is equal to zero. The test statistics are calculated using the standard errors from Fama-MacBeth approach (FM and FM GLS), Jagannathan and Wang (1998) generalized Shanken's errors-in-variables allowing for conditional heteroskedasticity (Shanken and Shanken-GLS), as well as misspecification-robust standard errors in Kan et al. (2013) (CSR and CSR-GLS), in Gospodinov et al. (2014) based on Hansen-Jagannathan distance (HJ) and in Gospodinov et al. (2017) in linear SDF estimated using continuously updated GMM (CU-GMM). We also report the percentages of rejection with Giglio and Xiu (2021) three-pass approach and Burnside (2011) bootstrap approach.

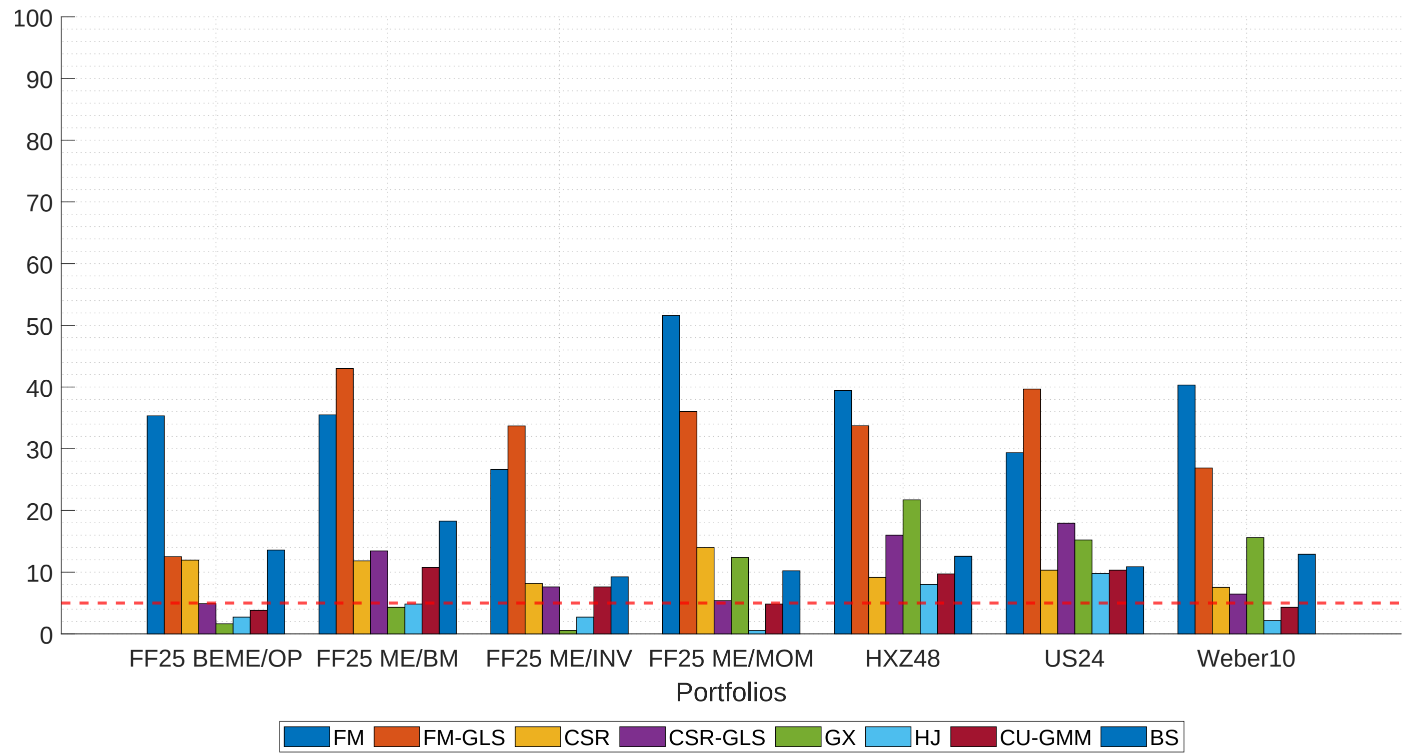


Figure A.2: **The percentages of priced factors at a 5% significance level with equity portfolios**
For each set of equity portfolios, the figure reports the percentages of factors with nonzero risk premia at the 5% level, suggested by each test. The test statistics are calculated using the standard errors from Fama-MacBeth approach (FM and FM-GLS), Jagannathan and Wang (1998) generalized Shanken's errors-in-variables allowing for conditional heteroskedasticity (Shanken and Shanken-GLS), as well as misspecification-robust standard errors in Kan et al. (2013) (CSR and CSR-GLS), in Gospodinov et al. (2014) based on Hansen-Jagannathan distance (HJ) and in Gospodinov et al. (2017) in linear SDF estimated using continuously updated GMM (CU-GMM). We also report the percentages of rejection with Giglio and Xiu (2021) three-pass approach and Burnside (2011) bootstrap approach.

