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, Université PSL, CNRS, DRM, Finance, 75016 Paris, France*
[b,c,d] *Warwick Business School, Coventry, CV4 7AL, United Kingdom*

October 2026

Abstract

We conduct a broad investigation of whether nontraded factors price the cross-section of returns, accounting for weak identification and global model misspecification. We collect 102 nontraded factors from 51 papers in mainstream finance and economics journals and replicate 145 published specifications using widely used test portfolios. Because these factors are often weakly correlated with asset returns and their models are likely misspecified, conventional risk-premium tests may overstate their pricing ability. Standard econometric methods classify up to 39% as priced, whereas robust procedures reduce that share to 4%–13%; multiple-testing adjustments weaken the evidence, and adding the market factor worsens identification. Complementary beta-sorted portfolio analysis provides similar results: only 11.2% of long–short strategies earn statistically significant average returns, with small premia. Our results do not necessarily imply that nontraded factors are unpriced, but most published proxies receive little support from robust pricing tests or portfolio sorts immune to weak identification.

Keywords: Factor zoo; Nontraded factors; Global misspecification; Weak identification; Risk premia; Portfolio sorting.

JEL classification: C58, G12.

*We thank seminar participants at the 2024 SoFiE conference for helpful comments. We are also indebted to Bertille Antoine, Elise Gourier, Arie Gozluklu, Christian Julliard, Bryan Kelly, Ilaria Piatti, and Ao Wang for their insightful suggestions.

†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

Most published nontraded factors do not retain their pricing ability once weak identification and global (fixed) model misspecification are taken into account. We collect 102 nontraded factors from 51 papers in top academic journals, replicate 145 published model specifications, and compare standard pricing tests with procedures that remain asymptotically valid when factor exposures are weak and/or the model is misspecified. Standard methods classify as many as 39% of the proposed nontraded factors as priced, but that share falls to about 4%–13% under robust methods. The pricing evidence weakens further when multiple-testing adjustments are applied and when the market factor enters as a control.

While the recent debate in the literature has focused primarily on traded factors, the universe of nontraded factors – those not constructed as portfolio returns or return spreads – has received comparatively little attention. These factors are potentially attractive because they provide the main empirical link between many asset-pricing theories and observed financial asset returns. Macroeconomic shocks, intermediary proxies, sentiment indexes, and news measures are usually nontradable. Their appeal is clear: they offer a direct economic interpretation that traded return spreads often do not. That same feature makes them harder to evaluate. Since these variables are not returns on traded portfolios, their factor exposures are often only weakly related to the returns they are meant to price. A nontraded factor can therefore look economically appealing while remaining statistically difficult to evaluate in the cross-section. Even so, their proliferation has continued unabated over the past four decades, as reflected in the steady rise in newly proposed nontraded factors (see Fig. B.9).

Weak identification is the first obstacle. In a two-pass asset-pricing framework, the risk premium is obtained from cross-sectional differences in estimated betas. When those betas are close to zero and barely vary across test assets, or when those betas are close to one and the zero-beta rate is included as a free parameter in the second-pass cross-sectional regression, the true premium is undefined and its estimate diverges as the time-series sample size goes to infinity. Conventional t-statistics, constructed under the assumption of correct model specification, can then be too high and wrongly suggest that the underlying factors are priced (Gospodinov et al., 2014a, 2019). We show that roughly 34% of published specifications fail favorable identification diagnostics, and the share of weakly identified models rises by 3 to

14 percentage points depending on the test once the market factor is included in the model.

Global misspecification creates a second obstacle. Conventional two-pass inference is typically performed under the assumption of correct model specification, that is, under the restriction that the cross-sectional pricing errors are zero (Kan et al., 2013). When that restriction fails, the standard errors of the premium estimates can understate uncertainty and conventional significance tests can substantially inflate the evidence of factor pricing. Procedures that are robust to global model misspecification address this problem directly, and in our sample they remove most of the pricing evidence from standard two-pass tests. Robust distance-based procedures remain asymptotically valid under global misspecification and in the presence of weak (spurious) factors (Gospodinov et al., 2014a), at least in the generalized least squares (GLS) case. Therefore, the main message of our work is that the pricing evidence for most nontraded factors is much weaker than the existing literature suggests.

Following Harvey et al. (2016), we collect 102 distinct nontraded factors from 51 papers published between July 1986 and February 2023.[1] By adopting a taxonomic approach similar to Jensen et al. (2023), we classify these factors into six themes: *News*, *Sentiment*, *Consumption*, *Macroeconomics*, *Liquidity*, and *Intermediary & Aggregate Firm-level Risk*. We then scientifically replicate 145 published model specifications and evaluate them on 13 sets of test portfolios that span equity and non-equity markets.

We evaluate the nontraded factor literature at three distinct levels. We first analyze the credibility of the published model specifications by asking whether their pricing claims survive once weak identification and global misspecification are properly accounted for. We then study the economic relevance of the proposed factors by asking whether exposures to them generate meaningful return spreads in implementable beta-sorted portfolios. Finally, we provide an interpretation for the widespread failure of individual factors. The key question is whether the various published nontraded factors individually price returns, and what their pricing failure implies about the way asset pricing tests accommodate nontraded risks.

We find that the published nontraded factor specifications share the same statistical pat-

[1]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, and G. Zhou for publicly sharing their data.

tern: weak identification is widespread and global model misspecification occurs frequently. Conservative identification-robust procedures deliver an even more skeptical assessment of these nontraded factor models and flag more than 94% of published models as weakly identified. The model misspecification evidence points in the same direction. Conventional specification tests reject only a small fraction of models, whereas more robust distance-based procedures reject many more. These results show that the assumptions underlying standard risk-premium inference often fail in the published nontraded factor zoo.

The pricing evidence we provide is not entirely negative, but it is more limited than the published literature suggests. A small set of published nontraded factors continues to price selected cross-sections based on our identification- and misspecification-robust schemes. More importantly, the pattern of our results points to a broader interpretation. The main message is not that nontraded risk is unrelated to asset prices. It is that most published empirical proxies for nontraded risks are weak, noisy, and difficult to validate with the standard factor-by-factor testing paradigm.

This paper makes three main contributions to the existing literature. First, it brings the work on weak identification, global model misspecification, and robust risk-premium inference to the nontraded factor zoo in a systematic and thorough way, showing how sharply the conclusions change once the relevant statistical issues are properly taken into account. Second, it connects our statistical reassessment to economic relevance by asking whether the same factors generate meaningful return spreads in beta-sorted portfolios. Third, it uses the combined evidence to clarify how empirical researchers should interpret published nontraded factor claims. We therefore reassess a large set of factors/models and provide a framework for evaluating asset pricing findings built around nontraded risks.

Two strands of literature shape our contribution. The first relates to weak identification, global model misspecification, and robust inference in linear asset-pricing models. Kleibergen (2009), Kan et al. (2013), Gospodinov et al. (2014a), Gospodinov et al. (2019), Gospodinov and Robotti (2021), Kleibergen and Zhan (2020), and Bryzgalova et al. (2023b) show that conventional inference can be unreliable when factors are weak and/or evaluated under global model misspecification. These concerns are especially relevant for nontraded factors that are often only weakly related to returns. We build on this literature by applying these insights systematically to a large set of published nontraded factors and original model specifications

as opposed to just a handful of them.

A second strand of literature evaluates factor proliferation through selection, replication, and model comparison. Following Cochrane (2011), this literature investigates whether the growing factor zoo reflects genuine priced risks or fragile empirical findings (Harvey et al., 2016; McLean and Pontiff, 2016; Linnainmaa and Roberts, 2018; Hou et al., 2020; Jensen et al., 2023; Giglio et al., 2021). Feng et al. (2020) use LASSO-based methods to assess whether candidate factors add incremental explanatory power relative to existing factors. Related Bayesian methods formally compare combinations of candidate factors. For example, Barillas and Shanken (2018) provide a framework for comparing asset-pricing models based on their maximum squared Sharpe ratios, and Bryzgalova et al. (2023a) search over 2.25 quadrillion models formed from 51 traded and nontraded factors.

Our exercise should be viewed as complementary. Selection and model-comparison approaches are concerned with which factors or factor combinations perform best within a candidate set. We instead take published specifications as the objects of interest and ask whether the evidence reported for those specifications survives identification and misspecification diagnostics. This design reveals a different margin of fragility: how much of the published nontraded-factor evidence depends on conventional inference in the original model, rather than on the choice of factor combination within a large model space.

Section 2 describes the diagnostics and pricing tests. Section 3 describes the factors, test portfolios, and original model specifications. Section 4 presents the main results for identification, misspecification, and pricing. Section 5 examines economic relevance with beta-sorted portfolios. Section 6 interprets what the widespread failure of published nontraded factors means for empirical asset pricing. Section 7 concludes. For the sake of a concise discussion, details on assumptions, methods, data, and supplementary results are in the Appendix and the Internet Appendix.

2 Diagnostic Testing and Robust Inference Procedures

A 'conventional' pricing test is informative only if two conditions are satisfied in the data. The first relates to model identification: factor exposures must vary enough across test assets

to enable us to identify the risk premium. The second is correct specification: pricing errors must be small enough such that inference under correct model specification is asymptotically justified. Nontraded factors often strain both conditions because they are weakly correlated with returns and the corresponding models are globally misspecified. We therefore organize the methodology around a simple sequence: first identification, then model specification, and only then inference on risk premia.

This ordering is not just expositional. The statistical properties of the estimated premium depend on the strength of the factor exposure matrix and on the magnitude of the pricing errors. Weak exposure variation creates an identification problem; large pricing errors create a model specification problem. The aim of this section is to make those distinctions explicit before outlining our main results.

Consider the panel dataset $\{R_t, f_t\}_{t=1,..,T}$, with R_t denoting an N-vector of gross test-asset returns and f_t a K-vector of candidate risk factors at time t. With respect to the $(K+1)$-vector of stochastic discount factor (SDF) parameters, $\lambda = [\lambda_0, \lambda_1']'$, we define a candidate SDF at time t as $y_t(\lambda) = \lambda_0 + f_t'\lambda_1$. When the asset pricing model is correctly specified and well-identified, there exists a unique set of SDF parameters $\lambda^\star$ such that the following condition holds:

$$\mathbb{E}[e_t(\lambda^\star)] = 0_N, \tag{1}$$

where $e_t(\lambda^\star) = R_t y_t(\lambda^\star) - 1_N$, and 1_N and 0_N are N-vectors of ones and zeros, respectively. Similarly, we can express the linear asset pricing model in beta-pricing form. Let $\beta = \text{Cov}(R_t, f_t)(\text{Var}[f_t])^{-1}$ be the $N \times K$ matrix of factor exposures. With respect to the $K+1$ vector of beta-pricing parameters, $\gamma = [\gamma_0, \gamma_1']'$, define $\alpha(\gamma) = \mathbb{E}[R_t] - 1_N\gamma_0 - \beta\gamma_1$. When the asset-pricing model is correctly specified and well identified, there exists a unique $\gamma^\star = [\gamma_0^\star, \gamma_1^{\star\prime}]'$ such that the following condition holds:

$$\alpha(\gamma^\star) = 0_N, \tag{2}$$

where $\gamma_1^\star$ is the K-vector of true risk premia and $\gamma_0^\star$ is the true zero-beta rate parameter.[2]

In the beta representation, the standard estimation and testing approach is the two-pass

[2]There is a direct link between the two sets of parameters, that is, $\gamma_0 = (\lambda_0 + \mu_f'\lambda_1)^{-1}$ and $\gamma_1 = -V_f\lambda_1(\lambda_0 + \mu_f'\lambda_1)^{-1}$, with $\mu_f = \mathbb{E}[f_t]$ and $V_f = \text{Var}[f_t]$ (Gospodinov et al., 2014b).

procedure of Fama and MacBeth (1973) and Shanken (1992). For the SDF formulation of the model, the Hansen (1982) J (HJ) and Hansen and Jagannathan (1997) distance (HJD) have emerged as the preferred criteria for model estimation and testing. For nontraded factors, researchers often rely on mimicking portfolios by replacing the nontraded factor with its projection onto a set of base asset excess returns that should have maximum correlation with the factor itself (Giglio and Xiu (2021) and Kleibergen and Zhan (2018)).[3] Internet Appendix IA.1 provides a technical discussion of the null hypotheses, assumptions, and interpretations of the various methods.

2.1 Overview of the Empirical Tests

Table 1 groups the various methods into three main blocks: identification, specification, and risk premium inference. The first block includes the rank tests that aim at determining whether the risk premium/SDF parameters are well-identified. Specification tests make up the second block. The third block collects various inference procedures for model parameters. Fama-MacBeth (FM) and Shanken (Shank) are the main benchmarks: they summarize what a researcher would conclude if identification and correct specification both held. The remaining procedures relax those conditions progressively, so that moving down the block traces out how much of the pricing evidence survives once weak identification and/or global model misspecification are properly accounted for.

In the identification stage, we test whether the (augmented) beta or $D = \mathbb{E}[R_t x_t']$ matrices, with $x_t = [1, \ f_t']'$, are of deficient column rank using the four test statistics proposed by Kleibergen and Paap (2006), Gospodinov et al. (2017), Chen and Fang (2019), and Kleibergen and Zhan (2020). The specification stage is concerned with testing $\alpha = 0_N$ and $D\lambda = 1_N$ using the asymptotic tools developed in Kan et al. (2013) for the ordinary least squares (OLS) cross-sectional regression (CSR) R-squared and the HJ and HJD statistics, respectively. The pricing stage is about inference on the model's parameters based on standard two-pass procedures, model misspecification-robust procedures, bootstrap inference, latent-

[3] Cochrane (2009) notes that expressing a pricing model in terms of a traded payoff rather than a nontraded variable such as consumption growth circumvents the measurement difficulties of nontradables, but cautions that projecting the discount factor onto a set of payoffs preserves pricing implications only for those payoffs, not for assets outside the projection basis.

factor methods, and identification-robust confidence intervals (CIs). Throughout the paper, FM-OLS and FM-GLS refer to inference based on Fama and MacBeth (1973), while Shank-OLS and Shank-GLS refer to Shanken (1992)'s inference based on the errors-in-variables (EIV) correction under general distributional assumptions proposed by Jagannathan and Wang (1998).

Table 1 about here

The comparison across pricing procedures is the core of our paper. The CSR-OLS and CSR-GLS (Kan et al., 2013) approaches are employed to determine how much of the FM pricing evidence survives after accounting for global model misspecification. The HJD and continuously-updated GMM (CU-GMM) methods used by Gospodinov et al. (2014a) and Gospodinov et al. (2014b) pose a similar question in the SDF framework and are also robust to the presence of weak/spurious factors in the analysis. BS (Burnside, 2011) and WB (wild bootstrap of Djogbenou and Hounyo (2025)) help to determine whether finite-sample resampling changes those conclusions. In contrast, GX (Giglio and Xiu, 2021) analyze whether an observed factor loads on priced latent variation once omitted factors are captured by a few principal components. We also include the p-values of the GRS-FAR (Kleibergen and Zhan, 2020) and MPAR (Kleibergen and Zhan, 2018) tests, where a nonempty and bounded confidence set that excludes zero is used as evidence that the underlying model's factors are priced. Internet Appendix IA.1 lists the technical assumptions, null hypotheses, and confidence-set interpretations.

These procedures form the first evaluation layer in the paper and help us determine whether published nontraded factor claims are statistically credible once weak identification and global model misspecification are properly accounted for. The next section applies this diagnostic protocol to the model specifications in the original papers. The later sections then investigate whether the proposed factors generate economically meaningful return spreads, and discuss where the widespread pricing inability of many individual factors stems from.

2.1.1 Testing for Weak Identification

Weak identification is the condition most closely tied to the economic nature of nontraded factors. If factor exposures do not vary much across test assets, then the data cannot pin down the risk premium with enough precision. This is a situation where the parameter estimates become inconsistent and standard t-statistics unreliable.

