

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 $N, T \rightarrow \infty$ with $N/T \rightarrow c > 0$.

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

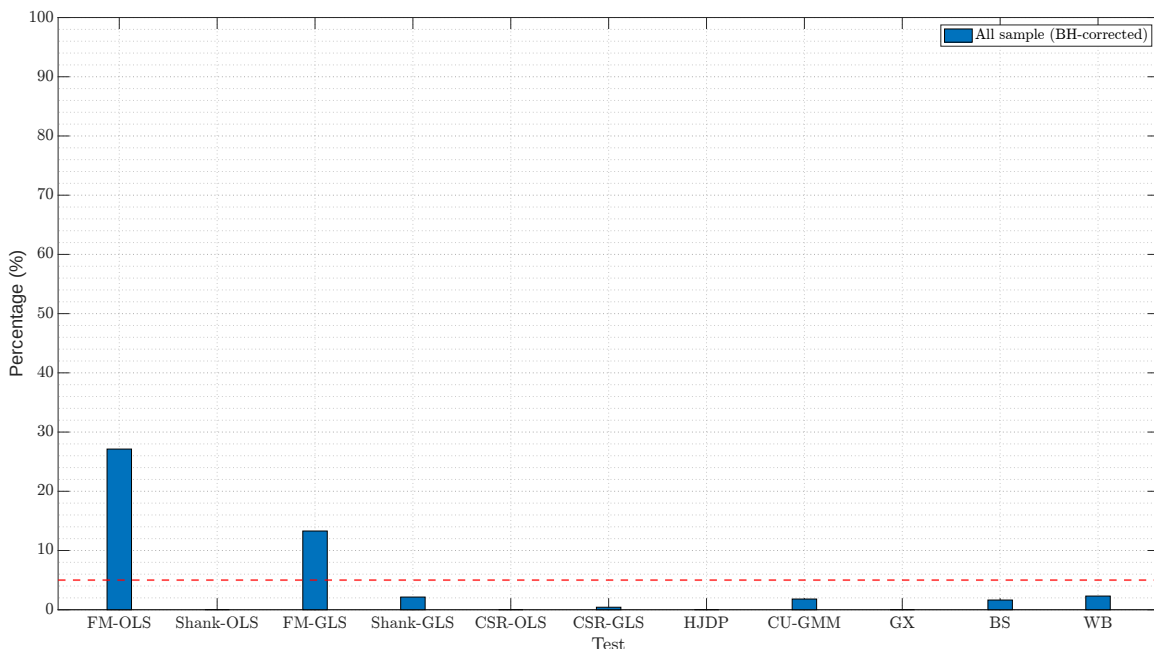


Figure IA.1: The percentages of priced factors at the 5% significance level controlling for multiple hypothesis testing

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).

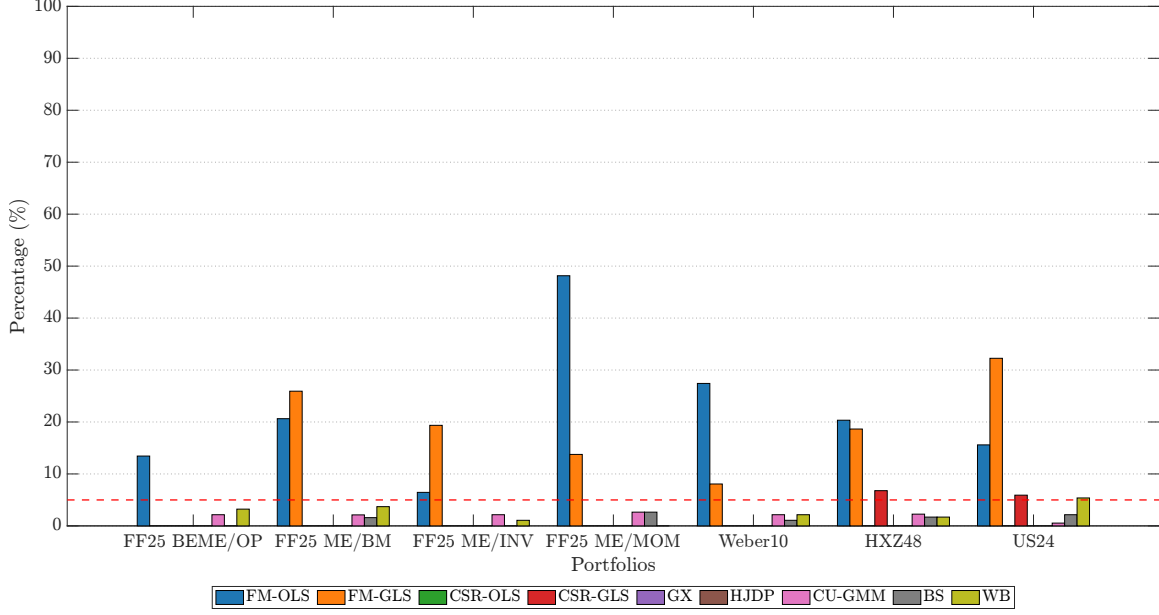


Figure IA.2: The percentages of priced factors at a 5% significance level with equity portfolios controlling for multiple hypothesis testing

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).