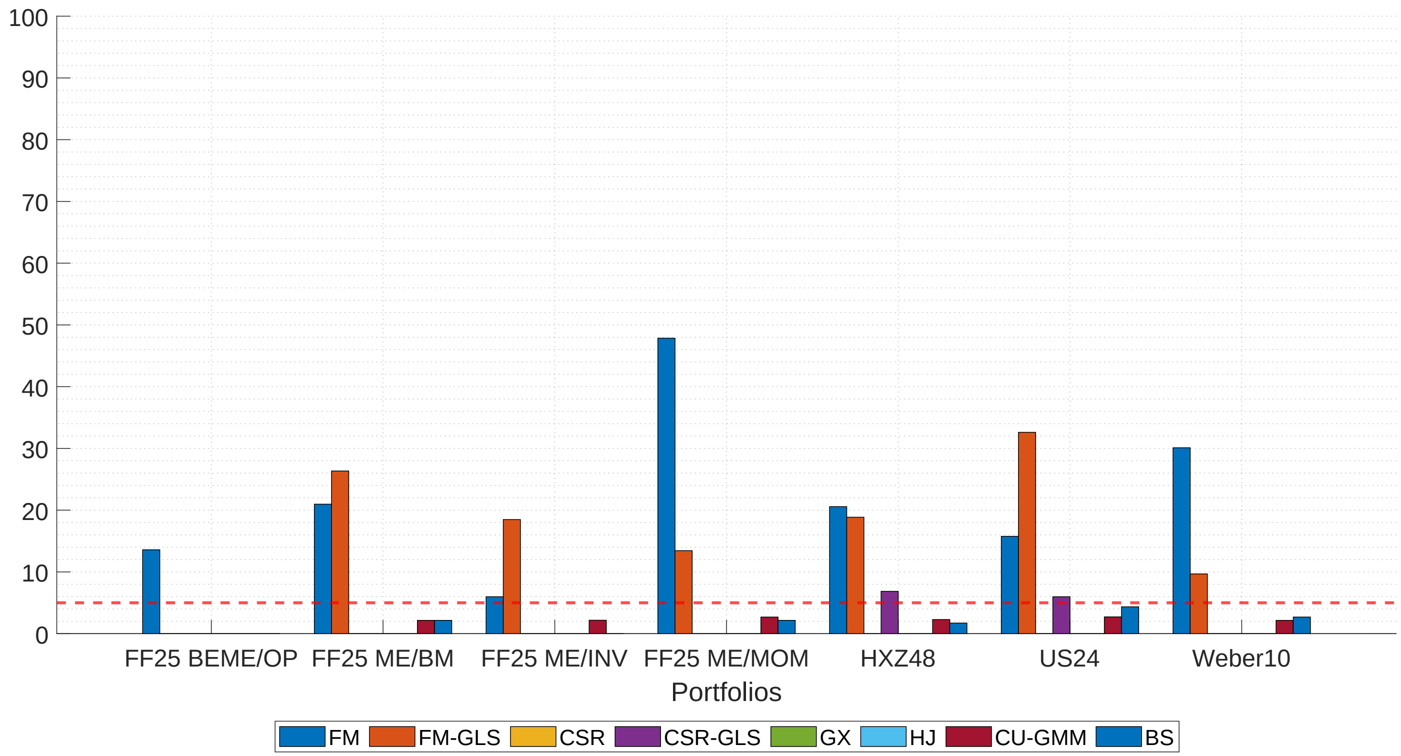


Figure A.3: **The percentages of priced factors at a 5% significance level with equity portfolios controlling for multiple hypothesis testing**

For each set of equity portfolios, the figure reports the percentages of factors with nonzero risk premia at the 5% level, controlling the false discovery rate at 5% in each type of test portfolios with Benjamini and Hochberg (1995) technique. The test statistics are calculated using the standard errors from Fama-MacBeth approach (FM and FM GLS), Jagannathan and Wang (1998) generalized Shanken's errors-in-variables allowing for conditional heteroskedasticity (Shanken and Shanken-GLS), as well as misspecification-robust standard errors in Kan et al. (2013) (CSR and CSR-GLS), in Gospodinov et al. (2014) based on Hansen-Jagannathan distance (HJ) and in Gospodinov et al. (2017) in linear SDF estimated using continuously updated GMM (CU-GMM). We also report the percentages of rejection with Giglio and Xiu (2021) three-pass approach and Burnside (2011) bootstrap approach.

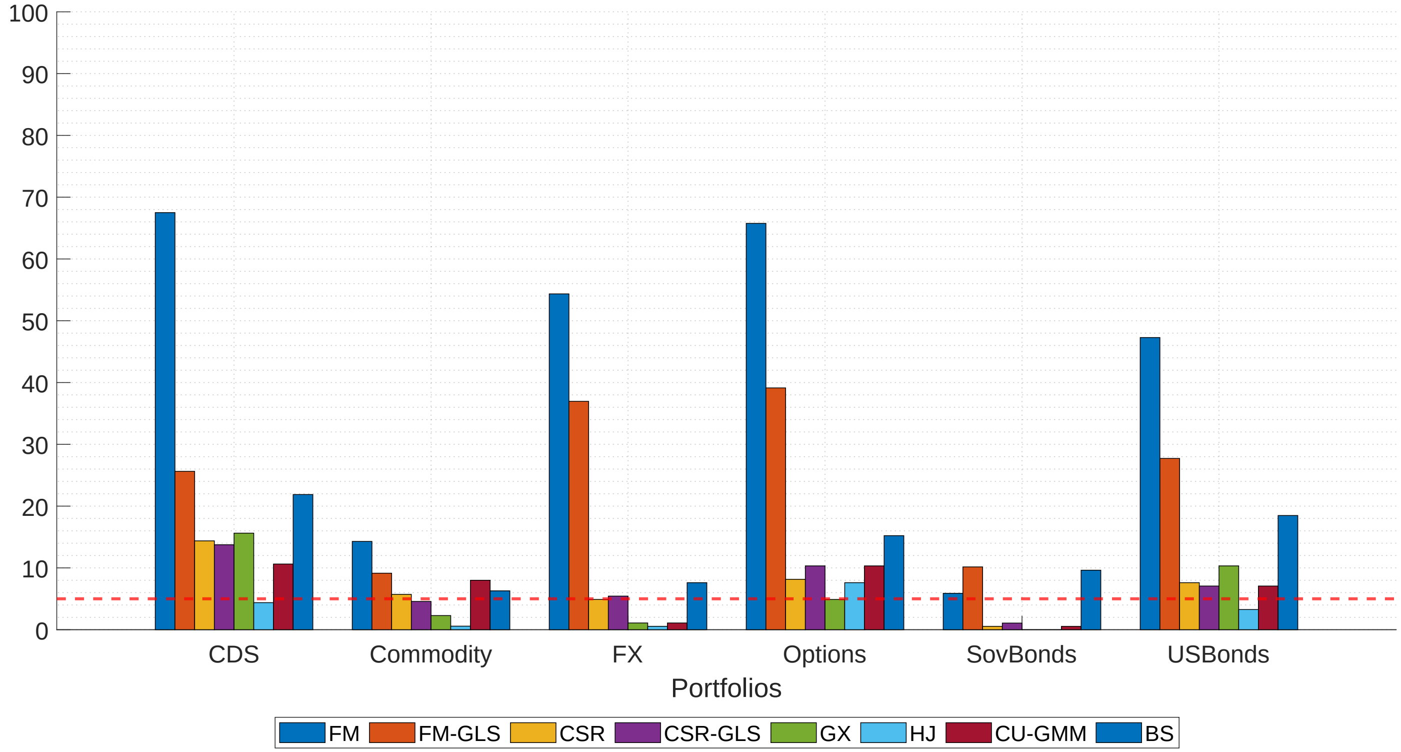


Figure A.4: **The percentages of priced factors at a 5% significance level with non-equity portfolios**
The figure shows the percentages of priced factors for each set of non-equity portfolios at a 5% significance level. The percentages are based on the rejection of the null hypothesis that the risk premium is equal to zero. The test statistics are calculated using the standard errors from Fama-MacBeth approach (FM and FM GLS), Jagannathan and Wang (1998) generalized Shanken's errors-in-variables allowing for conditional heteroskedasticity (Shanken and Shanken-GLS), as well as misspecification-robust standard errors in Kan et al. (2013) (CSR and CSR-GLS), in Gospodinov et al. (2014) based on Hansen-Jagannathan distance (HJ) and in Gospodinov et al. (2017) in linear SDF estimated using continuously updated GMM (CU-GMM). We also report the percentages of rejection with Giglio and Xiu (2021) three-pass approach and Burnside (2011) bootstrap approach.