The four rank tests we consider are designed for settings that are similar but not identical to ours. Therefore, we specialize the various test statistics to our testing designs. The Kleibergen-Paap (KP) test is the most general one, as it only relies on mild regularity conditions on the beta estimator. Cragg-Donald (CD) specializes the Cragg and Donald (1997) rank test to linear asset pricing models by analyzing the limiting behavior of the CU-GMM specification test when the moment-condition derivative matrix is rank-deficient. Chen-Fang (CF) recasts the rank hypothesis as $\phi_r(\beta) = 0$, where $\phi_r(\cdot)$ is the sum of the $K - r$ smallest squared singular values of β, and uses bootstrap inference, which improves reliability when heteroskedasticity makes KP-style asymptotics fragile. The finite-sample F-statistic (F-stat) of Kleibergen and Zhan (2020) is closest to a strength diagnostic and is informative when the available time series is short.

Since no single rank test exhibits superior size and power properties in every data environment, we report all four of them in our empirical investigations. In our conservative (toward the null of rank deficiency) summary, we treat a model as well-identified only when all four tests reject the null hypothesis, while in the least conservative summary, we treat a model as well-identified when we reject the null based on at least one test statistic.[4]

2.1.2 Misspecification-Robust Standard Errors

Kan et al. (2013) (KRS) and Gospodinov et al. (2014a) (GKR) show how to conduct asymptotically valid inference when the fundamental asset-pricing restrictions do not hold, that is, the model is globally misspecified. KRS consider the two-pass beta-pricing representation and develop inference on risk premium parameters and CSR R-squared that is robust to global model misspecification (with GLS being also robust to the presence of one or more

[4]We acknowledge that failure to reject a rank test can, in general, reflect i) rank deficiency, but also ii) a weak alternative or iii) low power.

weak/spurious factors in the model). GKR work with a linear SDF representation of the asset pricing model and allow for potential identification failure induced by irrelevant factors (including those that either completely spurious or weak).

This distinction matters for interpretation. Standard FM/Shank inference is asymptotically reliable when the asset pricing model is well-identified and correctly specified. KRS demonstrate that the pricing evidence for nontraded factors substantially weakens once model misspecification is properly accounted for in the beta-pricing model. They also derive the asymptotic distributions of the CSR R^2 for all the possible scenarios, that is, when the model holds perfectly, does not hold at all, and the more general case in which the model is misspecified. GKR employ the HJD framework for the SDF representation of the model, and show that standard inference on the risk premium parameters is asymptotically valid even in the presence of rank failure when misspecification-robust standard errors are employed. From a pricing perspective, we investigate whether the conclusions change drastically going from FM/Shank to KRS/GKR, which would suggest that previous findings based on FM/Shank overstated the importance on nontradables in asset pricing research.

2.1.3 Bootstrap Inference

Burnside (2011) proposes a pairwise bootstrap that resamples factors and test-asset returns jointly. We employ this method to construct bootstrap confidence intervals for risk premia. Relative to FM or KRS, this procedure is appealing because it does not require a separate analytical correction to address the generated-regressor problem. By jointly resampling factors and returns, it is also likely to remain informative under potential weak identification and global model misspecification. We interpret it as a finite-sample check on the asymptotic results rather than as a replacement for the robust parametric procedures.

2.1.4 Wild Bootstrap Inference

The wild bootstrap of Djogbenou and Hounyo (2025) perturbs only the tested factor using Rademacher weights while keeping fixed the test-asset returns and the remaining factors in the model. The resulting test targets the asymptotic GLS misspecification-robust standard error, which is robust to spurious/weak factors, with improved size and power properties

relative to the asymptotic bounds in moderate samples. Because the test applies to factors one at a time conditional on the remaining factors being useful, we treat it as a supporting robustness check rather than as the paper's main criterion for deciding whether a factor is priced.

2.1.5 Three-Step Inference

Giglio and Xiu (2021) estimate latent factors and their loadings from a large panel of observable returns by using principal component analysis (PCA), compute latent factor risk premia, project observed factors onto these latent factors, and then recover the risk premia for the observed factors by combining the latent factor risk premia with the mimicking portfolio weights from the last step. The authors investigate whether the proposed factor loads on priced latent variation after omitted factors have been absorbed by the PCA step. We employ this estimator because the omitted-factor question is central to the nontraded factor zoo, where many observed factors may be weak proxies for broader priced risks.

2.1.6 Inference via GRS-FAR Confidence Intervals

Kleibergen (2009) and Kleibergen and Zhan (2020) construct confidence sets by inverting Anderson-Rubin-style pricing and specification test statistics. Their confidence sets have a direct interpretation. Unbounded sets point to weak identification. Empty sets point to global model misspecification. Bounded sets that include zero mean that a zero premium cannot be rejected under a model that is itself not rejected. Bounded sets that exclude zero provide clean pricing evidence because they require the model to clear the identification and specification hurdles simultaneously.

2.1.7 Identification-Robust Mimicking Portfolio Approach

Kleibergen and Zhan (2018) adapt this logic to the mimicking-portfolio framework. Their procedure is highly suitable for nontraded factors because the relevant empirical question is often whether a factor's projected traded payoff earns a nonzero premium. We construct the spanning assets from principal components rather than from hand-picked portfolios, which

reduces sensitivity to a particular choice of base assets.

3 Data and Model Specifications

3.1 Nontraded Risk Factors

A traded risk factor is a source of systematic risk that is usually expressed as a return (in excess of the risk-free rate) on a traded portfolio or as a return spread. Standard examples of traded factors include the market, size, value, profitability, investment, and momentum factors from Fama and French (1993), Carhart (1997), and Fama and French (2015). In contrast, a nontraded factor is not itself a traded payoff. These factors are often referred to as macroeconomic or economic factors (Balduzzi and Robotti, 2010; Kleibergen and Zhan, 2018), although the category extends way beyond macro variables alone.[5] Many macroeconomic variables fall in this class, but the category is broader: some studies construct tradable and nontradable versions of the same underlying risk (He et al., 2017; Pastor and Stambaugh, 2019).

This distinction matters for inference. Under a correctly specified model, the risk premium on a traded factor equals the expected excess return on its exact mimicking portfolio. For a nontraded factor, the risk premium generally differs from the factor's mean because the factor is not itself a traded payoff. Researchers therefore evaluate nontraded factor pricing either by means of the CSR and HJD methods or through mimicking-portfolio procedures that project the nontraded factor onto the augmented span of basis asset payoffs (Vassalou, 2003; Balduzzi and Robotti, 2008, 2010; Adrian et al., 2014).

3.2 The Nontraded Factor Zoo

We collect 102 distinct nontraded factors from 51 papers published between July 1986 and February 2023 in leading economics, finance, and management journals: *Journal of Finance* (19), *Journal of Financial Economics* (18), *Journal of Political Economy* (4), *Jour-*

[5] We remain agnostic about why the underlying risk is not traded, whether due to market incompleteness, frictions, or theory.

nal 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), and *Review of Accounting Studies* (1). The factors fall into six themes: Consumption (18), Macroeconomics (32), News (11), Sentiment (17), Liquidity (16), and Intermediary and Aggregate Firm-level Risk (8). Tables A.1–A.6 in Appendix A list the distinct factors and their interactions when proposed in the original model specifications, together with their short names, descriptions, and references.

The sample is designed to track proposed factors rather than repeated applications of the same factor. In the spirit of Harvey et al. (2016), we focus on published papers that introduce new factors or new empirical proxies for a priced source of risk.[6] We classify a factor as nontraded if the literature treats it as nontraded or if it is not empirically constructed as a return-based payoff. When the two criteria conflict, we follow the literature's classification.

The zoo includes factors that have been proposed for both cross-sectional pricing and time-series predictability. We therefore include 11 studies that employ nontraded factors primarily in the time-series dimension. This choice broadens the sample of published proxies for priced macroeconomic and financial risks rather than limiting attention to factors originally developed for Fama-MacBeth-style tests.

For factors that are not publicly shared, we reconstruct the series from original data sources. The macroeconomic inputs come from the Federal Reserve Economic Data (FRED) of the Federal Reserve Bank of Saint Louis, the Survey of Professional Forecasters, the University of Michigan inflation-expectation series, and Amit Goyal's data library.[7] Of the 102 distinct factors, 97 are observed at a single frequency: 67 monthly, 21 quarterly, and 9 annual. The remaining five are observed at multiple frequencies: Parker and Julliard (2005)'s consumption growth and Grammig and Jank (2016)'s patenting activity at both annual and quarterly, Campbell and Vuolteenaho (2004)'s cash-flow and discount-rate news

[6] We report the first reference proposing the factor for explaining the cross-sectional or time-series variation in returns.

[7] 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), and Industrial Production.

at both monthly and quarterly, and He et al. (2017)'s intermediary capital factor at both monthly and quarterly. The equity market factor (MKT) and risk-free rate are from Kenneth French's data library.

3.3 Test Portfolios

Nontraded factors are generally not associated with a specific asset class. Because asset-pricing model performance can vary across test assets under model misspecification (Kan et al., 2013), and because conclusions about factor pricing can crucially depend on the choice of test portfolios (Lewellen et al., 2010), we employ multiple test portfolios that are routinely used in the literature. Our scientific replication scheme therefore examines the cross-sectional pricing ability of nontraded factors on thirteen sets of portfolios, covering seven major asset classes. To match the factors at quarterly and annual frequencies, we compound monthly portfolio returns.

We employ 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 on Kenneth French's data library as well as two long-short portfolios of "quality-minus-junk" (Asness et al., 2019) and "betting-against-beta" (Frazzini and Pedersen, 2014) from AQR's data library (US24 anomalies) spanning the sample period July 1963 – August 2022; vi) 48 portfolios formed with the bottom and top portfolios of 24 anomalies studied by Hou et al. (2020) (HXZ48 anomalies) spanning the sample period January 1967 – December 2021; vii) 10 equity portfolios sorted by cash flow duration used by Weber (2018) (Weber10 duration) from July 1963 to June 2014.[8]

[8]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

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 credit default swap portfolios from February 2001 to December 2012 (CDS). The data is available on Asaf Manela's website.

3.4 Replicating the Model Specifications

Our object of interest is the published model specification, not just the standalone factor. We therefore replicate the right-hand-side variables of the original papers as closely as possible. For papers that estimate cross-sectional relations at the individual-asset level, we include the factors that proxy for the characteristics or anomalies used in the original study. Starting from the set of 102 nontraded factors, we obtain 145 distinct published model specifications.

4 Main Results

We report, in the order, identification, specification, and pricing tests. The ordering matters because the finding of a nonzero premium is hard to interpret when the underlying model is weakly identified and/or globally misspecified. In the published nontraded factor specifications, both problems are pervasive, and standard procedures therefore tend to overstate the pricing performance of the candidate factors.

We also study single-factor models and models that add the market factor to a single nontraded factor. These exercises are aimed at detecting whether the identification problem is tied to the published specification or to the proposed factor itself, and whether the

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, see Appendix A.1 of Bandi et al. (2021).

nontraded factor captures information not contained in the market factor. Unless stated otherwise, all tests use a 5% significance level.

4.1 Testing for Weak Identification

Weak identification is widespread in published nontraded factor specifications. We test the rank of factor exposures for the original model specifications on 13 sets of test portfolios. To summarize the information from the four rank tests, we consider a conservative case in which all tests reject rank deficiency and a favorable case in which at least one test rejects it.[9]

Figure 1 about here

Fig. 1 reports the share of original specifications for which we fail to reject the null of weak identification. The KP and finite-sample F-tests classify about 92% and 74% of the specifications as weakly identified, respectively. Based on the CD and CF tests, we still conclude that at least half of the specifications are weakly identified (53% and 58%, respectively). We reject rank deficiency with all four tests for only 5.5% of the models.[10] Weak identification is therefore not a second-order importance issue. It is the dominant scenario in the published nontraded-factor literature.

Sample size naturally matters for these diagnostics. In short samples ($T \leq 100$), the CD test rejects weak identification much more often than the other tests, which points to either relatively good power or rather size distortion due to misspecification. In moderate samples ($100 < T \leq 400$), the KP test still classifies 92% of specifications as weakly identified, versus about 77% for the F-test and 54% and 59% for CF and CD. In large samples ($T > 400$), the gap between the F and CD tests narrows, but weak identification remains widespread. More data helps, but it does not eliminate the problem.

[9]The conservative scenario provides an upper bound on identification failure, while the favorable scenario provides a lower bound. The upper and lower bounds are obtained by outer and inner joins of the results. We therefore do not consider alternative statistics such as the minimum and maximum proportion of rejections from each test.

[10]This figure reports one minus the conservative share of weakly identified models.

Figs. 2 and 3 display the proportion of models that are prone to identification failure across asset classes. The issue is pervasive across portfolios and most pronounced in the non-equity portfolio classes. Among equity portfolios, specifications with the 10 Weber portfolios sorted by cash flow duration appear to be the least identified. Empirical asset pricing studies often use the FF25 ME/BM portfolios as test portfolios, making them the most widely explored portfolios in factor pricing. However, we still observe identification failure for at least 44% of the models. Among non-equity portfolios, sovereign bonds and FX portfolios cause the biggest identification failures across models.

Figures 2, 3, and 4 about here

Fig. 4 groups the identification results by factor theme. The liquidity, sentiment, and intermediary and aggregate firm-level proxies are less likely to fail identification tests than other groups. Consumption factors also look good in some tests, but much of that pattern comes from short samples, which might signal potential finite-sample distortions.

Multiple-testing adjustments make the problem more severe. After controlling the false discovery rate at the 5% level, the share of models classified as weakly identified rises further, from 58% in CD to nearly 99% in KP (Fig. IA.5 in Internet Appendix IA.2.2). Figs. IA.6, IA.7, and IA.8 of the Internet Appendix show that the correction strengthens the weak-identification conclusion within portfolio classes and factor themes as well.

Removing the intercept cuts the share of rank-deficient models by only about 5 percentage points based on the F-test. The restriction on the cross-sectional intercept allows nearly constant betas to identify risk premia (Kleibergen and Zhan, 2020), but this correction does not materially change the overall picture.

4.1.1 Single-Nontraded Factor Models

Adding the market factor to a single nontraded factor usually worsens identification. Fig. 5 shows that the share of weakly identified models rises by 3 to 14 percentage points depending on the test once the market factor is included. This pattern might signal that many nontraded factors are noisy proxies for information already captured by the market factor (Gospodinov and Robotti, 2021).

Figure 5 about here

4.2 Testing for Global Model Misspecification

Misspecification-robust inference leads to a similar conclusion. Fig. 6 shows a large difference in results between the HJ and the HJD tests. The HJ test rejects only 13% of the models, whereas the HJD test rejects about 66% of them. The gap is largest in settings where identification is weak, which is also where the conventional HJ test is least reliable. Two facts explain the divergence. First, the HJ test is heavily distorted when factors are weak. In more than 80% of the original models, the HJ test is close in magnitude to the Anderson-Rubin statistic, which is a warning sign in this setting (Kleibergen and Zhan, 2022). Second, when spurious factors enter the model, one cannot cleanly separate misspecification from identification failure (Gospodinov et al., 2017). Figs. B.6 and B.7 show that the differences in results based on the HJ and HJD tests exist in both equity and non-equity portfolios. Moreover, the CSR R-squared test rejects 36% of the specifications.[11]

Figure 6 about here

4.3 Robust Inference on Risk Premia

When accounting for weak identification and global model misspecification, the pricing evidence for individual nontraded factors becomes sparse. No factor is priced across the full range of tests and test assets. Fig. 7 summarizes the share of priced factors under each method.

Figure 7 about here

The contrast across methods is sharp. FM-OLS classifies about 39% of factors as priced, and FM-GLS lowers that share by about 10 percentage points. Once we allow for model misspecification, CSR-OLS and CSR-GLS reduce the share to about 9%. HJDP lowers it

[11]The GLS variant of the CSR R-squared test provides nearly identical conclusions to those based on the HJD test (rejection rates around 65%; per-model agreement on the 5% reject decision is 98.6%).

to about 4%, while CU-GMM and GX leave it near 7–8%. Burnside's bootstrap yields about 13%, and the wild bootstrap of Djogbenou and Hounyo (2025) yields about 9%. Both bootstrap procedures therefore classify roughly twice as many factors as priced relative to the asymptotic robust methods, but far fewer than FM. Standard methods therefore overstate pricing success by a wide margin.

Figs. B.1 and B.2 show that the decline in pricing performance under robust inference methods appears in both equity and non-equity test portfolios. Fig. B.5 and, in the Internet Appendix, Fig. IA.4 show that the decline is not concentrated in one economic category. Across non-equity portfolios, robust test statistics yield only 4–22% of priced factors for CDS, 1–9% for commodities, 1–8% for FX, 5–15% for options, and 3–17% for US bonds. After multiple-testing adjustments, the pricing evidence nearly disappears (Fig. IA.1 of the Internet Appendix; see also Figs. IA.2 and IA.3).

Table 2 lists the factors that remain priced under the consensus of three misspecification-robust tests, without multiple-testing adjustment, and that also pass the CF identification test.

Table 2 about here

The wild bootstrap p-values of Djogbenou and Hounyo (2025) are also reported. Because the wild bootstrap tests each factor individually, conditional on the remaining factors being useful, it answers a narrower question than the joint misspecification-robust procedures. Among the monthly factors reported in the table, the wild bootstrap rejects the null of no pricing contribution at the 5% level for all specifications.[12] The priced factors include Chan et al. (1996)'s monetary aggregates, Campbell and Vuolteenaho (2004)'s cash-flow news, Parker and Julliard (2005)'s consumption factor, Baker and Wurgler (2006)'s sentiment, Sadka (2006)'s permanent illiquidity, Huang et al. (2015)'s PLS sentiment indexes, Bali et al. (2017)'s macroeconomic uncertainty, He et al. (2017)'s intermediary capital factor, several of Chen et al. (2018)'s adjusted illiquidity measures and Chen et al. (2022)'s attention index.

GRS-FAR and MPAR leave even less pricing evidence. GRS-FAR classifies only about 4% of nontraded factors as priced with bounded confidence intervals that exclude zero. More

[12]The wild bootstrap rejects at the 5% level for at least one proposed factor in at least one set of test assets in about two thirds of the original models.

than 90% of the models produce either unbounded intervals, which signals weak identification, or empty intervals, which signals global model misspecification. Based on MPAR, we detect even fewer priced factors: at most 1.4% survive once the proposed factor is mapped into a traded payoff (the share drops to 0.5% when test assets cover non-equity portfolios). Fig. B.3 in Appendix B.1 summarizes our results.

Restricting the zero-beta rate to equal the risk-free rate by dropping the cross-sectional intercept increases the share of priced factors substantially[13] (Fig. B.4): for FM-OLS from 39% to 54%, for FM-GLS from 29% to 80%, for CSR-OLS from 9% to 15%, for GX from 8% to 16%, and for BS from 13% to 19% . However, we should treat this finding with caution because excluding the intercept from the model can assign pricing power to factors with nearly constant exposures when that variation should be absorbed by the large (estimated) zero-beta rate (Kleibergen and Zhan, 2020; Kroencke and Thimme, 2022). Unless there are strong reasons to believe the model is correctly specified (a large zero-beta rate is a symptom of model misspecification), including the cross-sectional intercept is essential to avoid further nontraded factor proliferation.

These results show that most published nontraded factor claims do not survive once identification and misspecification are accounted for. This conclusion does not yet tell us whether exposures to the proposed factors generate economically meaningful return spreads when they are mapped into implementable trading strategies. A factor may be deemed as irrelevant in an asset pricing test and still produce meaningful return spreads, just like a factor may be found to be priced but fail to produce meaningful trading opportunities. The next section therefore turns from model-level credibility to factor-level economic relevance.

5 Economic Relevance of Nontraded Factors

Section 4 evaluates whether published nontraded factor specifications are statistically credible once identification and global misspecification are properly accounted for. This section examines a different question. Even if formal pricing claims weaken sharply under robust inference, exposures to the proposed factors could still generate economically mean-

[13] We use excess return for this analysis.

ingful return spreads when translated into implementable strategies.

Across 4,428 beta-sorted long–short strategies, only 496, or 11.2%, earn statistically significant average returns at the 5% level, and the corresponding spreads are small in magnitude. The formal tests show that most published nontraded factor claims are statistically fragile; the portfolio evidence shows that the same factors rarely translate into economically meaningful realized spreads.

Beta-sorted portfolios complement the in-sample tests because they map estimated factor exposures into implementable return spreads. For each asset, we estimate rolling betas with respect to a proposed nontraded factor and sort assets on those ex-ante exposures. This approach has a long tradition in empirical asset pricing: Black et al. (1972) and Fama and French (1992) sort stocks into beta-sorted portfolios to test the CAPM, and the original proponents of several nontraded factors in our sample use the same design to establish economic relevance (e.g., Baker and Wurgler, 2006; Adrian et al., 2014; Herskovic et al., 2016; He et al., 2017).

The portfolio sorting exercise is concerned with whether the factor generates a realized return spread, not whether it is priced under maintained identification and correct-specification assumptions. A zero spread indicates either that the ex-ante betas with the factor do not predict subsequent returns, or that the estimation error is too large to recover the return variation associated with betas.

In each month t, we sort assets into portfolios by their estimated rolling betas and compute the return from holding these portfolios. The resulting portfolio returns provide a nonparametric estimate of $\mathbb{E}[r_{i,t+1}|\widehat{\beta}_{i,t}]$. We focus on the high-minus-low beta spread: the return an investor would realize by buying the portfolio with the highest estimated exposure and selling the portfolio with the lowest. Because portfolio formation uses only past information, the long–short spreads are formed based on ex-ante information. As hinted above, the potential drawback of this approach is estimation error: sorting on estimated betas is noisier than sorting on firm characteristics, and that noise might be large for nontraded factors. Poor portfolio performance could therefore reflect measurement error rather than the irrelevance of the underlying sorting variable. We treat the portfolio evidence as a complement to the formal asset pricing tests illustrated above.[14]

[14]Although portfolio formation uses only past information, some proposed nontraded factors, such as

5.1 Constructing Portfolios

The analysis in Section 4 uses standard pre-sorted test portfolios to reduce estimation error in asset pricing tests. We now require test assets to have dispersion in their estimated exposures to each nontraded factor. Because sorted portfolios may compress that dispersion, we estimate betas based on individual CRSP stocks, sort stocks based on those betas, and average returns within beta-sorted portfolios. The resulting portfolios use stock-level variation in ex-ante betas while reducing some idiosyncratic return noise through aggregation.

For each proposed factor at month t, we estimate individual-stock betas from rolling time-series regressions of stock returns on that factor, including an intercept and the controls implied by the beta specification: the other factors in the original published model, no controls in the single-factor specification, or MKT in the single-factor plus MKT specification. We employ the preceding 60 months of returns and require at least 36 nonmissing observations for each stock. When the original model includes multiple proposed factors, we estimate their betas jointly and sort separately on each factor. Stocks are sorted at month t using only information available through that month, and portfolio returns are measured over month $t+1$, so the resulting returns reflect strictly ex-ante and implementable out-of-sample strategies.

We then compute the average return on the high-minus-low beta spread and test for its statistical significance with Newey and West (1987) standard errors (allowing for three lags to account for the serial correlation induced by overlapping returns).[15]

We vary the number of portfolios, the weighting scheme, and the holding period, that is, we form 5, 10, or 25 portfolios, use both equal and value weights, and hold the portfolio composition fixed for 1, 3, or 6 months. These choices generate 18 strategies for each model specification.

principal components or macroeconomic aggregates, rely on full-sample estimates. This source of look-ahead bias is inherited from factor construction, not from the portfolio-sorting procedure.

[15] For six-month holding periods, using more than three lags leaves the overall significance rate essentially unchanged. We do not employ more elaborate inference procedures such as Cattaneo et al. (2023). The reported significance should therefore be interpreted as an upper bound.

5.2 Empirical Results

Our monthly sample contains 70 distinct nontraded factors from 31 papers, corresponding to the monthly subset of the 102 factors documented in Section 3. We augment this set with 13 robustness variants from these papers and two additional studies, yielding 83 monthly factor specifications from 33 papers. For each factor specification, we estimate betas in three different ways, that is, using the original published specification, a single-factor specification, and the single-factor plus MKT specification. We then apply the 18-strategy beta-sorted grid to each case, producing 4,428 long–short strategies.[16]

Fig. 8 reports the distribution of average long–short portfolio returns (% per month) across all specifications for each proposed nontraded factor, showing how research design choices affect estimated premia. Each horizontal box corresponds to a distinct factor, and the width of the box measures the dispersion in average monthly long–short returns obtained under alternative empirical configurations.

Figure 8 about here

Narrow boxes indicate factors whose estimated premia are stable across specifications, whereas wide boxes indicate large design sensitivity (Walter et al., 2024). Fig. 8 shows substantial heterogeneity across factors, but most distributions remain close to zero.

Only 496 of the 4,428 strategies, or 11.2%, produce average returns different from zero at the 5% significance level.[17] This share is low relative to the number of proposed factors, and the estimated premia are usually small.

Table 3 summarizes the results of the beta-sorted portfolio analysis across the 33 papers that propose monthly nontraded factors as sources of systematic risk. For each paper,

[16]The 13 robustness variants are: all three of Bali et al. (2017)'s uncertainty measures, macroeconomic, financial, and real, at three forecast horizons, 1, 3, and 12 months, of which only the one-month horizon of macroeconomic uncertainty enters the main cross-sectional tests; two polynomial expansions of labor income growth used by Dittmar (2002); Kuehn et al. (2017)'s labor tightness factor; and the two risk-neutral variance factors of Martin and Wagner (2019).

[17]The significance rate is similar across design choices: 11.6%, 11.4%, and 10.6% for one-, three-, and six-month holding periods; 11.4% and 11.0% for equal- and value-weighted portfolios; and 12.3%, 11.0%, and 10.2% for 5, 10, and 25 portfolios. No strategy survives a BH correction. We therefore interpret the unadjusted significance rates as upper bounds.

we report the number of model specifications considered, the percentage of statistically significant long—short investment strategies, and the absolute mean and median investment premia among those significant cases.

Table 3 about here

Across studies, significant strategies remain rare and the corresponding premia are small. The absolute mean and median returns in Table 3 rarely exceed 50 basis points.[18] Among the significant strategies, the average absolute long–short spread is about 0.48% per month, or 5.7% annualized. Because this calculation conditions on statistical significance, it describes the magnitude of the surviving spreads rather than the typical premium to factor exposure.

One concern with beta-sorted portfolios applied to nontraded factors is that estimated betas may lack cross-sectional dispersion, making the long–short spread mechanically small. We address this concern by computing the cross-sectional standard deviation of estimated rolling betas for each factor-specification pair. The correlation between beta dispersion and the fraction of significant strategies is essentially zero: factors with larger estimated beta dispersion do not produce significant spreads more often. The median R-squared is 0.18, and factors with R-squared above the median have a significance rate of 12.4%, compared with 10% for those below. These diagnostics suggest that the small spreads are not simply a consequence of mechanically compressed estimated betas, although beta-estimation error remains a potential concern.

To illustrate how the estimated factor premia vary across alternative beta specifications, Table A.7 in the Appendix decomposes the results for each study into the original, single-factor, and single-factor plus MKT specifications, while Fig. B.8 plots the distribution of p-values across all 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 weak beta-sort evidence is not driven by one beta specification alone.

The portfolio evidence therefore matches the in-sample pricing evidence: most nontraded factors fail to generate meaningful returns once we map estimated exposures into implementable trading strategies.

[18] Some studies generate more significant strategies than others because sample periods, estimation windows, and portfolio-construction choices differ across them.

5.3 Robust Pricing Evidence and Return Spreads

The formal tests in Section 4 identify a small set of published nontraded factors whose pricing evidence remains intact under misspecification-robust inference and identification diagnostics. Then, a natural question is whether those same factors also look stronger in the beta-sorting portfolio exercise. This comparison matters because the two parts of the paper answer different questions. Section 4 investigates whether published pricing claims are statistically credible. The portfolio analysis asks whether estimated exposures translate into economically large realized return spreads.

Table 2 lists 17 factor entries that survive the formal tests. One of them is quarterly and cannot be tested in the monthly portfolio exercise. Of the remaining 16 monthly factors, 10 generate at least one significant long–short spread at the 5% level. These 10 factors span six distinct economic families: monetary aggregates (Chan et al., 1996), cash-flow news (Campbell and Vuolteenaho, 2004), sentiment (Baker and Wurgler, 2006; Huang et al., 2015), macroeconomic uncertainty (Bali et al., 2017), intermediary capital (He et al., 2017), and liquidity (Chen et al., 2018). Sentiment and intermediary capital produce the strongest portfolio results, with significance rates of 43% and 34% across design choices. The remaining six monthly factors, including Sadka's permanent illiquidity, several of Chen et al. (2018)'s adjusted illiquidity measures, and Chen et al. (2022)'s attention index, generate no significant return spreads.

The overall agreement between the two exercises strengthens the evidence for this narrow subset of nontraded factors. Most proposed factors fail at both levels: they are not statistically credible under robust inference and they do not generate economically meaningful return spreads. Since the in-sample tests and the out-of-sample portfolio exercise point in the same direction, our conclusion that most nontraded factor proxies may be irrelevant from a financial markets perspective is much harder to dismiss as an artifact of a particular testing environment.

6 Interpreting the Fragility of the Nontraded Factor Zoo

We do not claim that nontraded factors are necessarily unrelated to asset prices. Our evidence shows that published proxies for those risks are difficult to validate in robust cross-sectional pricing tests, and they rarely generate sizable return spreads. The issue is not whether nontraded risk matters, but whether the published proxies isolate that risk well enough to produce reliable pricing evidence.

One possibility is weak beta dispersion. The identification tests in Section 4 show that many published specifications do not generate enough cross-sectional variation in betas to reliably identify risk premia. When factor exposures vary little across assets, or are estimated imprecisely, the data does not allow us to separate a true premium from sampling error. This concern is most relevant for factors that are observed over short samples and are only weakly correlated with the test-asset returns. In this setting, conventional t-statistics are likely to overstate the pricing ability of the proposed nontraded proxies.

A related possibility is that relevant factor exposures are conditional. The robust cross-sectional tests in Section 4 evaluate published specifications using unconditional portfolio-level exposures, while our beta-sorted portfolios allow for time variation only through rolling stock-level betas. Different from the instrumented PCA framework of Kelly et al. (2019), where loadings vary with firm characteristics, our design choices do not account for the full conditional factor structure in asset pricing tests.

A second interpretation is redundancy induced by broader priced variation. When we add the market factor to a single nontraded-factor model, the share of weakly identified models rises by 5 to 14 percentage points depending on the test. This pattern might suggest that many published nontraded factors are not isolated sources of priced risk but noisy empirical proxies for broader variation that is already captured, at least partially, by the market or by other common factors. This interpretation is also consistent with the broader factor zoo literature, where pricing information may reside in latent low-dimensional combinations of candidate factors rather than in many individually strong standalone factors (Kozak et al., 2020). As an exploratory check, Table 4 reports how principal components extracted from the

nontraded factor zoo correlate with common traded-factor benchmarks, and Tables 5 and 6 report the corresponding pricing results. Some aggregate components price selected cross-sections, but the results are not strong enough to overturn the factor-by-factor evidence. These results are consistent with the possibility that some nontraded factors summarize common shocks jointly rather than provide independent sources of pricing information.

The portfolio evidence points in the same direction. If the very weak pricing evidence merely reflected limitations of the formal in-sample tests, one might still expect many nontraded factors to generate meaningful return spreads when assets are sorted on their estimated exposures. Across 4,428 long–short strategies, significant spreads are rare and the associated premia are generally small. The portfolio evidence is thus not very supportive even for factors that survive the robust in-sample tests.

Most empirical applications evaluate nontraded factor claims by two-pass CSR or SDF methods that ignore potential model misspecification and weak identification. Our results suggest that this approach can be misleading when factors are weakly related to test-asset returns and global model misspecification cannot be rejected. At a minimum, empirical work on nontraded factors should report identification diagnostics, misspecification-robust inference, and evidence on economic relevance that does not rely solely on in-sample tests.

7 Conclusion

We conduct a broad investigation of whether published nontraded factors price the cross-section of asset returns once weak identification and fixed model misspecification are taken seriously into account. We collect 102 nontraded factors from 51 papers in top academic journals, reconstruct 145 published model specifications, and evaluate them with both robust cross-sectional tests and beta-sorted portfolio evidence.