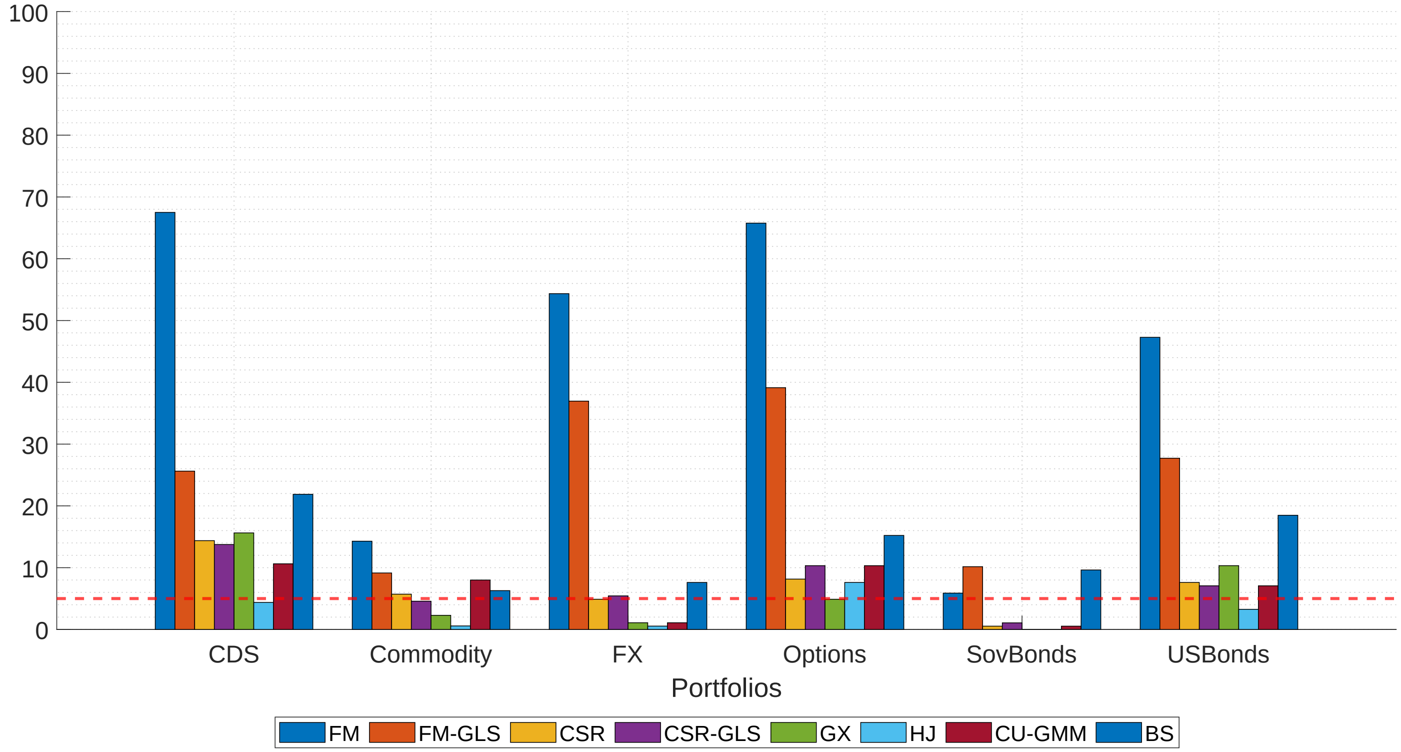


Figure A.5: **The percentages of priced factors at a 5% significance level with non-equity portfolios controlling for multiple hypothesis testing**

The figure shows the percentages of priced factors for each set of non-equity portfolios at a 5% significance level, controlling the false discovery rate at 5% in each type of test portfolios with Benjamini and Hochberg (1995) technique. The percentages are based on the rejection of the null hypothesis that the risk premium is equal to zero. The test statistics are calculated using the standard errors from Fama-MacBeth approach (FM and FM-GLS), Jagannathan and Wang (1998) generalized Shanken's errors-in-variables allowing for conditional heteroskedasticity (Shanken and Shanken-GLS), as well as misspecification-robust standard errors in Kan et al. (2013) (CSR and CSR-GLS), in Gospodinov et al. (2014) based on Hansen-Jagannathan distance (HJ) and in Gospodinov et al. (2017) in linear SDF estimated using continuously updated GMM (CU-GMM). We also report the percentages of rejection with Giglio and Xiu (2021) three-pass approach and Burnside (2011) bootstrap approach.

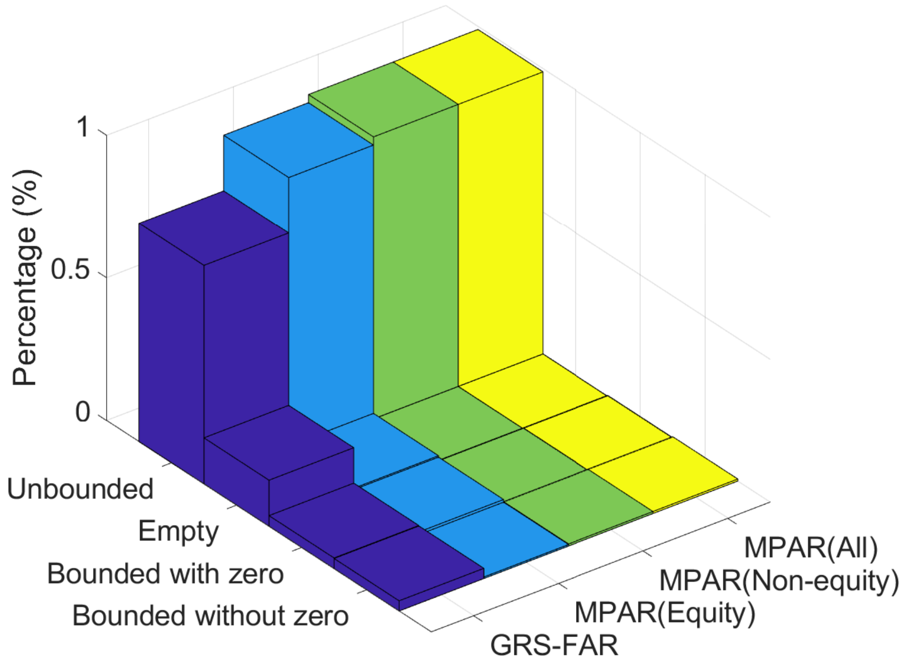


Figure A.6: **Confidence intervals of risk premia**

We employ the techniques proposed by Kleibergen and Zhan (2020) to construct confidence intervals by inverting the GRS-FAR test and using the mimicking portfolio approach (MPAR), which is robust to weak identification as described in Kleibergen and Zhan (2018). For the MPAR, we use three sets of base assets: all equity portfolios (Equity), non-equity portfolios (Non-Equity), and all assets (All) included in this paper as detailed in Section 3. There are four types of confidence intervals: unbounded, empty, bounded including zero, and bounded excluding zero.

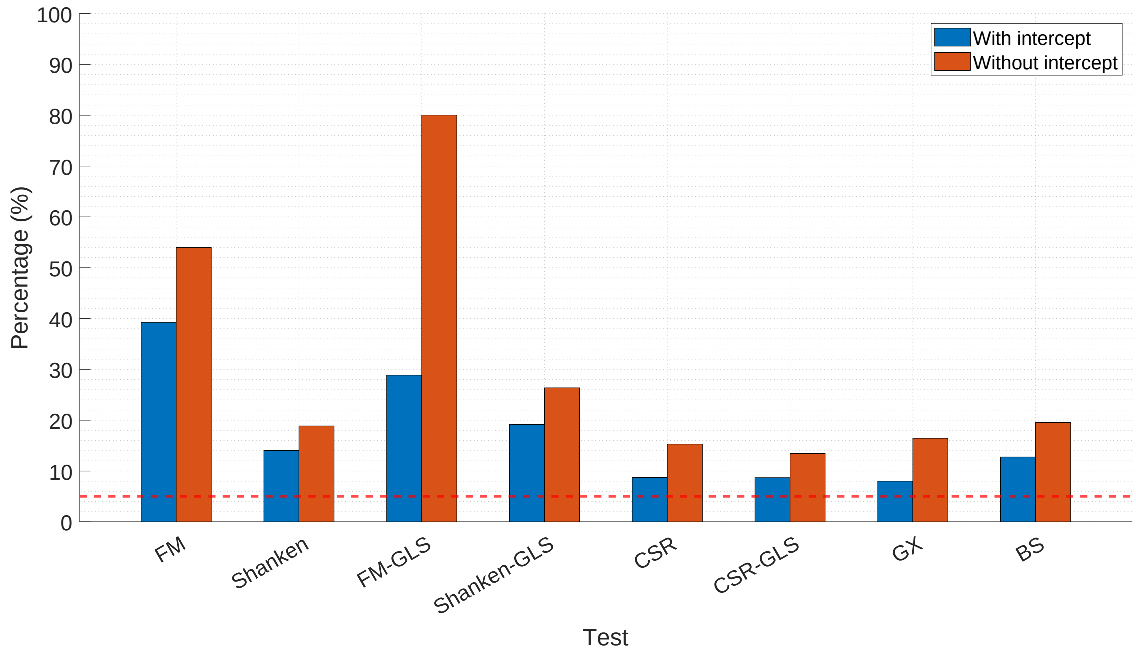


Figure A.7: **The percentages of priced factors with and without intercept**
This figure compares the percentages of priced factors at the significance level of 5% including and excluding the cross-sectional intercepts. The percentages are based on the rejection of the null hypothesis that the risk premium is equal to zero. The test statistics are calculated using the standard errors from Fama-MacBeth approach (FM and FM-GLS), Jagannathan and Wang (1998) generalized Shanken's errors-in-variables allowing for conditional heteroskedasticity (Shanken and Shanken-GLS), as well as misspecification-robust standard errors in Kan et al. (2013) (CSR and CSR-GLS). We also report the percentages of rejection with Giglio and Xiu (2021) three-pass approach and Burnside (2011) bootstrap approach.

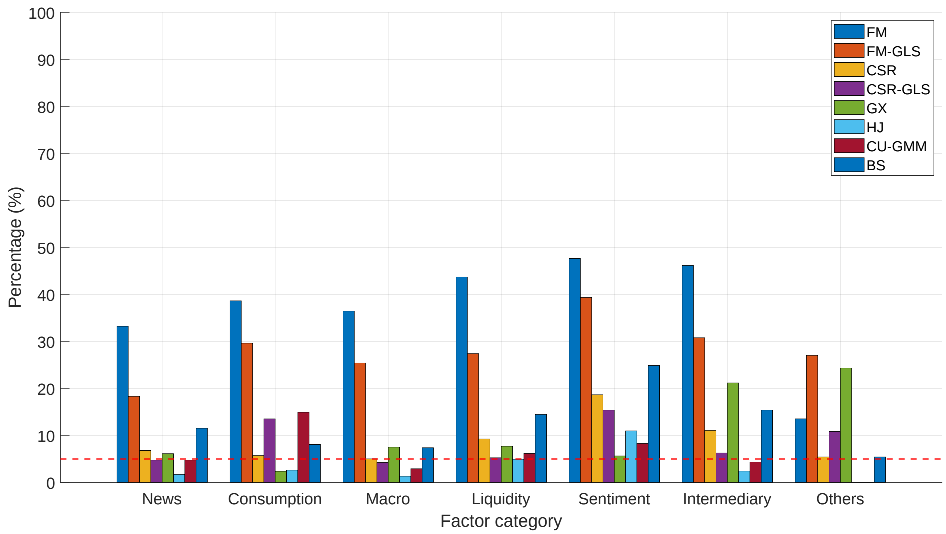


Figure A.8: **The percentages of priced factors by factor category**
This figure shows the percentages of priced factors for each factor category at 5% significance level. The test statistics are calculated using the standard errors from Fama-MacBeth approach (FM and FM-GLS), Jagannathan and Wang (1998) generalized Shanken's errors-in-variables allowing for conditional heteroskedasticity (Shanken and Shanken-GLS), as well as misspecification-robust standard errors in Kan et al. (2013) (CSR and CSR-GLS). We also report the percentages of rejection with Giglio and Xiu (2021) three-pass approach and Burnside (2011) bootstrap approach.

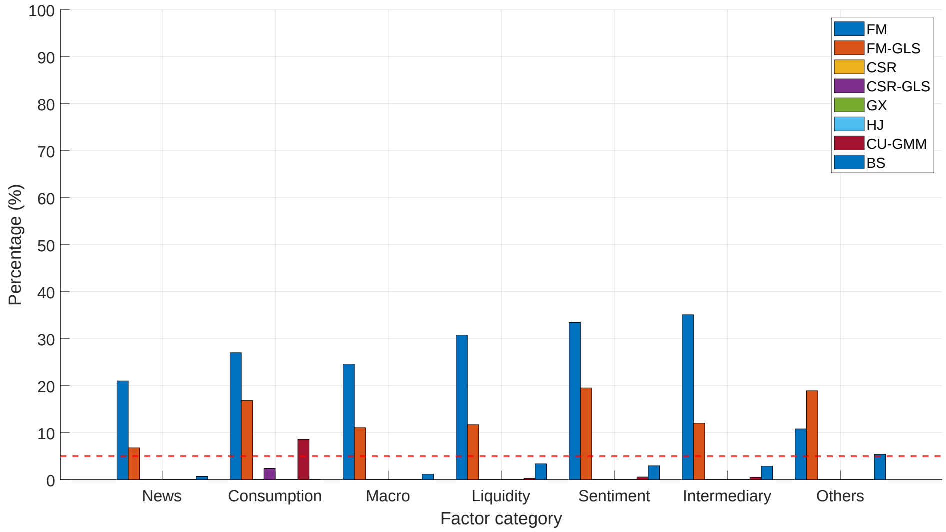


Figure A.9: **The percentages of priced factors by factor category controlling for multiple hypothesis testing**

This figure shows the percentages of priced factors for each factor category at 5% significance level controlling false discovery rate at 5% level. The test statistics are calculated using the standard errors from Fama-MacBeth approach (FM and FM-GLS), Jagannathan and Wang (1998) generalized Shanken's errors-in-variables allowing for conditional heteroskedasticity (Shanken and Shanken-GLS), as well as misspecification-robust standard errors in Kan et al. (2013) (CSR and CSR-GLS). We also report the percentages of rejection with Giglio and Xiu (2021) three-pass approach and Burnside (2011) bootstrap approach.