Overall, the evidence of pricing and economic gains from nontradables is disappointing. Across the published model specifications we consider, only about 5.5% of them survive the full battery of rank tests that aim at establishing whether the various models are properly identified. Moreover, the Hansen and Jagannathan (1997) distance test rejects correct specification for roughly 66% of the models. Standard Fama–MacBeth procedures classify about

39% of nontraded factors as priced, but that share falls to about 4%–13% once identification- and global misspecification-robust inference is employed. We are also unable to identify nontraded factors that exhibit robust pricing ability across asset classes, although a small subset of nontraded factors has some explanatory power for selected cross-sections.

Our out-of-sample beta-sorting evidence points in the same direction. Across 4,428 long–short strategies, only 11.2% of them earn average returns that are statistically different from zero at the 5% significance level, and the corresponding spreads are small. The consistency between the pricing tests and the portfolio results strengthens our conclusion that the pricing ability of most published nontraded factor proxies is underwhelming from a statistical and economic perspective.

References

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

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, 2023, Climate change concerns and the performance of green vs. brown stocks, *Management Science* 69, 7607–7632.

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.

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.

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.

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.

Black, Fischer, Michael C. Jensen, and Myron Scholes, 1972, The capital asset pricing model: Some empirical tests, in Michael C. Jensen, ed., *Studies in the Theory of Capital Markets*, 79–121 (Praeger, New York).

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.

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, Working Paper.

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

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, 2023, Beta-sorted portfolios, Working Paper.

Chan, K. C., Silverio Foresi, and Larry H. P. 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, 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* 59, 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.

Cochrane, John H., 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.

Djogbenou, Antoine A., and Ulrich Hounyo, 2025, Misspecification-robust bootstrap t-test for irrelevant factor in linear stochastic discount factor models, *Journal of Econometrics* 252, 106097.

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.

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

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.

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, 2014a, 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, 2014b, Spurious inference in unidentified asset-pricing models, Working Paper.

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

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

Herskovic, Bernard, Bryan T. 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 T. Kelly, and Lasse Heje Pedersen, 2023, Is there a replication crisis in finance?, *Journal of Finance* 78, 2465–2518.

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.

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

Kelly, Bryan T., Seth Pruitt, and Yinan Su, 2019, Characteristics are covariances: A unified model of risk and return, *Journal of Financial Economics* 134, 611–643.

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

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, 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, Working Paper.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2020, Shrinking the cross-section, *Journal of Financial Economics* 135, 271–292.

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

Kroencke, Tim A., and Julian Thimme, 2022, A skeptical appraisal of robust asset pricing tests, Working Paper.

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 C. 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.

Lewellen, Jonathan, Stefan Nagel, and Jay Shanken, 2010, A skeptical appraisal of asset pricing tests, *Journal of Financial Economics* 96, 175–194.

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

McLean, David R., 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.

Pastor, Lubos, 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, *Critical Finance Review* 8, 277–299.

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, 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, Working Paper.

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

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

Table 1: **Main test statistics.** The table presents the list of test statistics used in the main paper. For each test statistic, we report its mnemonic, a brief description, and its reference to the 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	Kleibergen and Zhan (2020)
	CF	Rank test	Chen and Fang (2019)
Specification	HJ	Hansen J-test	Hansen (1982)
	HJD	HJD test	Hansen and Jagannathan (1997)
	CSR R^2	OLS CSR R^2	Kan et al. (2013)
Pricing	FM-OLS	Fama-MacBeth CIs	Fama and MacBeth (1973)
	Shank-OLS	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-OLS	Misspecification-robust CIs	Kan et al. (2013)
	CSR-GLS	Misspecification-robust GLS CIs	Kan et al. (2013)
	HJDP	HJD CIs	Gospodinov et al. (2014a)
	CU-GMM	Continuously-updated GMM CIs	Gospodinov et al. (2014b)
	GX	Giglio-Xiu CIs	Giglio and Xiu (2021)
	BS	Burnside bootstrap CIs	Burnside (2011)
	WB	Wild bootstrap CIs	Djogbenou and Hounyo (2025)
	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.** The table presents all factors whose risk premia are uniformly significant under three misspecification-robust tests (CSR-GLS, HJDP, and CU-GMM), together with the associated model specifications, test assets, time-series sample size (T), number of test portfolios (N), and 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, the table also reports the p-values from the FM-OLS, GX, BS, and WB approaches.

								Misspecification-robust						
Reference	Model	Test Portfolios	Factors	T	N	K	FM-OLS	CSR-GLS	HJDP	CU-GMM	GX	BS	WB	CF
Chan et al. (1996)	$M2C_{adj}$	Options	$M2C_{adj}$	310	18	1	0.00	0.01	0.03	0.03	0.32	0.08	0.00	0.00
Chan et al. (1996)	$M2C_{unadj}$	Options	$M2C_{unadj}$	310	18	1	0.00	0.03	0.05	0.03	0.32	0.12	0.02	0.00
Chan et al. (1996)	$M3C_{adj}$	Options	$M3C_{adj}$	239	18	1	0.00	0.01	0.02	0.00	0.02	0.01	0.00	0.00
Chan et al. (1996)	$M3C_{unadj}$	Options	$M3C_{unadj}$	239	18	1	0.00	0.02	0.02	0.00	0.03	0.03	0.01	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	0.00
Parker and Julliard (2005)	Quart_PJconsumption	HXZ48 anomalies	PJconsumption	208	48	1	0.00	0.00	0.01	0.00	0.06	0.05	0.01	0.00
Baker and Wurgler (2006)	$SENT$	FF25 ME/BM	$SENT$	681	25	1	0.32	0.00	0.00	0.01	0.30	0.19	0.00	0.04
Baker and Wurgler (2006)	$SENT$+FF3+MOM	HXZ48 anomalies	$SENT$	660	48	5	0.00	0.02	0.04	0.03	0.11	0.00	0.02	0.00
Baker and Wurgler (2006)	$SENT$	HXZ48 anomalies	$SENT$	660	48	1	0.07	0.03	0.04	0.01	0.11	0.07	0.03	0.02
Sadka (2006)	$ILLIQ_perm$+FF3	USBonds	$ILLIQ_perm$	345	20	4	0.01	0.02	0.02	0.04	0.28	0.10	0.00	0.03
Sadka (2006)	$ILLIQ_perm$+MKT	USBonds	$ILLIQ_perm$	345	20	2	0.03	0.02	0.02	0.02	0.28	0.23	0.00	0.00
Sadka (2006)	$ILLIQ_trans$+$ILLIQ_perm$+MKT	USBonds	$ILLIQ_perm$	345	20	3	0.01	0.02	0.04	0.02	0.28	0.14	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	0.15	0.07	0.00	0.01
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	0.25	0.09	0.00	0.02
Bali et al. (2017)	UNC_1+MKT+controls	CDS	UNC_1	143	20	8	0.00	0.03	0.05	0.04	0.04	0.01	0.02	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	0.00
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	0.01	0.00
Chen et al. (2018)	$ILLIQ_{PC,adj}$	Options	$ILLIQ_{PC,adj}$	310	18	1	0.00	0.00	0.00	0.00	0.84	0.00	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	0.73	0.00	0.00	0.00
Chen et al. (2018)	$ILLIQ_{TO,adj}$	Options	$ILLIQ_{TO,adj}$	310	18	1	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00
Chen et al. (2018)	$ILLIQ_{HM,adj}$	Options	$ILLIQ_{HM,adj}$	310	18	1	0.00	0.00	0.02	0.03	0.78	0.00	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	0.09	0.02	0.00
Chen et al. (2022)	$Attention_{PCA}$	CDS	$Attention_{PCA}$	143	20	1	0.00	0.01	0.02	0.02	0.18	0.05	0.00	0.00

Table 3: **Significance of nontraded beta portfolios.** The table reports summary statistics for beta-sorted portfolios of the nontraded factors proposed in the literature. For each study, we construct high-minus-low portfolios based on rolling beta estimates 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 premia across specifications. All tests are based on monthly returns.

Paper	# Specs	% Sign.	Abs. Avg. (%)	Abs. Med. (%)
Chen et al. (1986)	360	9.17	0.296	0.285
Sweeney and Warga (1986)	36	0.00		
Campbell (1996)	54	0.00		
Chan et al. (1996)	144	11.11	0.380	0.339
Jagannathan and Wang (1996)	72	26.39	0.582	0.628
Dittmar (2002)	612	9.97	0.474	0.507
Pastor and Stambaugh (2003)	72	0.00		
Campbell and Vuolteenaho (2004)	72	20.83	0.408	0.432
Vassalou and Xing (2004)	54	9.26	0.604	0.552
Baker and Wurgler (2006)	252	43.25	0.494	0.435
Sadka (2006)	180	5.00	0.465	0.453
Berkman et al. (2011)	54	0.00		
Hu et al. (2013)	36	0.00		
Da et al. (2015)	162	0.00		
Huang et al. (2015)	72	11.11	0.473	0.452
Bali et al. (2017)	360	19.44	0.591	0.599
Herskovic et al. (2016)	54	50.00	0.420	0.410
He et al. (2017)	108	34.26	0.353	0.339
Kuehn et al. (2017)	54	5.56	0.393	0.442
Manela and Moreira (2017)	36	22.22	0.406	0.383
Boguth and Simutin (2018)	54	18.52	0.276	0.251
Chen et al. (2018)	360	1.39	0.289	0.249
Martin and Wagner (2019)	72	0.00		
Boons et al. (2020)	54	0.00		
Chen et al. (2021)	54	1.85	0.943	0.943
Goldberg and Nozawa (2021)	144	7.64	0.823	0.729
Huang et al. (2021)	144	0.00		
Huynh and Xia (2021)	108	0.00		
Ardia et al. (2023)	72	0.00		
Birru and Young (2022)	216	16.67	0.521	0.458
Chen et al. (2023)	54	16.67	0.457	0.437
Chen et al. (2022)	108	0.93	0.562	0.562
Liu and Matthies (2022)	144	2.08	0.482	0.449

Table 4: **Correlation between principal components of nontraded factors and popular pricing factors.** The 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.** The table reports risk premia (γ_1) 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 (2011) bootstrap confidence intervals for risk premia are also reported. We further report the SDF parameter (λ_1) 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_1(FM-OLS)	-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_1(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_1(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)
BS 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_1(FM-OLS)	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_1(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_1(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)
BS 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_1(FM-OLS)	-0.76^{**}	-1.09	0.78^{**}	-0.26	0.35										
	(0.04)	(0.12)	(0.02)	(0.54)	(0.33)										
γ_1(GX)	-0.49^{**}	-0.44^{**}	0.61^{**}	0.19^{**}	-0.19^{**}										
	(0.01)	(0.01)	(0.01)	(0.04)	(0.04)										
λ_1(CU-GMM)	221.28	827.58	-238.31	510.19	-418.47										
	(0.74)	(0.75)	(0.70)	(0.77)	(0.77)										
BS 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.** The table reports risk premia (γ_1) 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 (2011) bootstrap confidence intervals for risk premia are also reported. We further report the SDF parameter (λ_1) 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_1(FM-OLS)	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_1(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_1(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)
BS 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_1(FM-OLS)	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_1(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_1(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)
BS 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 specifications.** Share of original model specifications for which the null of rank deficiency cannot be rejected. KP denotes the rank test of Kleibergen and Paap (2006); F-stat and F-stat (excl. intercept) denote the finite-sample rank tests used in Kleibergen and Zhan (2020), with and without an intercept; CD denotes the Cragg–Donald rank test used in Gospodinov et al. (2017); and CF denotes the rank test of Chen and Fang (2019). "Conservative" requires all four tests to reject rank deficiency; "Favourable" requires at least one test to reject. Bars are split by time-series sample size: small ($T \leq 100$), medium ($100 < T \leq 400$), and large ($T > 400$).

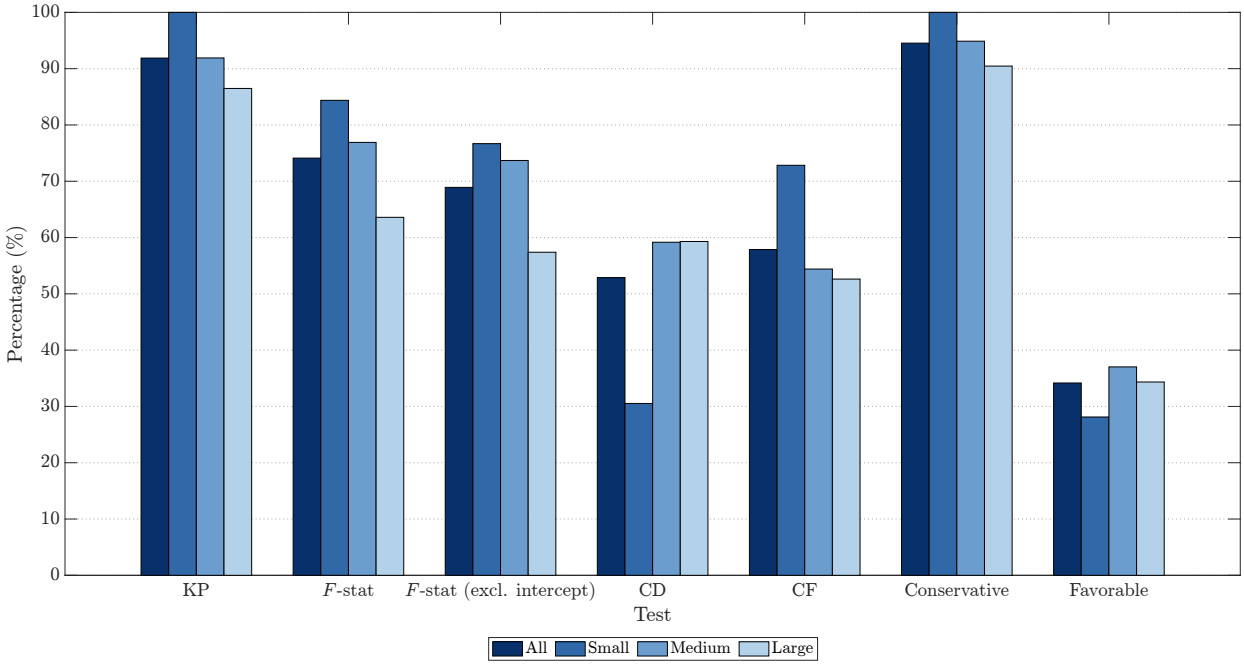


Figure 2: **Identification failure of original model specifications with equity portfolios.** Share of original model specifications for which rank deficiency cannot be rejected in equity portfolios. KP is the rank test of Kleibergen and Paap (2006); F-stat and F-stat (excl. intercept) are the finite-sample rank tests of Kleibergen and Zhan (2020), with and without an intercept; CD is the Cragg–Donald rank test of Gospodinov et al. (2017); CF is the rank test of Chen and Fang (2019).

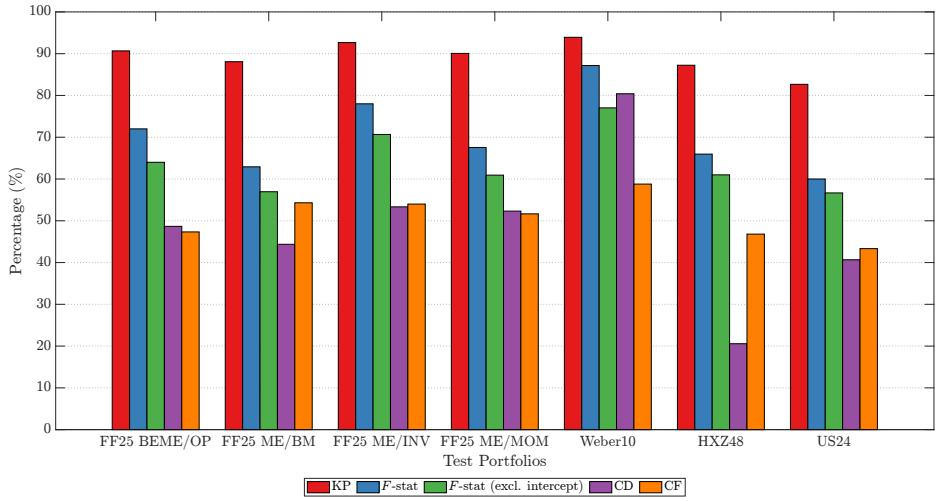


Figure 3: **Identification failure of original model specifications with non-equity portfolios.** Share of original model specifications for which the null of rank deficiency cannot be rejected across different sets of non-equity portfolios. KP is the rank test of Kleibergen and Paap (2006); F-stat and F-stat (excl. intercept) are the finite-sample rank tests of Kleibergen and Zhan (2020), with and without an intercept; CD is the Cragg–Donald rank test of Gospodinov et al. (2017); CF is the rank test of Chen and Fang (2019).