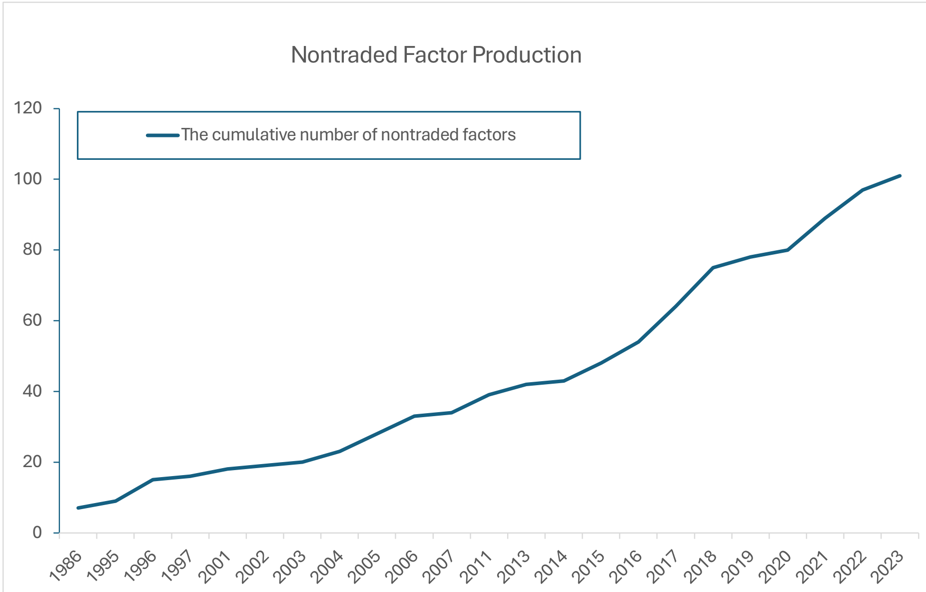


Figure A.10: **The cumulative number of nontraded factors proposed in the literature from 1986 to February 2023**

Identification

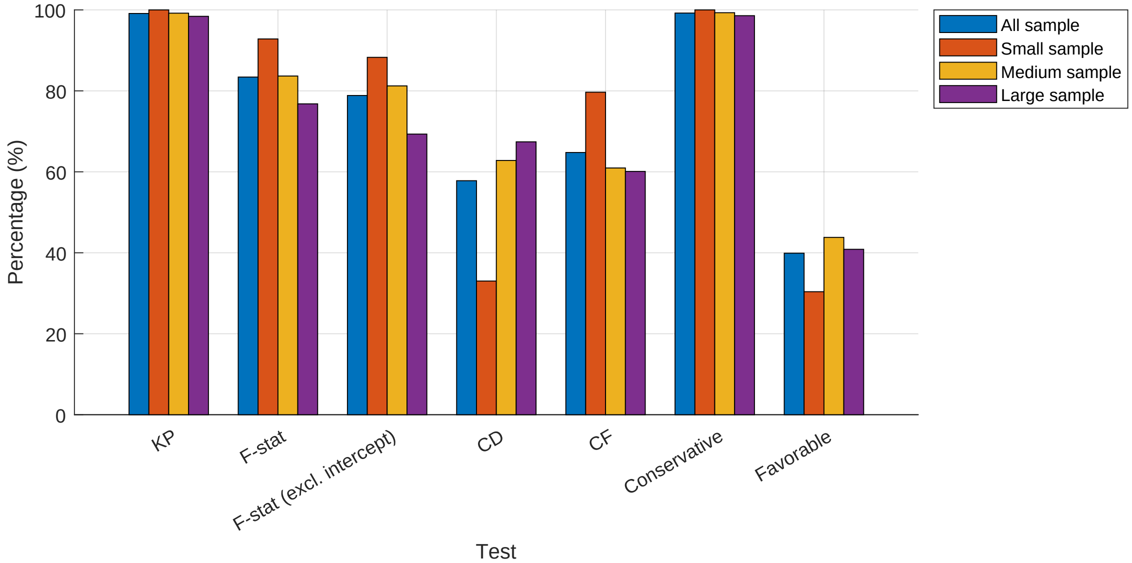


Figure A.11: **Identification failure of original models corrected for multiple hypothesis testing**

This figure shows The percentages 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.The rank tests used are Kleibergen and Paap (2006) rank test (KP), finite-sample rank test(F-stat) used in Kleibergen and Zhan (2020) as well as the test excluding intercept (F-stat (excl. intercept)), Cragg-Donald rank test (CD) used in Gospodinov et al. (2017), and Chen and Fang (2019) rank test (CF). We also consider the most conservative situation where we can conclude a model is identified only when all of four test statistics can reject the null of rank deficiency, and the favorable situation where any of the test statistics can reject. We also present the percentages across length of time series observations $T \leq 100$ (small), $100 < T \leq 400$ (medium) and $T > 400$ (large).

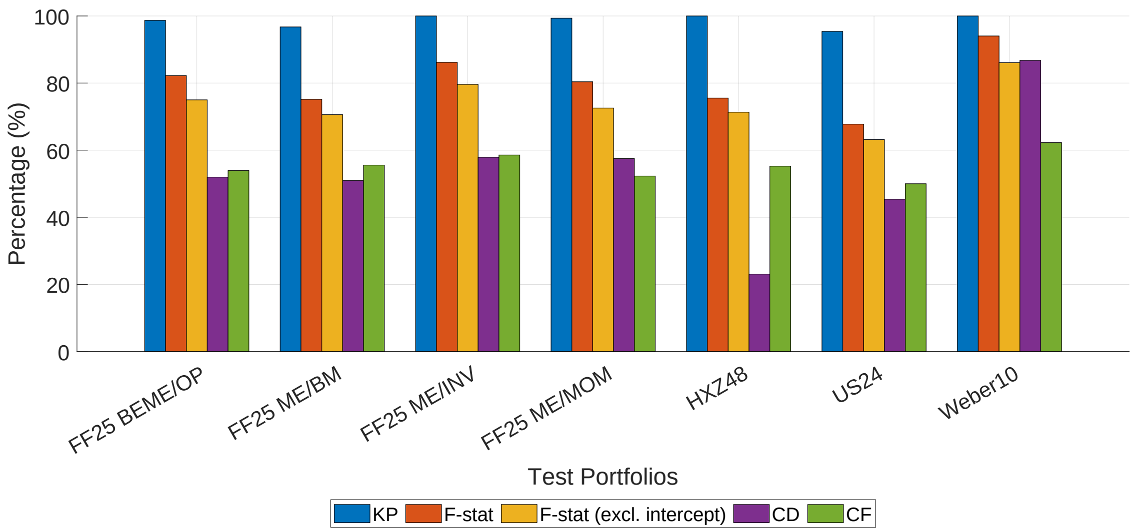


Figure A.12: **Identification failure of model specifications with equity portfolios corrected for multiple hypothesis testing**

This figure presents The percentages 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 (KP), Kleibergen and Zhan (2020) finite-sample rank test(F-stat) and the test excluding intercept (F-stat (excl. intercept)), Gospodinov et al. (2017) Cragg-Donald rank test (CD), as well as Chen and Fang (2019) rank test (CF).

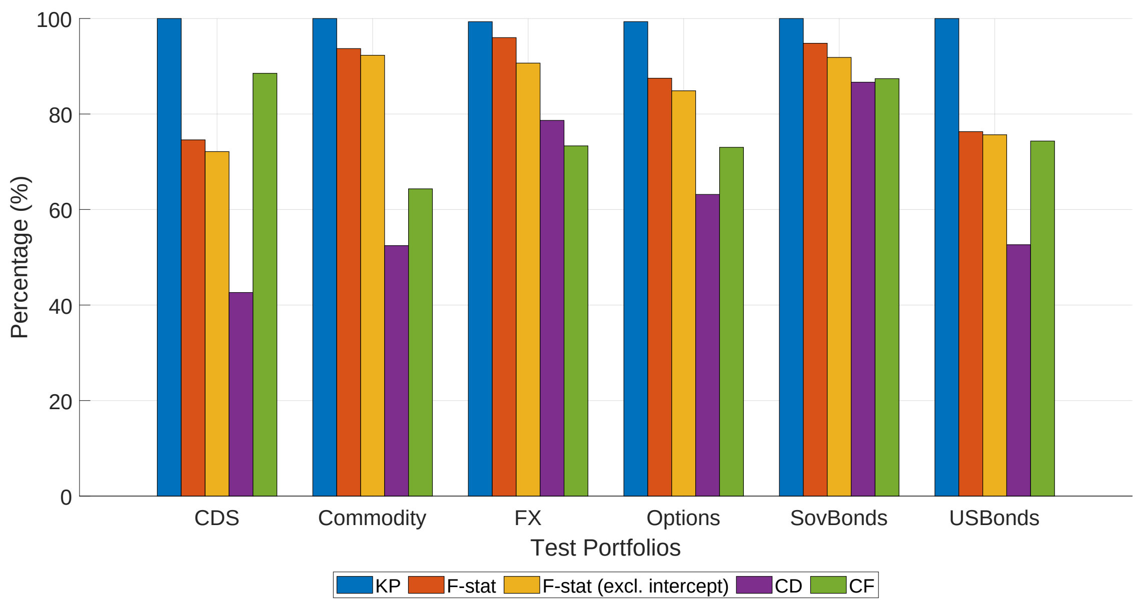


Figure A.13: **Identification failure of model specifications with non-equity portfolios corrected for multiple hypothesis testing**

This figure presents The percentages 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 (KP), Kleibergen and Zhan (2020) finite-sample rank test(F-stat) and the test excluding intercept (F-stat (excl. intercept)), Gospodinov et al. (2017) Cragg-Donald rank test (CD), as well as Chen and Fang (2019) rank test (CF).

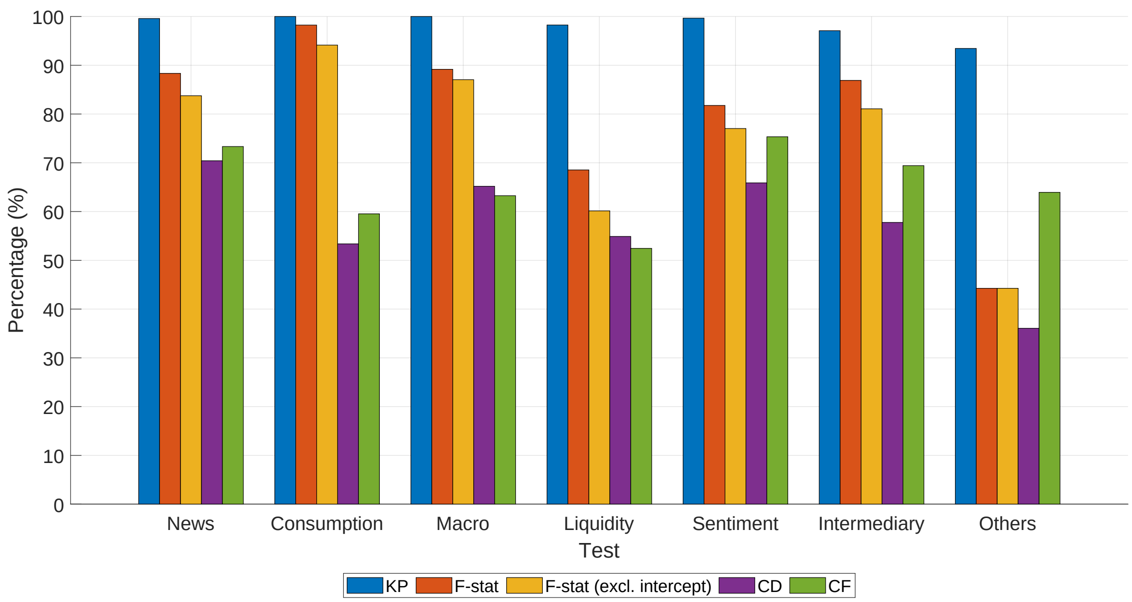


Figure A.14: **The percentages of identification failure for different categories of factors controlling for multiple hypothesis testing**

We control false discovery rate at 5% using Benjamini and Hochberg (1995) method. We use four rank tests: Kleibergen and Paap (2006) rank test (KP), Kleibergen and Zhan (2020) finite-sample rank test(F-stat) and the test excluding intercept (F-stat (excl. intercept)), Gospodinov et al. (2017) Cragg-Donald rank test (CD), as well as Chen and Fang (2019) rank test (CF).

Misspecification

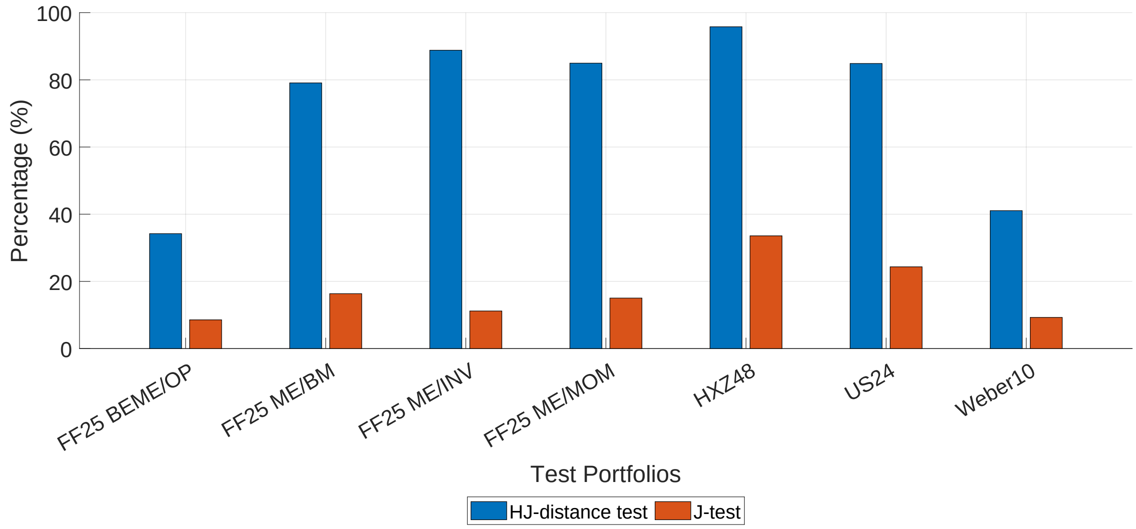


Figure A.15: **The percentages of model misspecification with equity portfolios**
This figure reports the percentages of model misspecification for equity portfolios in over-identifying restriction J-test and Hansen-Jagannathan (HJ) distance test. The percentages represent the proportion of models that reject the null hypothesis of correct model specification.