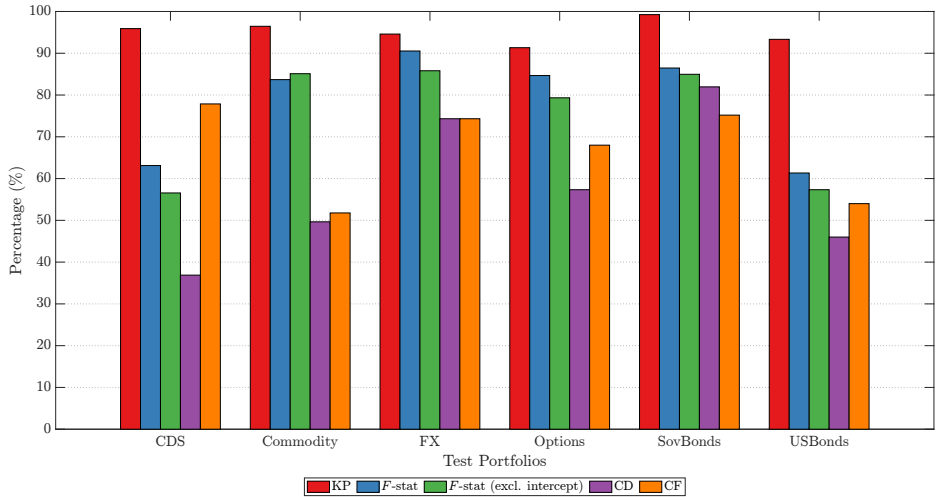


Figure 4: **Identification failure of original model specifications by theme.** Share of model specifications for which the null of rank deficiency cannot be rejected across different factor categories. KP is the rank test of Kleibergen and Paap (2006); F-stat and F-stat (excl. intercept) are the finite-sample rank tests of Kleibergen and Zhan (2020), with and without an intercept; CD is the Cragg–Donald rank test of Gospodinov et al. (2017); CF is the rank test of Chen and Fang (2019).

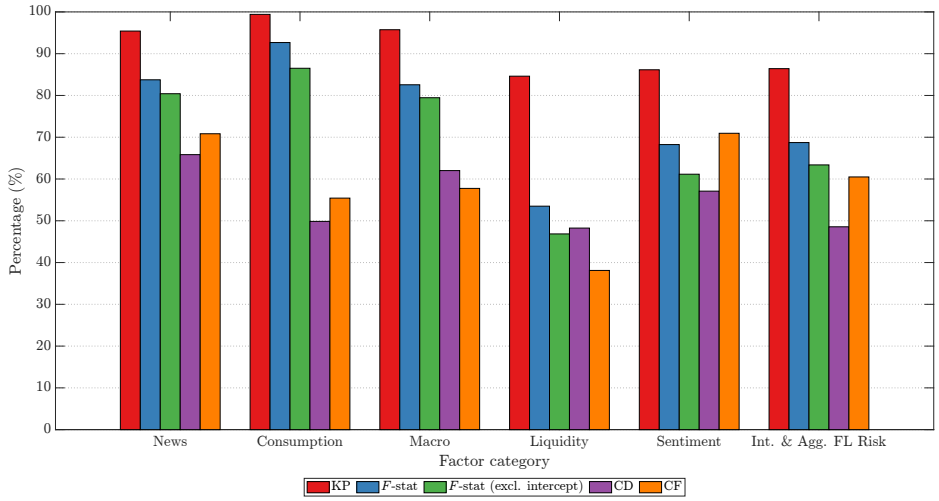


Figure 5: **Identification failure of models with and without controlling market factor.** Share of model specifications for which the null of rank deficiency cannot be rejected, separately for specifications with and without the market factor. KP is the rank test of Kleibergen and Paap (2006); F-stat and F-stat (excl. intercept) are the finite-sample rank tests of Kleibergen and Zhan (2020), with and without an intercept; CD is the Cragg–Donald rank test of Gospodinov et al. (2017); CF is the rank test of Chen and Fang (2019).

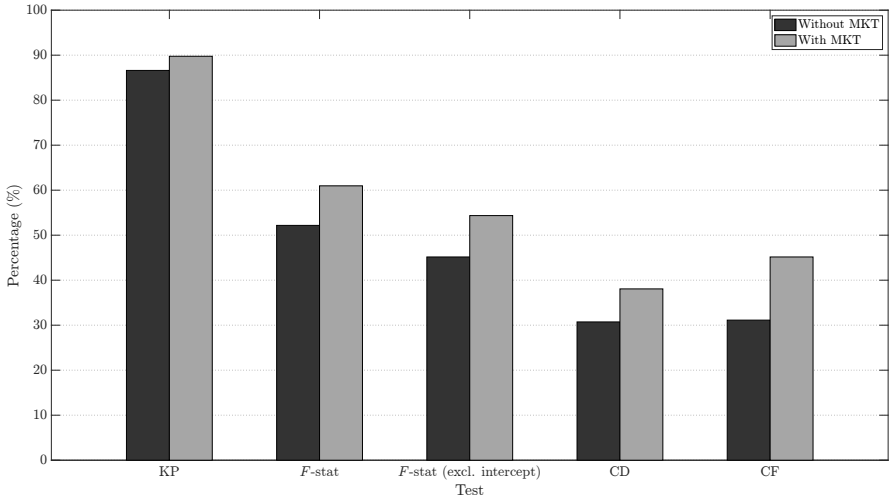


Figure 6: **Global misspecification of the original models.** Share of original model specifications for which the null of correct specification is rejected. HJ denotes the overidentifying-restrictions J test of Hansen (1982); HJD denotes the distance test of Hansen and Jagannathan (1997); and CSR R^2 denotes the cross-sectional R^2 test of Kan et al. (2013). Bars are split by time-series sample size: small ($T \leq 100$), medium ($100 < T \leq 400$), and large ($T > 400$).

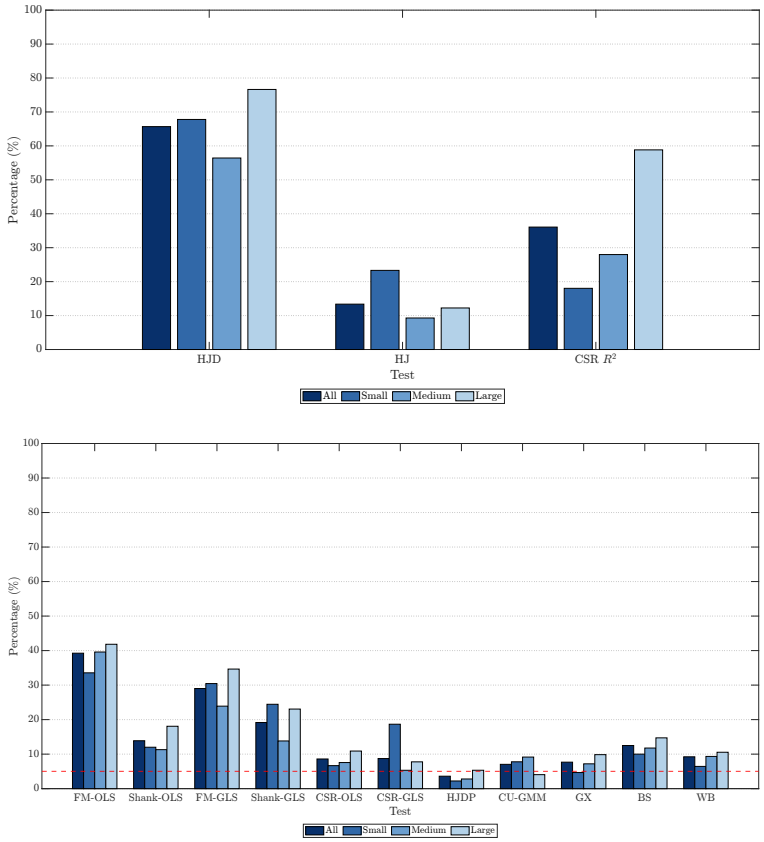


Figure 7: **Priced factors at the 5% significance level.** The figure shows the percentages of priced factors using the 5% significance level pooling across all model specifications and test assets. The percentages are based on the rejection of the null hypothesis that the risk premium is equal to zero. The test statistics are calculated using the standard errors from Fama-MacBeth approach (FM-OLS and FM-GLS), Jagannathan and Wang (1998) generalized Shanken's errors-in-variables allowing for conditional heteroskedasticity (Shank-OLS and Shank-GLS), as well as misspecification-robust standard errors in Kan et al. (2013) (CSR-OLS and CSR-GLS), in Gospodinov et al. (2014a) based on Hansen-Jagannathan distance (HJDP) 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 (GX) and Burnside (2011) bootstrap approach (BS), and the Djogbenou and Hounyo (2025) wild bootstrap (WB). 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).

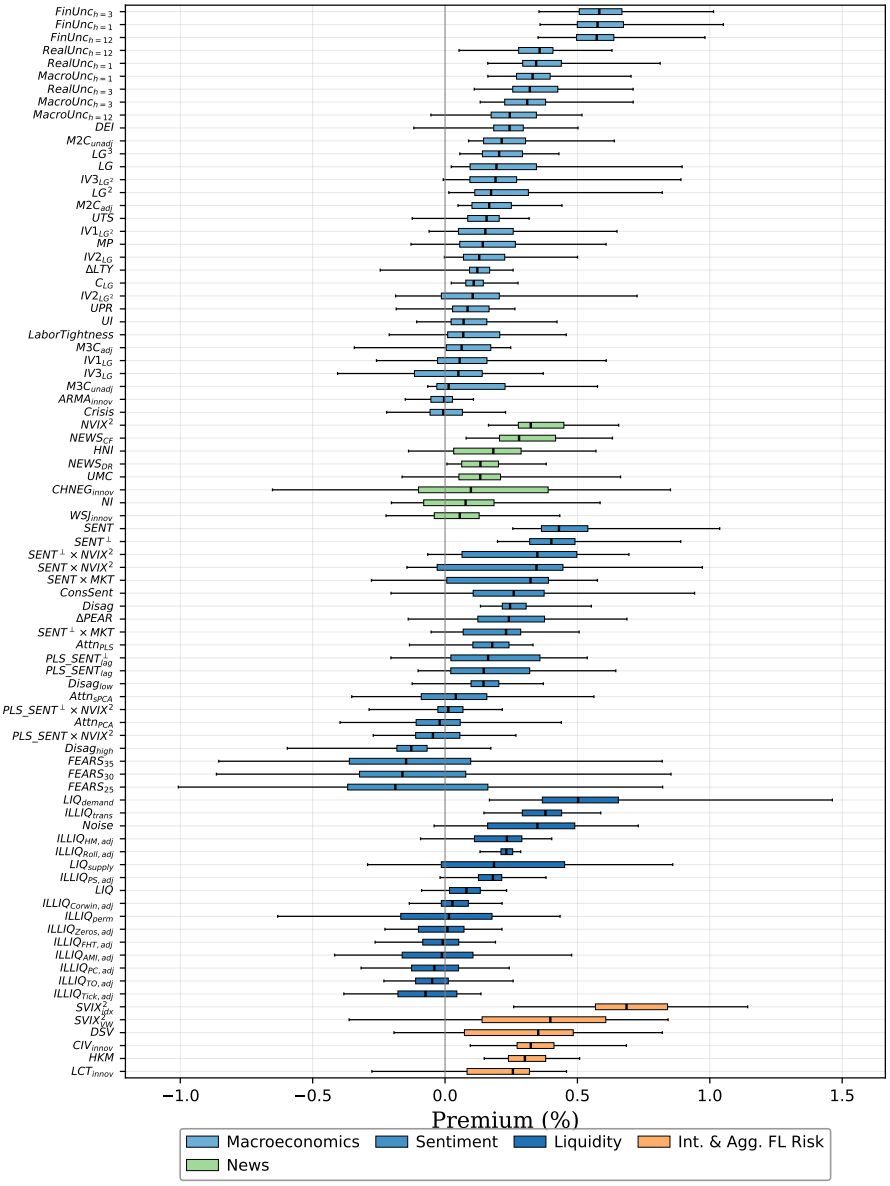


Figure 8: **Premia across empirical designs.**The figure shows the premia in box plots for the monthly nontraded factors per group. Each box covers the interquartile range (IQR) of average monthly returns (in percent) across all portfolio construction choices and model specifications. The maximum and minimum of each distribution is indicated by the lines at the end of each box. Narrow boxes indicate factors whose economic performance is stable across empirical implementations; wide boxes reveal strong sensitivity to design choices. Strategies are sign corrected to show a positive premium.

A List of Nontraded Factors

The following tables list the nontraded factors and interactions that we include in our analysis. We follow a scientific-replication approach. For factors that are publicly available from the original authors, we use the posted series directly; for the remaining factors, we reconstruct the series from the data sources and procedures described in the original papers. 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 nondurable 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

Continued on next page

Table A.2 — *continued*

Short name	**Factor description**	**Reference**	**Label**
$myfar$	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
$LaborTightness$	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. (2023)	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	Pastor 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 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 beta estimates 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 premia across originals, single-factor, and single + MKT 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	4.44	0.260	0.284
	single	90	10.00	0.260	0.262
	single + MKT	90	17.78	0.334	0.315
Chan et al. (1996)	single	72	9.72	0.413	0.409
	single + MKT	72	12.50	0.355	0.328
Jagannathan and Wang (1996)	original	36	36.11	0.513	0.388
	single	18	16.67	0.746	0.732
	single + MKT	18	16.67	0.717	0.735
Dittmar (2002)	original	324	11.73	0.375	0.358
	single	144	8.33	0.645	0.609
	single + MKT	144	7.64	0.628	0.613
Campbell and Vuolteenaho (2004)	original	36	5.56	0.528	0.528
	single	36	36.11	0.389	0.376
Vassalou and Xing (2004)	original	18	5.56	0.393	0.393
	single + MKT	18	22.22	0.657	0.632
Baker and Wurgler (2006)	original	108	46.30	0.442	0.387
	single	72	44.44	0.581	0.528
	single + MKT	72	37.50	0.488	0.405
Sadka (2006)	original	108	1.85	0.494	0.494
	single	36	5.56	0.513	0.513
	single + MKT	36	13.89	0.434	0.442
Huang et al. (2015)	single	36	11.11	0.501	0.499
	single + MKT	36	11.11	0.444	0.431
Bali et al. (2017)	original	36	33.33	0.452	0.419
	single	162	14.81	0.717	0.665
	single + MKT	162	20.99	0.552	0.522
Herskovic et al. (2016)	original	18	50.00	0.439	0.412
	single	18	50.00	0.462	0.419
	single + MKT	18	50.00	0.358	0.290
He et al. (2017)	original	72	36.11	0.343	0.334
	single + MKT	18	61.11	0.377	0.370
Kuehn et al. (2017)	original	18	5.56	0.281	0.281
	single	18	11.11	0.450	0.450
Manela and Moreira (2017)	single	18	5.56	0.656	0.656
	single + MKT	18	38.89	0.370	0.321
Boguth and Simutin (2018)	original	18	11.11	0.221	0.221
	single	18	27.78	0.279	0.256
	single + MKT	18	16.67	0.308	0.320
Chen et al. (2018)	single + MKT	180	2.78	0.289	0.249
Chen et al. (2021)	single	18	5.56	0.943	0.943
Goldberg and Nozawa (2021)	original	72	11.11	0.666	0.659
	single	36	8.33	1.241	1.222
Birru and Young (2022)	single	72	33.33	0.524	0.484
	single + MKT	72	16.67	0.515	0.452
Chen et al. (2023)	original	18	22.22	0.463	0.442
	single	18	22.22	0.440	0.436
	single + MKT	18	5.56	0.503	0.503
Chen et al. (2022)	single + MKT	54	1.85	0.562	0.562
Liu and Matthies (2022)	single	36	2.78	0.586	0.586
	single + MKT	36	5.56	0.429	0.429

B Figures

B.1 Risk Premia Inference

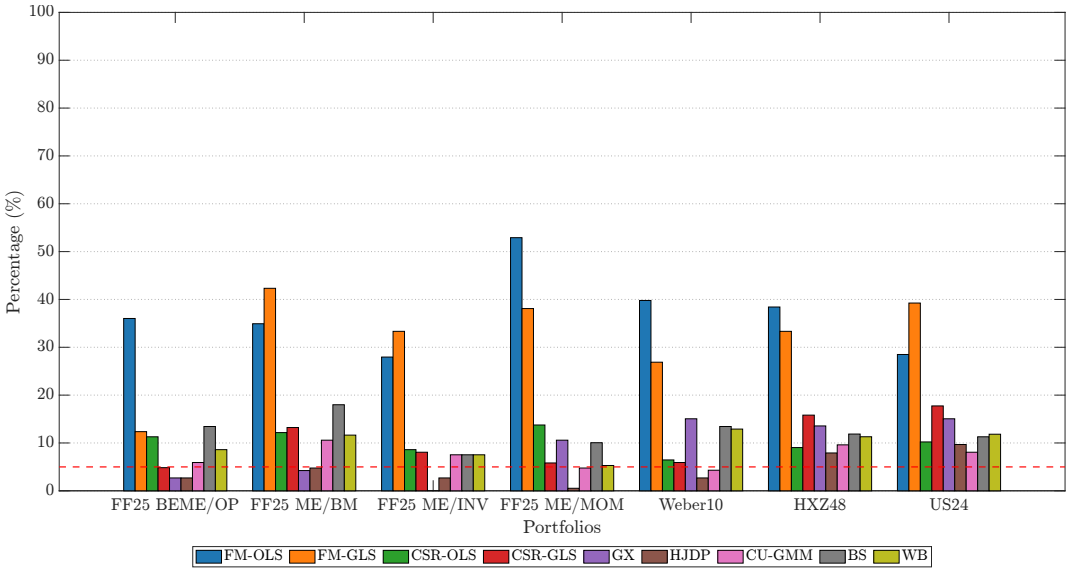


Figure B.1: **The percentages of priced factors at a 5% significance level with equity portfolios**
For each set of equity portfolios, the figure reports the percentages of factors with nonzero risk premia at the 5% level, suggested by each test. The test statistics are calculated using the standard errors from Fama-MacBeth approach (FM-OLS and FM-GLS), as well as misspecification-robust standard errors in Kan et al. (2013) (CSR-OLS and CSR-GLS), in Gospodinov et al. (2014a) based on Hansen-Jagannathan distance (HJDP) 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 (GX) and Burnside (2011) bootstrap approach (BS), and the Djogbenou and Hounyo (2025) wild bootstrap (WB).

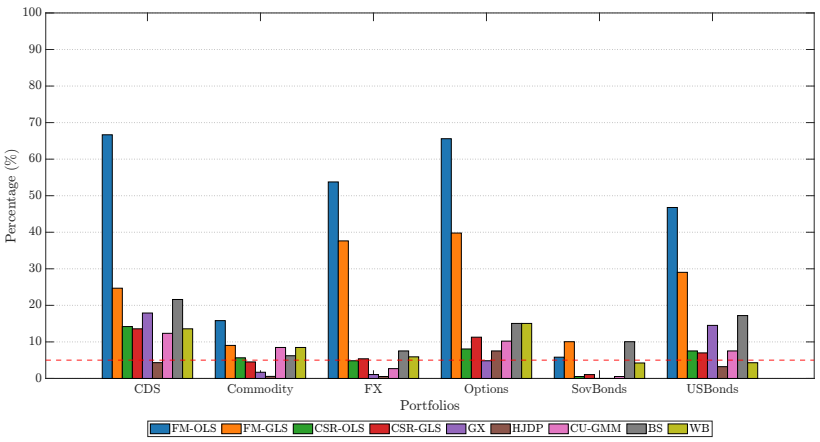


Figure B.2: **The percentages of priced factors at a 5% significance level with non-equity portfolios**

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-OLS and FM-GLS), as well as misspecification-robust standard errors in Kan et al. (2013) (CSR-OLS and CSR-GLS), in Gospodinov et al. (2014a) based on Hansen-Jagannathan distance (HJDP) 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 (GX) and Burnside (2011) bootstrap approach (BS), and the Djogbenou and Hounyo (2025) wild bootstrap (WB).

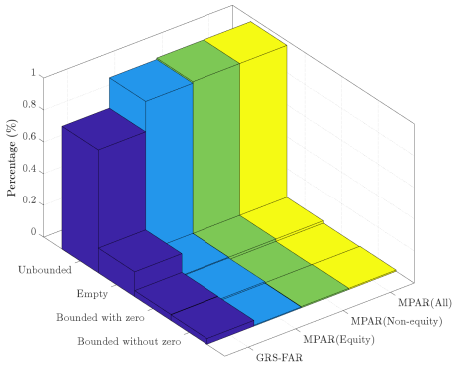


Figure B.3: **Confidence intervals of risk premia**

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.

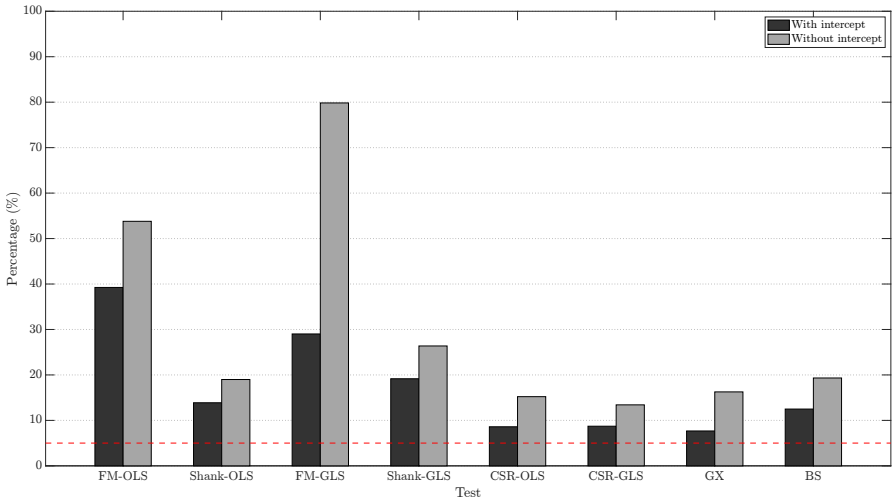


Figure B.4: **The percentages of priced factors with and without intercept**
This figure compares the percentages of priced factors at the significance level of 5% including and excluding the cross-sectional intercepts. The percentages are based on the rejection of the null hypothesis that the risk premium is equal to zero. The test statistics are calculated using the standard errors from Fama-MacBeth approach (FM-OLS and FM-GLS), Jagannathan and Wang (1998) generalized Shanken's errors-in-variables allowing for conditional heteroskedasticity (Shank-OLS and Shank-GLS), as well as misspecification-robust standard errors in Kan et al. (2013) (CSR-OLS and CSR-GLS). We also report the percentages of rejection with Giglio and Xiu (2021) three-pass approach (GX) and Burnside (2011) bootstrap approach (BS).

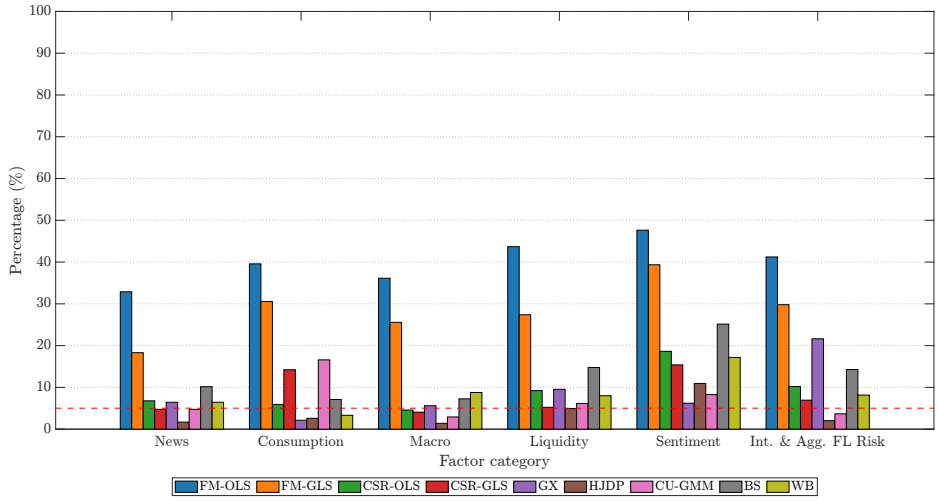


Figure B.5: **The percentages of priced factors by factor category**
This figure shows the percentages of priced factors for each factor category at 5% significance level. The test statistics are calculated using the standard errors from Fama-MacBeth approach (FM-OLS and FM-GLS), as well as misspecification-robust standard errors in Kan et al. (2013) (CSR-OLS and CSR-GLS). We also report the percentages of rejection with Giglio and Xiu (2021) three-pass approach (GX) and Burnside (2011) bootstrap approach (BS), and the Djogbenou and Hounyo (2025) wild bootstrap (WB).

B.2 Misspecification

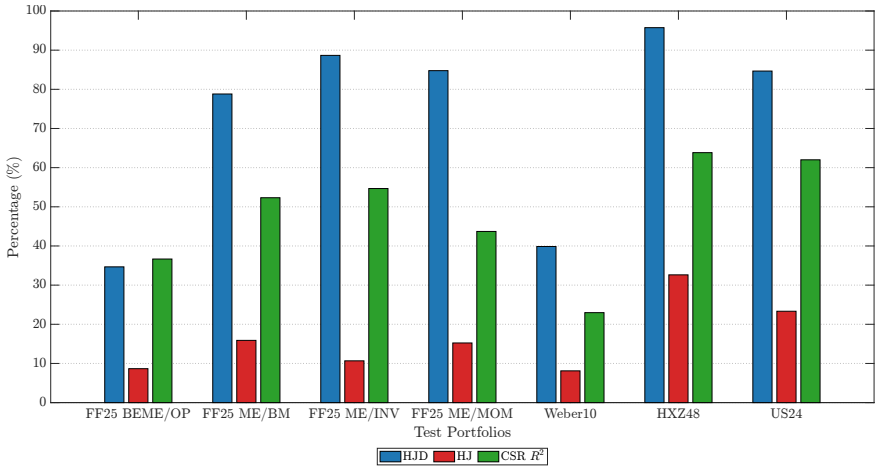


Figure B.6: **Global misspecification of model specifications with equity portfolios.** Share of model specifications for which the null of correct specification is rejected in equity portfolios. HJ is the overidentifying-restrictions J test of Hansen (1982); HJD is the distance test of Hansen and Jagannathan (1997); and CSR R^2 is the cross-sectional R^2 test of Kan et al. (2013).

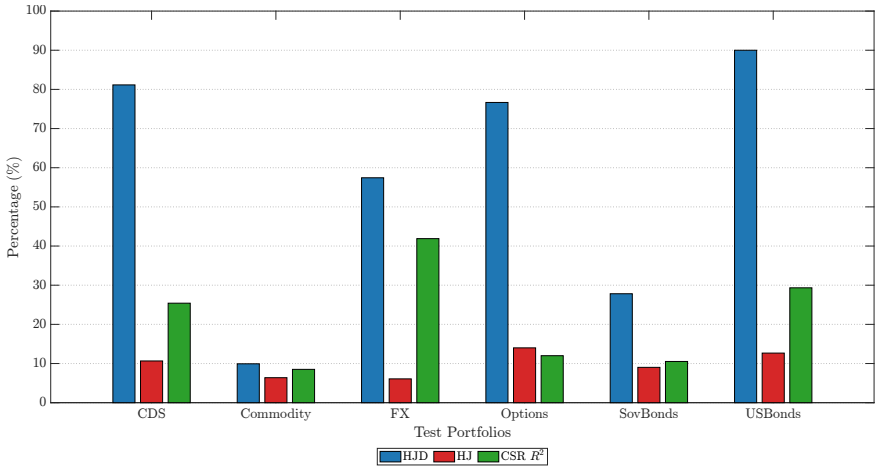


Figure B.7: **Global misspecification of original model specifications with non-equity portfolios.** Share of original model specifications for which the null of correct specification is rejected in non-equity portfolios. HJ is the overidentifying-restrictions J test of Hansen (1982); HJD is the distance test of Hansen and Jagannathan (1997); and CSR R^2 is the cross-sectional R^2 test of Kan et al. (2013).

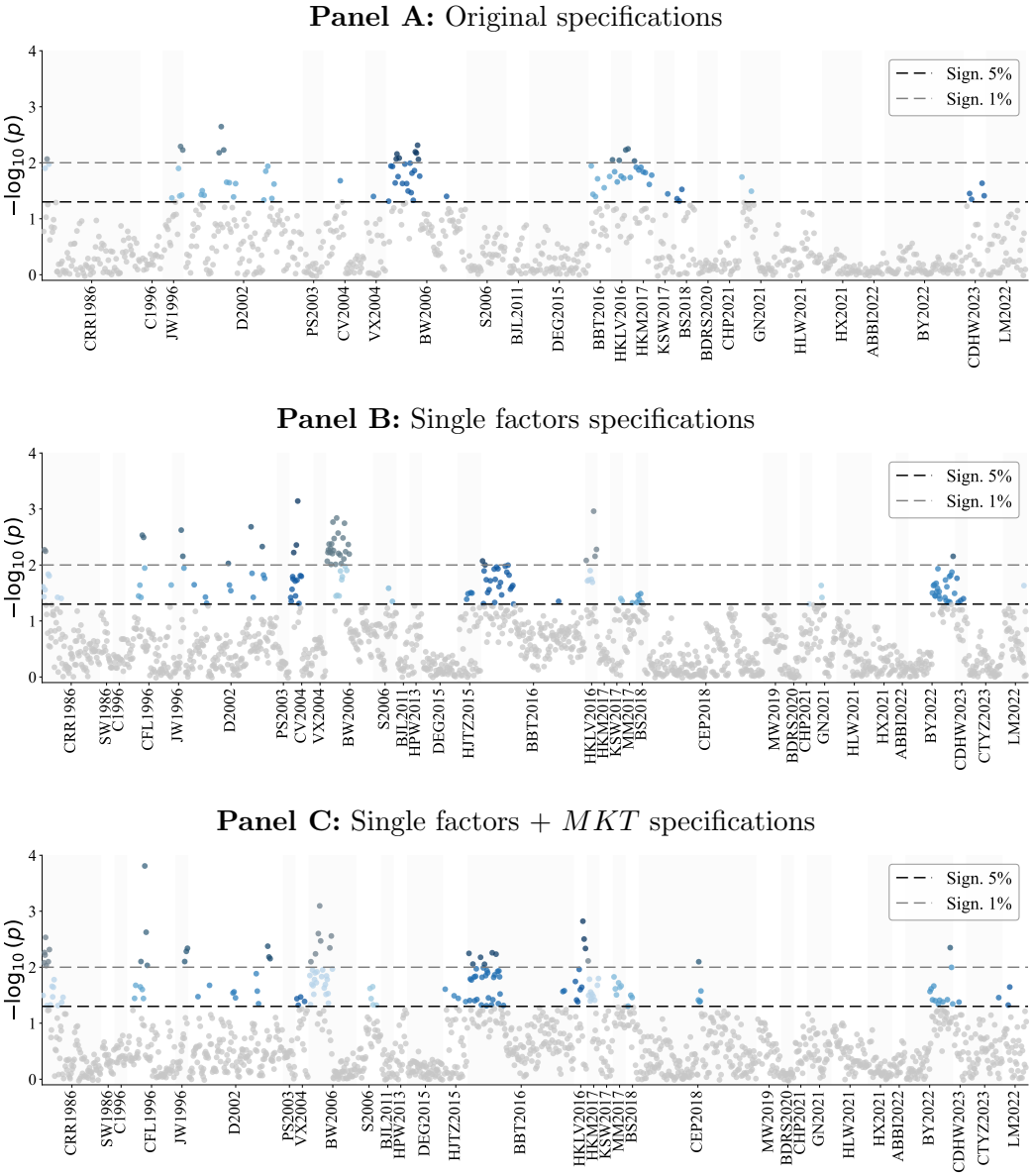


Figure B.8: **Significance of portfolio premia from β-sorted portfolios across papers.** This figure plots the $-\log_{10}(p\text{-value})$ from tests of whether average long-short portfolio expected returns differ from zero for the original specifications (Panel A), single-factor models (Panel B), and single-factor models augmented with the market factor (Panel C). Higher values indicate stronger statistical significance. The labels follow the factor definitions presented in Tables A.1–A.5.