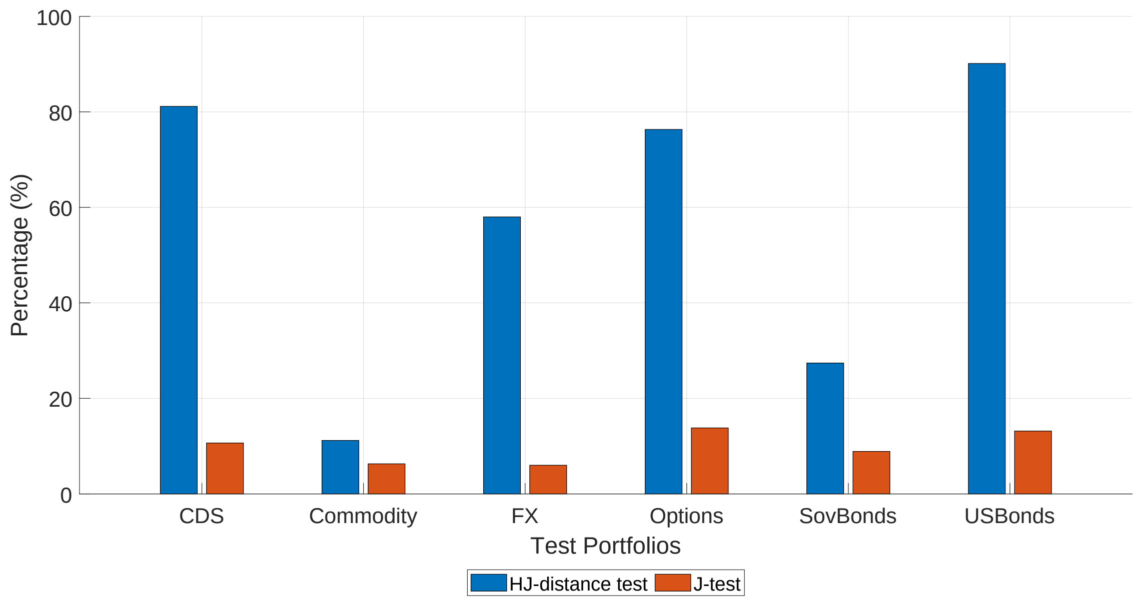


Figure A.16: **The percentages of model misspecification with non-equity portfolios**
This figure reports the percentages of model misspecification for non-equity portfolios in over-identifying restriction J-test and Hansen-Jagannathan (HJ) distance test. The percentages represent the proportion of models that reject the null hypothesis of correct model specification.

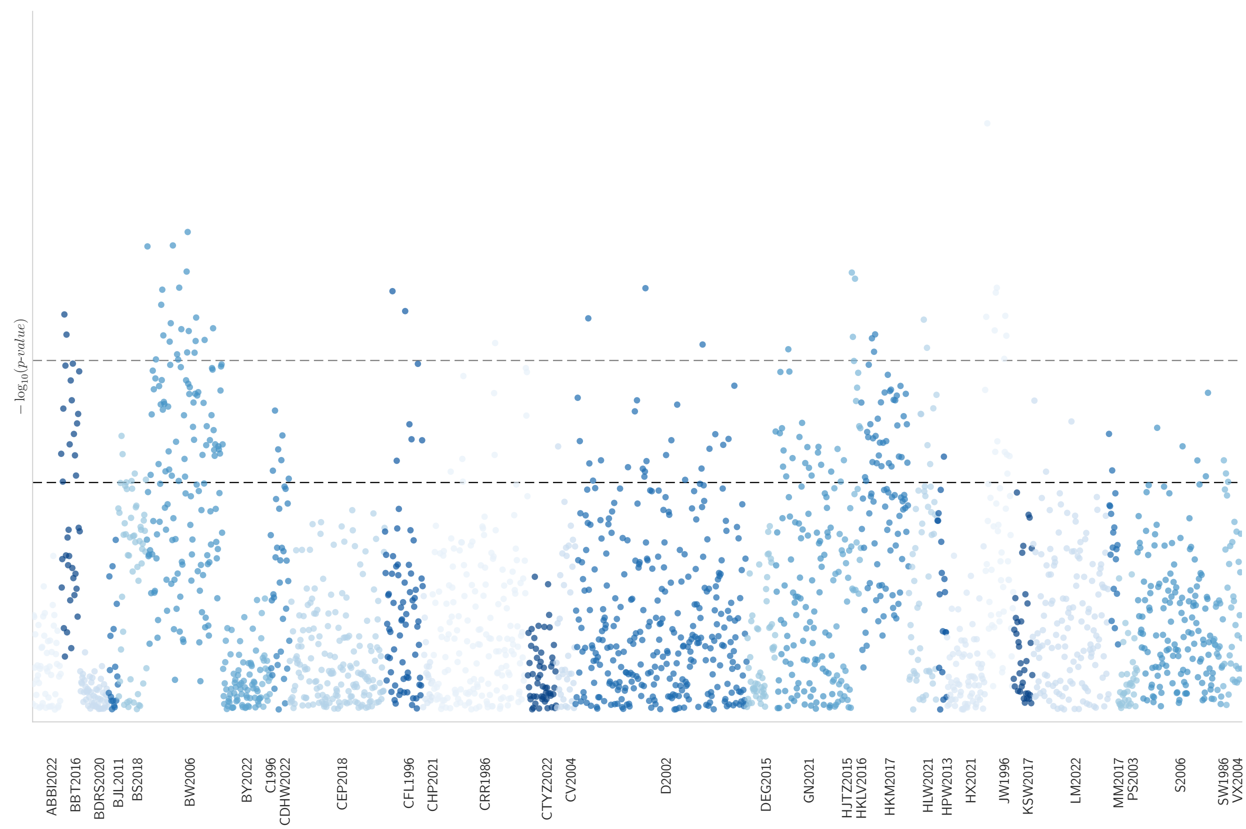


Figure A.17: **Significance of portfolio premiums from β-sorted portfolios across papers - Original specifications.** This figure plots the $-\log_{10}(p\text{-value})$ from tests of whether average long-short portfolio expected returns differ from zero for single-factor models. Higher values indicate stronger statistical significance. The labels follow the factor definitions presented in Tables A.1–A.5.

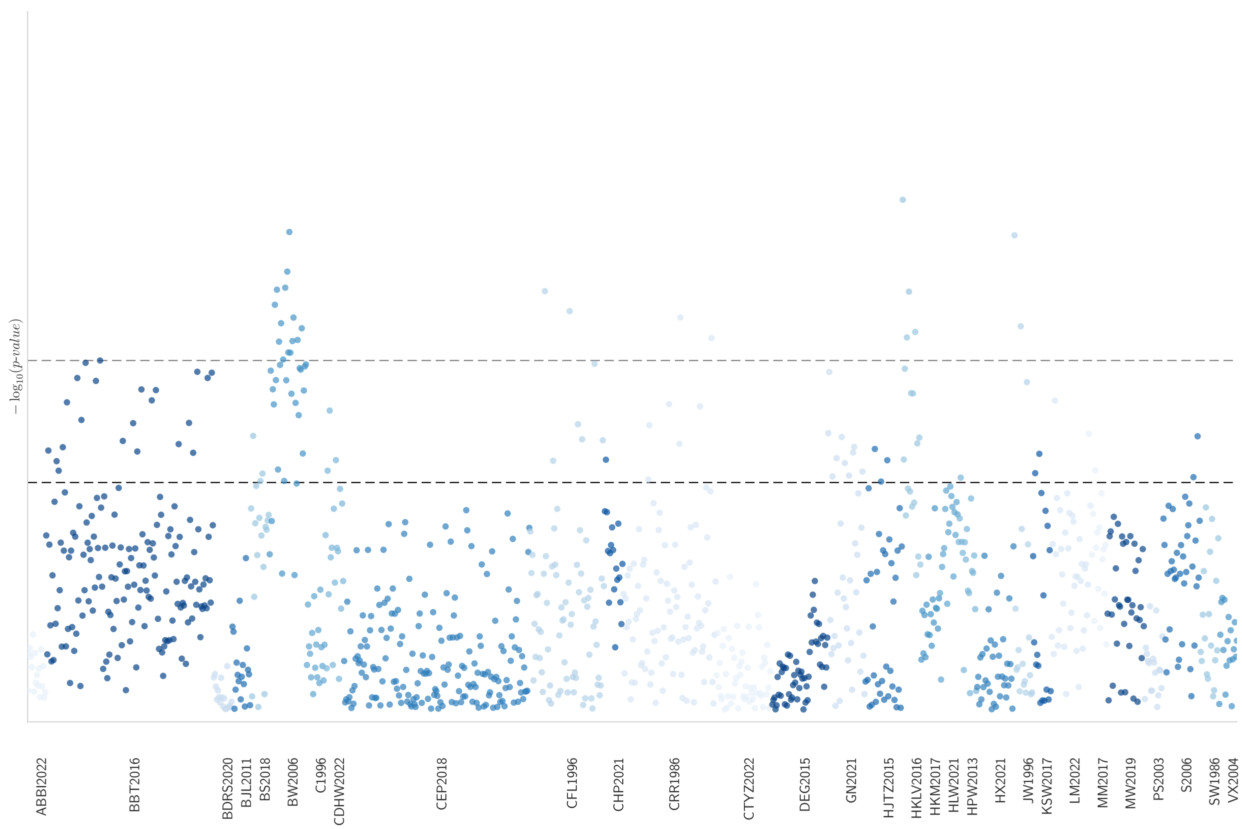


Figure A.18: **Significance of portfolio premiums from β-sorted portfolios across papers - Single factors specifications.** This figure plots the $-\log_{10}(p\text{-value})$ from tests of whether average long-short portfolio expected returns differ from zero for single-factor models. Higher values indicate stronger statistical significance. The labels follow the factor definitions presented in Tables A.1–A.5.

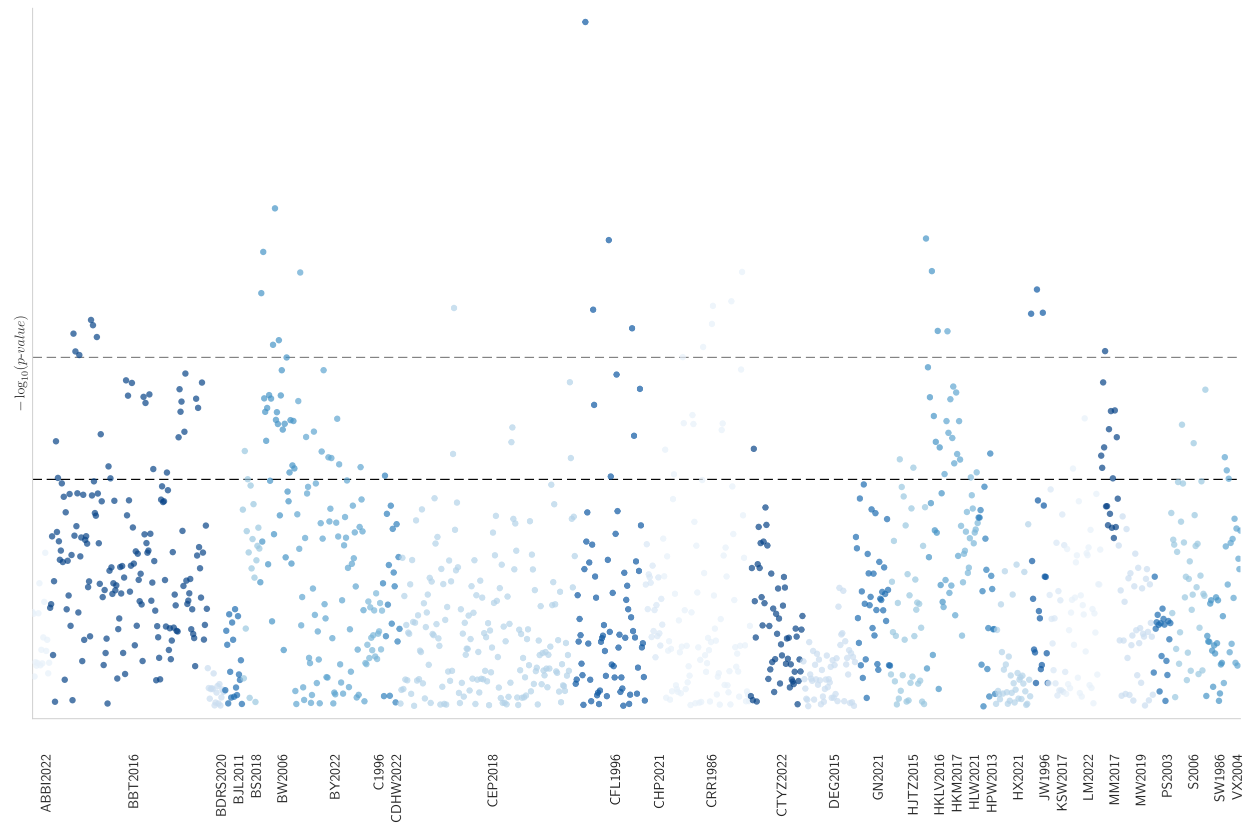


Figure A.19: **Significance of portfolio premiums from β-sorted portfolios across papers - Single factors + MKT specifications.** This figure plots the $-\log_{10}$(p-value) from tests of whether average long-short portfolio expected returns differ from zero for single-factor models augmented with the market factor. Higher values indicate stronger statistical significance. The labels follow the factor definitions presented in Tables A.1–A.5.