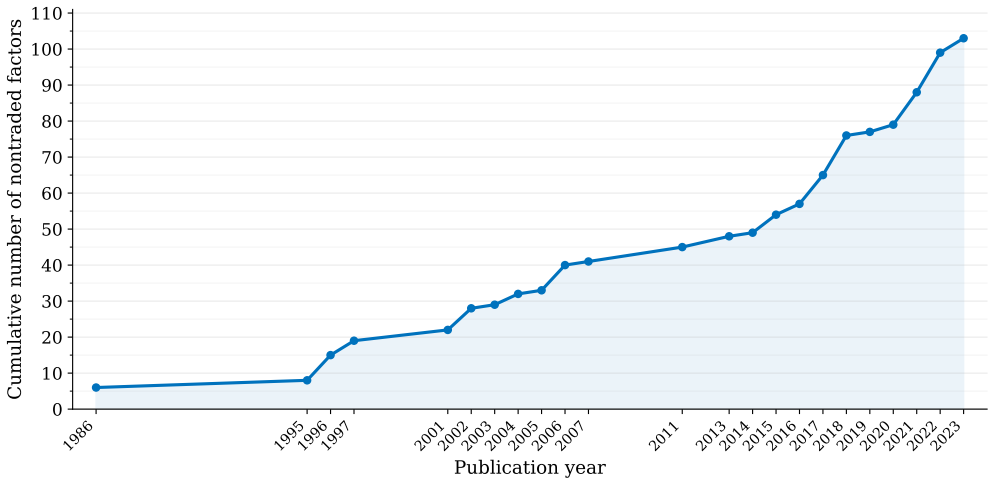


Figure B.9: **The cumulative number of nontraded factors proposed in the literature from 1986 to 2023**

Internet Appendix for

Misspecification and Weak Identification in the Nontraded Factor Zoo

Appendix IA.1 Methods Details

This appendix provides a summary of the methodologies used in the paper.

IA.1.1 Identification diagnostics

The four identification tests evaluate whether the matrix of beta loadings has full column rank. KP tests the rank of β using the singular-value decomposition of a $\sqrt{T}$-asymptotically normal estimator, under the assumptions that the asset returns and factor time series are jointly stationary and ergodic with positive-definite variance and finite fourth moments. CD applies the rank test framework in the GMM setting and it connects rank failure to the distortions that spurious factors induce in SDF-based specification framework. CF rewrites the rank hypothesis in terms of the smallest singular values and adopts a bootstrap procedure to improve reliability when heteroskedasticity makes KP-style asymptotics fragile. The finite-sample F-statistic of Kleibergen and Zhan (2020) is used as a first-stage strength diagnostic for the beta matrix.

IA.1.2 Specification diagnostics

Both HJ and HJD methodologies test whether pricing errors are zero. The HJ is the standard GMM test of overidentifying restrictions and is asymptotically chi-squared under correct specification and identification. The HJD instead measures the minimum distance between the candidate linear SDF and the set of admissible SDFs that price the test assets and it is more informative about economic misspecification, especially when weak factors are included.

IA.1.3 Standard and misspecification-robust risk-premium inference

FM and Shank are included as benchmarks. Shank corrects the resulting errors-in-variables problem arising when estimated betas are used as regressors in the second step of the FM regression. Both procedures still assume correct specification and strong enough identification for delta-method inference to be reliable.

KRS extends the two-pass beta representation to allow for nonzero pricing errors. Relative to Shank, the KRS methodology takes into consideration global misspecification. GKR

can be used in the SDF representation. The HJDP and CU-GMM statistics are designed to remain valid when the model is misspecified and at least one factor may be irrelevant. In the paper, the comparison between KRS and GKR helps separate the role of misspecification alone from the role of misspecification together with factor irrelevance.

IA.1.4 Bootstrap and latent factor procedures

The Burnside bootstrap approach jointly resamples factors and test asset returns and This appendix provides a summary of the methodologies used in the paper.

IA.1.5 Identification diagnostics

The four identification tests evaluate whether the matrix of beta loadings has full column rank. KP tests the rank of β using the singular-value decomposition of a $\sqrt{T}$-asymptotically normal estimator, under the assumptions that the asset returns and factor time series are jointly stationary and ergodic with positive-definite variance and finite fourth moments. CD applies the rank test framework in the GMM setting and it connects rank failure to the distortions that spurious factors induce in SDF-based specification framework. CF rewrites the rank hypothesis in terms of the smallest singular values and adopts a bootstrap procedure to improve reliability when heteroskedasticity makes KP-style asymptotics fragile. The finite-sample F-statistic of Kleibergen and Zhan (2020) is used as a first-stage strength diagnostic for the beta matrix.

IA.1.6 Specification diagnostics

Both HJ and HJD methodologies test whether pricing errors are zero. The HJ is the standard GMM test of overidentifying restrictions and is asymptotically chi-squared under correct specification and identification. The HJD instead measures the minimum distance between the candidate linear SDF and the set of admissible SDFs that price the test assets and it is more informative about economic misspecification, especially when weak factors are included.

IA.1.7 Standard and misspecification-robust risk-premium inference

FM and Shank are included as benchmarks. Shank corrects the resulting errors-in-variables problem arising when estimated betas are used as regressors in the second step of the FM regression. Both procedures still assume correct specification and strong enough identification for delta-method inference to be reliable.

KRS extends the two-pass beta representation to allow for nonzero pricing errors. Relative to Shank, the KRS methodology takes into consideration global misspecification. GKR can be used in the SDF representation. The HJDP and CU-GMM statistics are designed to remain valid when the model is misspecified and at least one factor may be irrelevant. In the

paper, the comparison between KRS and GKR helps separate the role of misspecification alone from the role of misspecification together with factor irrelevance.

IA.1.8 Bootstrap and latent factor procedures

The Burnside bootstrap jointly resamples factors and test asset returns and re-run the two-pass procedure in each bootstrapped sample, addressing the generated regressor issue. We interpret it as a finite-sample robustness check as bootstrap performance can still be sensitive to persistent dependence and near-boundary identification.

The wild bootstrap of Djogbenou and Hounyo (2025) methodology perturbs the factor under scrutiny with Rademacher weights while holding returns and the remaining factors fixed, and targets the misspecification-robust t-statistic when a useless factor is included in the model specification.

GX three-step procedure estimates latent factors from the covariance matrix of returns, projects the observed factors onto the space spanned by the estimated latent factors to estimate the risk premia of the set of observed factors. The procedure is robust to omitted factors and is derived under Bai (2003)-style assumptions, where factors are strong and pervasive, and $N, T \to \infty$ with $N/T \to c > 0$.

IA.1.9 Identification-robust confidence sets

FAR, GRS-FAR, and MPAR rely on test inversion rather than delta-method approximations. FAR is the Anderson-Rubin-style building block for identification-robust inference on risk premia while GRS-FAR extends this logic by jointly testing the premium and the model's pricing restrictions. Bounded sets excluding zero indicate pricing with identification and specification, bounded sets including zero indicate no pricing under an otherwise acceptable model, unbounded sets indicate weak identification, and empty sets indicate misspecification.

MPAR methodology applies the same idea to the mimicking portfolio approach. Since nontraded factors are projected on test assets, we use principal components extracted from the test assets as basis portfolios. In practice, MPAR is often the most conservative procedure in the paper because it requires the nontraded factor to survive both projection and identification-robust inference.

Appendix IA.2 Figures

IA.2.1 Risk Premia Inference

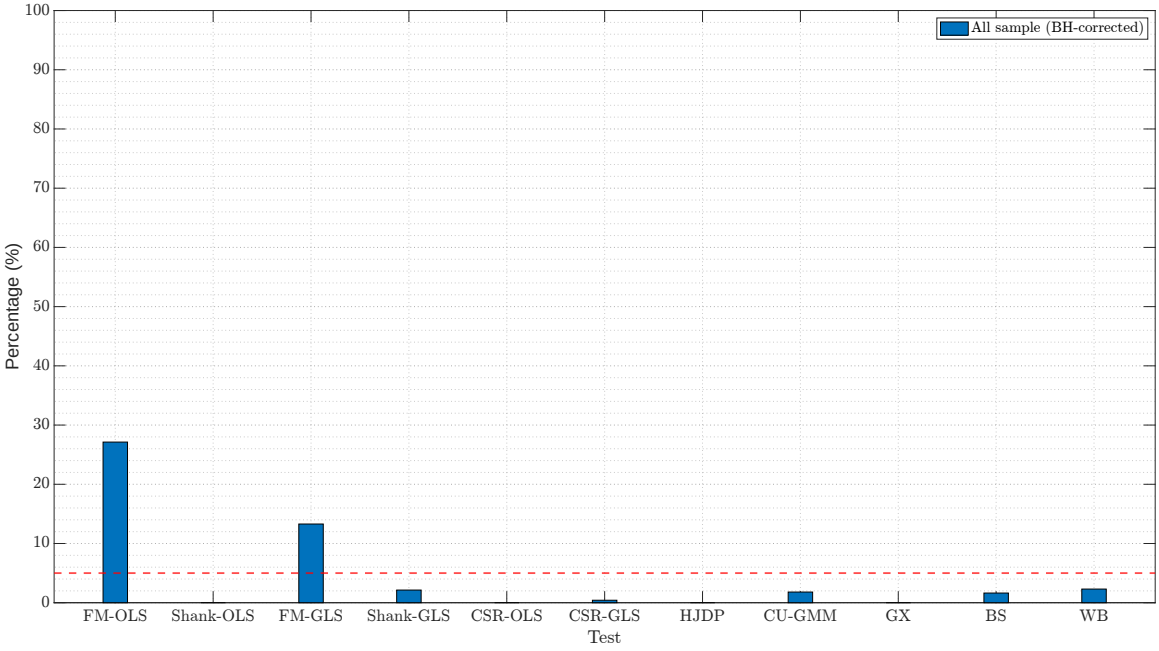


Figure IA.1: The percentages of priced factors at the 5% significance level controlling for multiple hypothesis testing

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) correction. 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-OLS and FM-GLS), Jagannathan and Wang (1998) generalized Shanken's errors-in-variables allowing for conditional heteroskedasticity (Shank-OLS and Shank-GLS), as well as misspecification-robust standard errors in Kan et al. (2013) (CSR-OLS and CSR-GLS), in Gospodinov et al. (2014) based on Hansen-Jagannathan distance (HJDP) 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 (GX) and Burnside (2011) bootstrap approach (BS), and the Djogbenou and Hounyo (2025) wild bootstrap (WB).

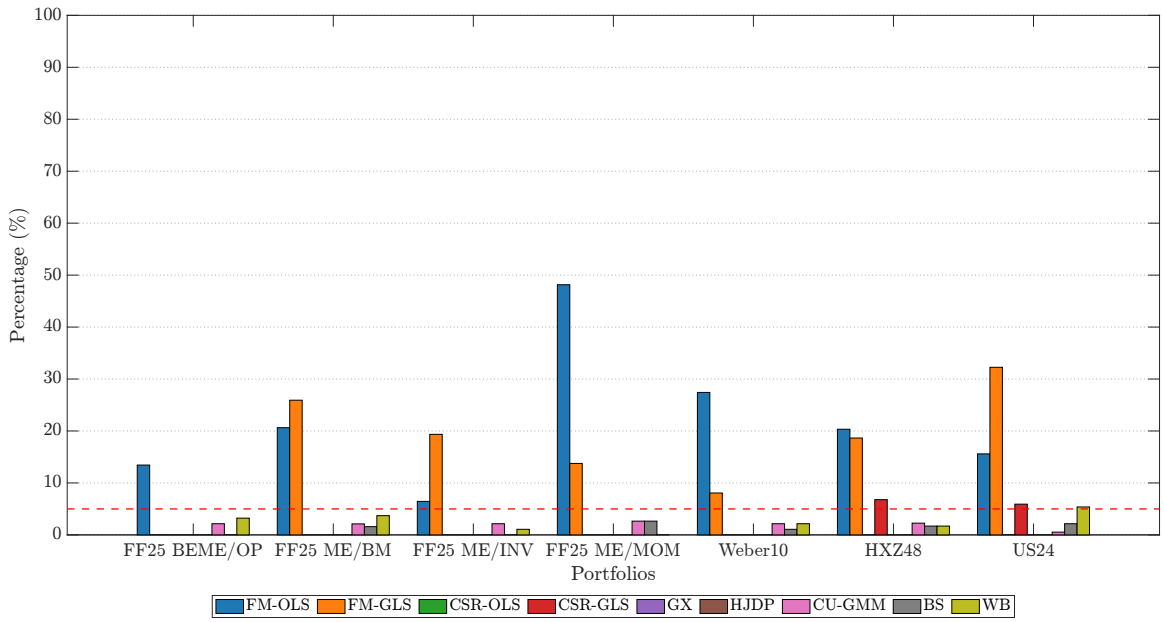


Figure IA.2: The percentages of priced factors at a 5% significance level with equity portfolios controlling for multiple hypothesis testing

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) correction. The test statistics are calculated using the standard errors from Fama-MacBeth approach (FM-OLS and FM GLS), as well as misspecification-robust standard errors in Kan et al. (2013) (CSR-OLS and CSR-GLS), in Gospodinov et al. (2014) based on Hansen-Jagannathan distance (HJDP) 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 (GX) and Burnside (2011) bootstrap approach (BS), and the Djogbenou and Hounyo (2025) wild bootstrap (WB).

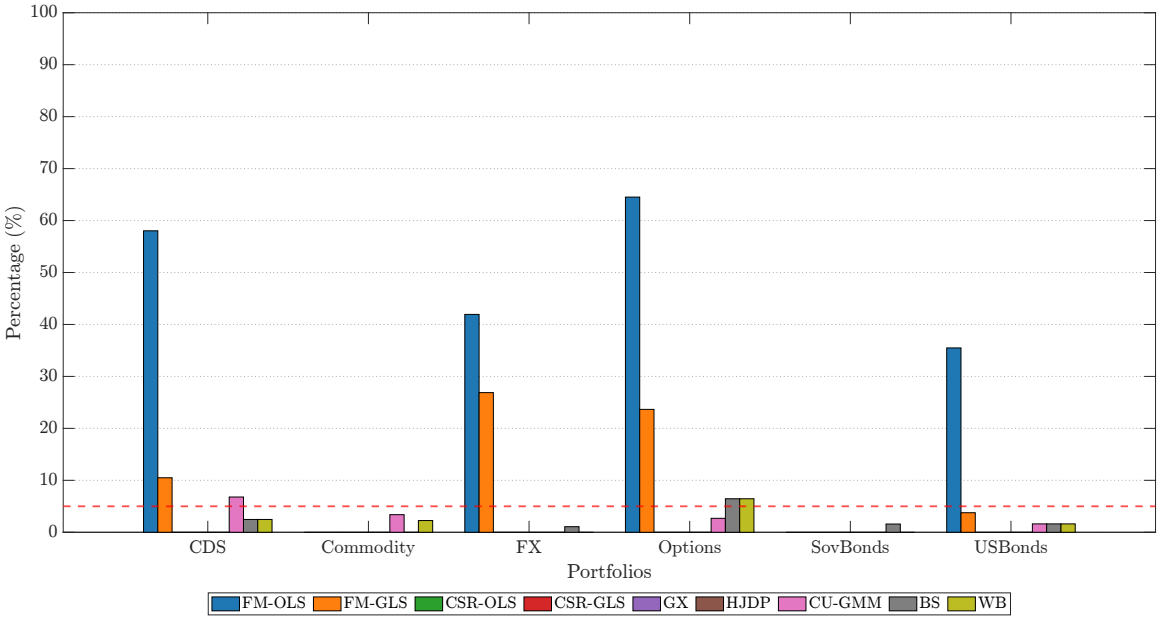


Figure IA.3: The percentages of priced factors at a 5% significance level with non-equity portfolios controlling for multiple hypothesis testing

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) correction. 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-OLS and FM-GLS), as well as misspecification-robust standard errors in Kan et al. (2013) (CSR-OLS and CSR-GLS), in Gospodinov et al. (2014) based on Hansen-Jagannathan distance (HJDP) 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 (GX) and Burnside (2011) bootstrap approach (BS), and the Djogbenou and Hounyo (2025) wild bootstrap (WB).

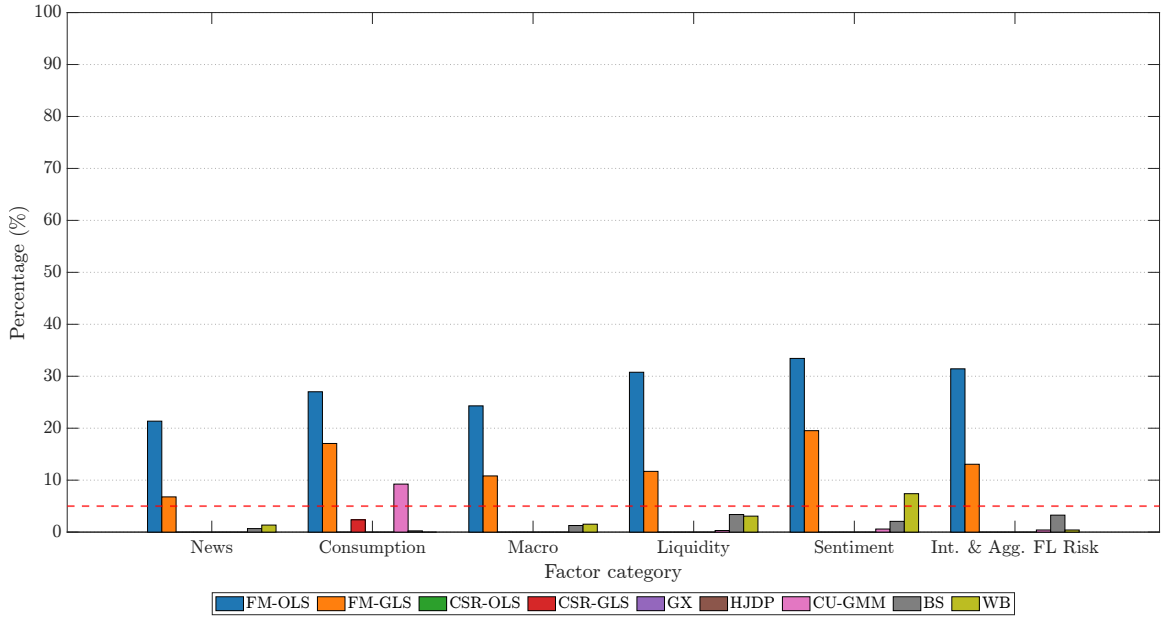


Figure IA.4: The percentages of priced factors by factor category controlling for multiple hypothesis testing

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-OLS and FM-GLS), as well as misspecification-robust standard errors in Kan et al. (2013) (CSR-OLS and CSR-GLS). We also report the percentages of rejection with Giglio and Xiu (2021) three-pass approach (GX) and Burnside (2011) bootstrap approach (BS), and the Djogbenou and Hounyo (2025) wild bootstrap (WB).

IA.2.2 Identification

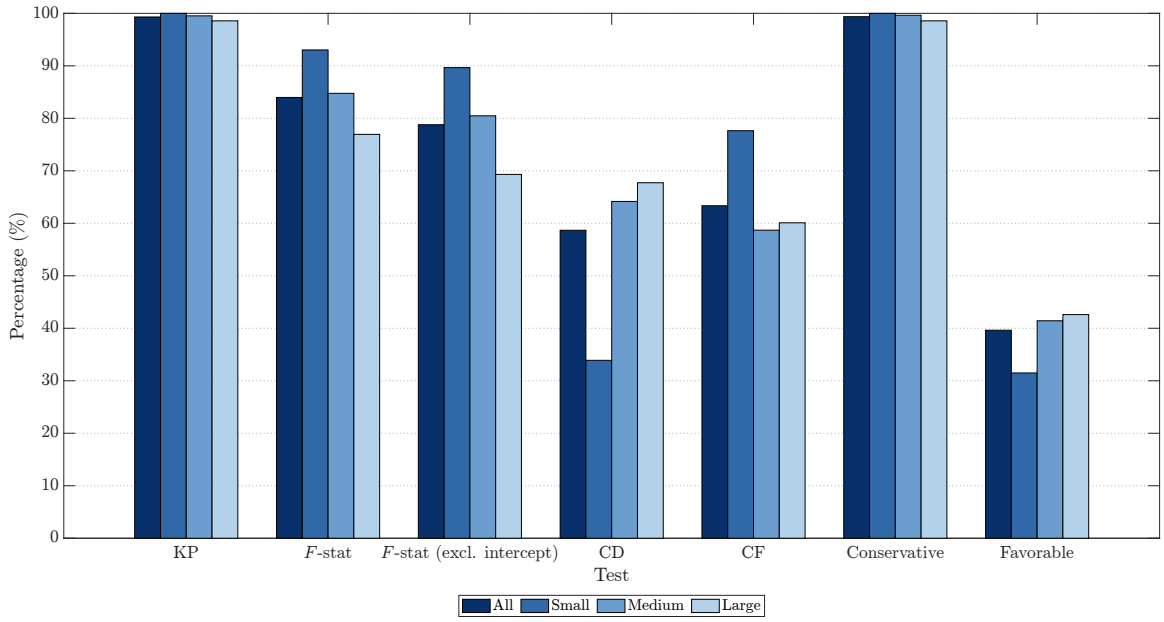


Figure IA.5: Identification failure of original model specifications after multiple-testing adjustment.

Share of original model specifications for which the null of rank deficiency cannot be rejected after controlling the false discovery rate at 5% using the procedure of Benjamini and Hochberg (1995). KP is the rank test of Kleibergen and Paap (2006); F-stat and F-stat (excl. intercept) are the finite-sample rank tests of Kleibergen and Zhan (2020), with and without an intercept; CD is the Cragg–Donald rank test of Gospodinov et al. (2017); CF is the rank test of Chen and Fang (2019). "Conservative" requires all four tests to reject rank deficiency after the multiple-testing adjustment; "Favourable" requires at least one test to reject. Bars are split by time-series sample size: small ($T \leq 100$), medium ($100 < T \leq 400$), and large ($T > 400$).

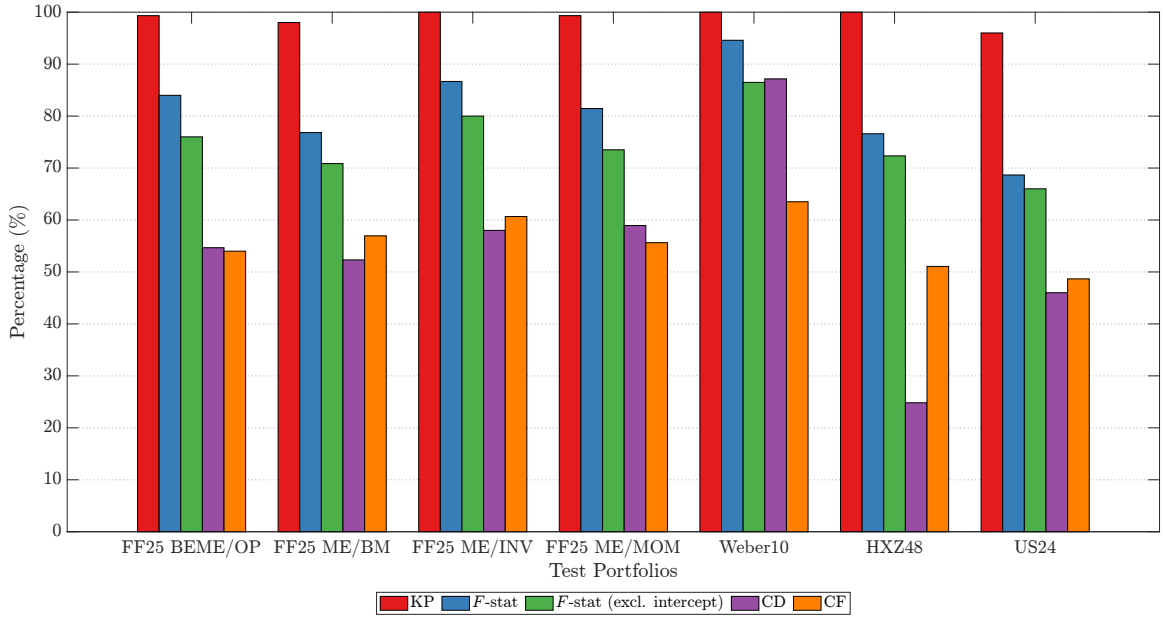


Figure IA.6: Identification failure of original model specifications with equity portfolios after multiple-testing adjustment.

Share of original model specifications for which the null of rank deficiency cannot be rejected across different sets of equity portfolios after controlling the false discovery rate at 5% using the procedure of Benjamini and Hochberg (1995). KP is the rank test of Kleibergen and Paap (2006); F-stat and F-stat (excl. intercept) are the finite-sample rank tests of Kleibergen and Zhan (2020), with and without an intercept; CD is the Cragg–Donald rank test of Gospodinov et al. (2017); CF is the rank test of Chen and Fang (2019).

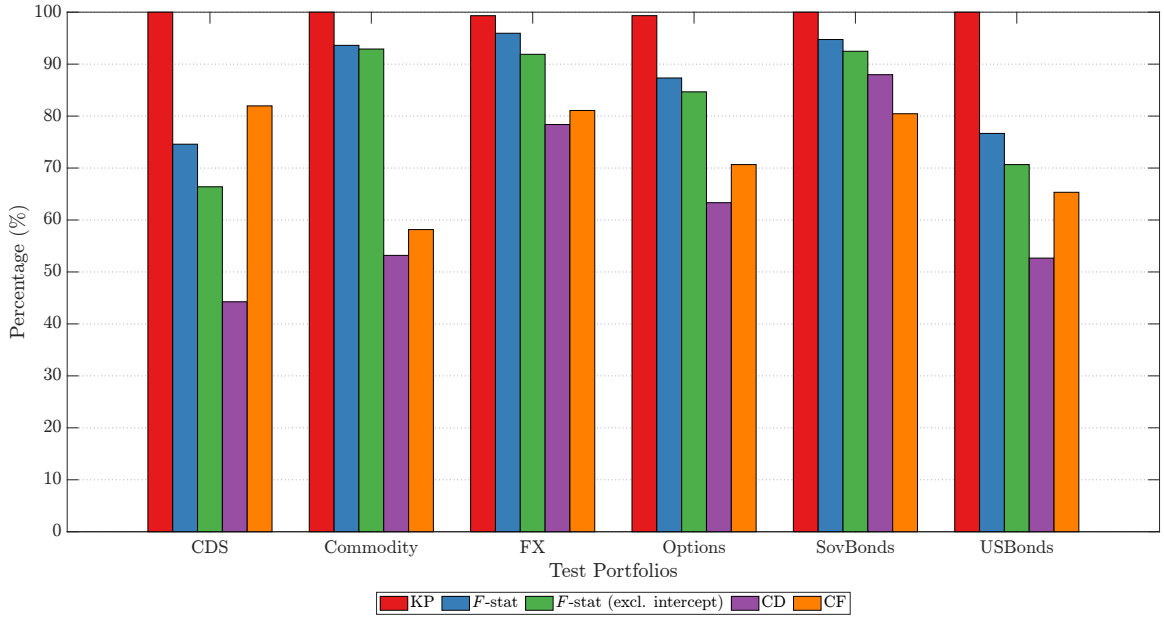


Figure IA.7: Identification failure of original model specifications with non-equity portfolios after multiple-testing adjustment.

Share of original model specifications for which the null of rank deficiency cannot be rejected across different sets of non-equity portfolios after controlling the false discovery rate at 5% using the procedure of Benjamini and Hochberg (1995). KP is the rank test of Kleibergen and Paap (2006); F-stat and F-stat (excl. intercept) are the finite-sample rank tests of Kleibergen and Zhan (2020), with and without an intercept; CD is the Cragg–Donald rank test of Gospodinov et al. (2017); CF is the rank test of Chen and Fang (2019).

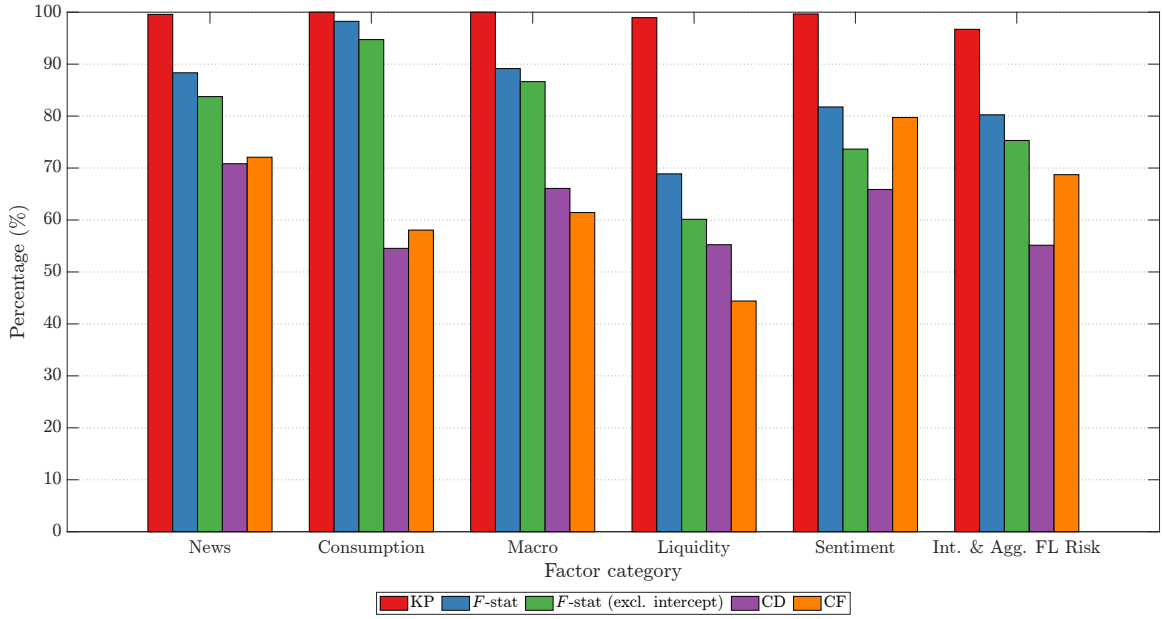


Figure IA.8: Identification failure of original model specifications by theme after multiple-testing adjustment.

Share of original model specifications for which the null of rank deficiency cannot be rejected across different factor categories after controlling the false discovery rate at 5% using the procedure of Benjamini and Hochberg (1995). KP is the rank test of Kleibergen and Paap (2006); F-stat and F-stat (excl. intercept) are the finite-sample rank tests of Kleibergen and Zhan (2020), with and without an intercept; CD is the Cragg–Donald rank test of Gospodinov et al. (2017); CF is the rank test of Chen and Fang (2019).

References

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

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.

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

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

Djogbenou, Antoine A., and Ulrich Hounyo, 2025, Misspecification-robust bootstrap t-test for irrelevant factor in linear stochastic discount factor models, *Journal of Econometrics* 252, 106097.

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

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.

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

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.

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, 2020, Robust inference for consumption-based asset pricing, *Journal of Finance* 75, 507–550.