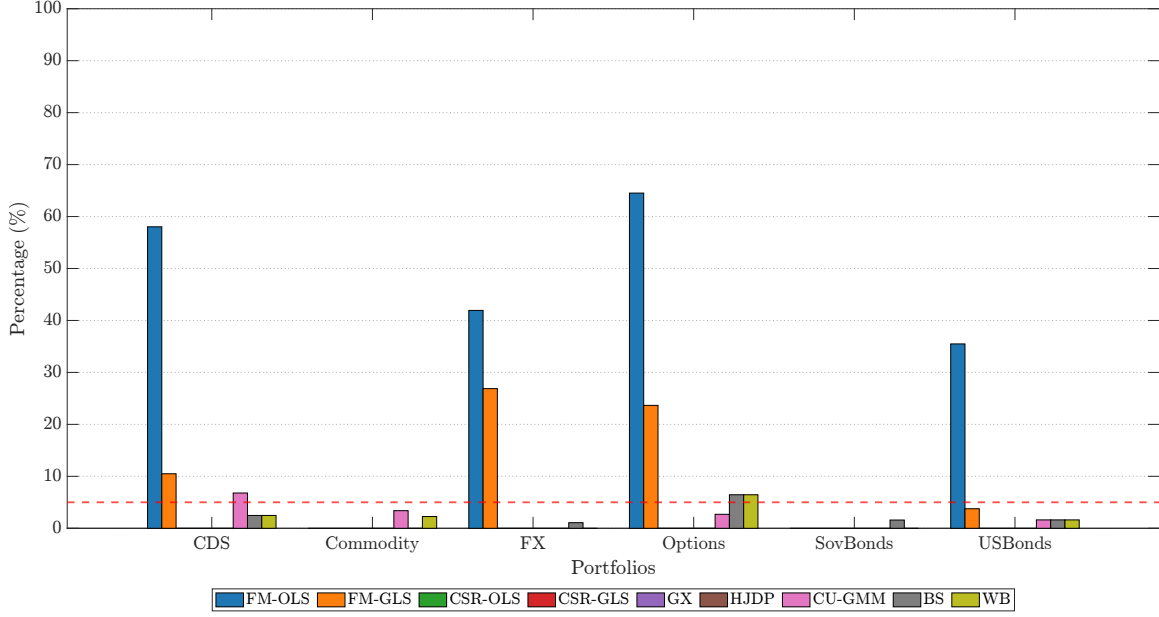


Figure IA.3: The percentages of priced factors at a 5% significance level with non-equity portfolios controlling for multiple hypothesis testing

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).

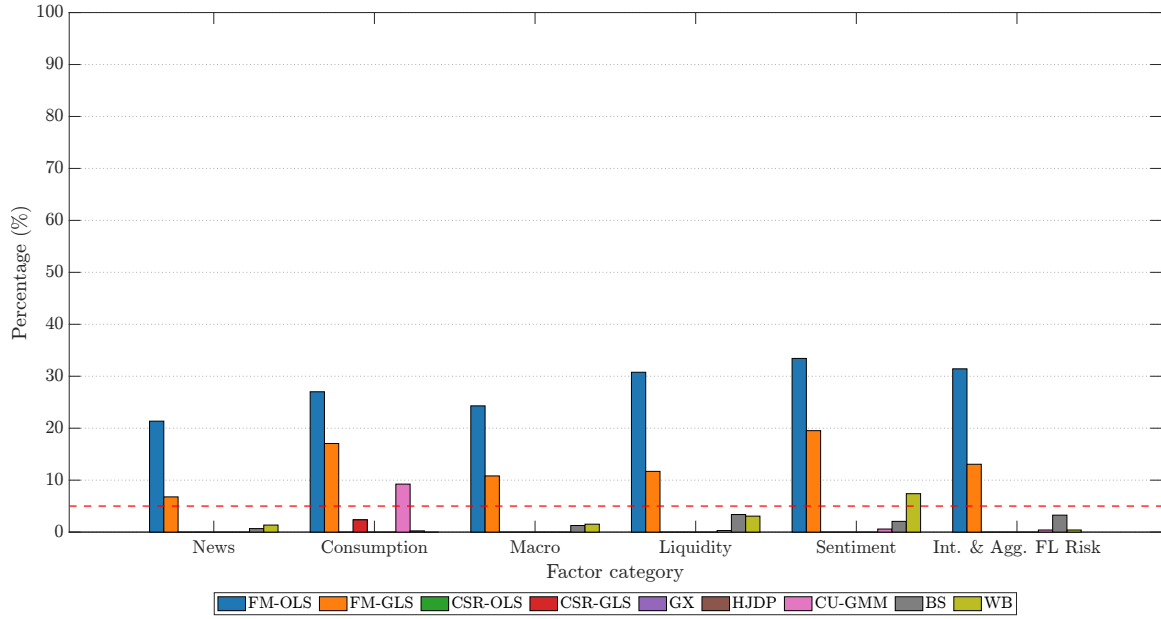


Figure IA.4: The percentages of priced factors by factor category controlling for multiple hypothesis testing

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

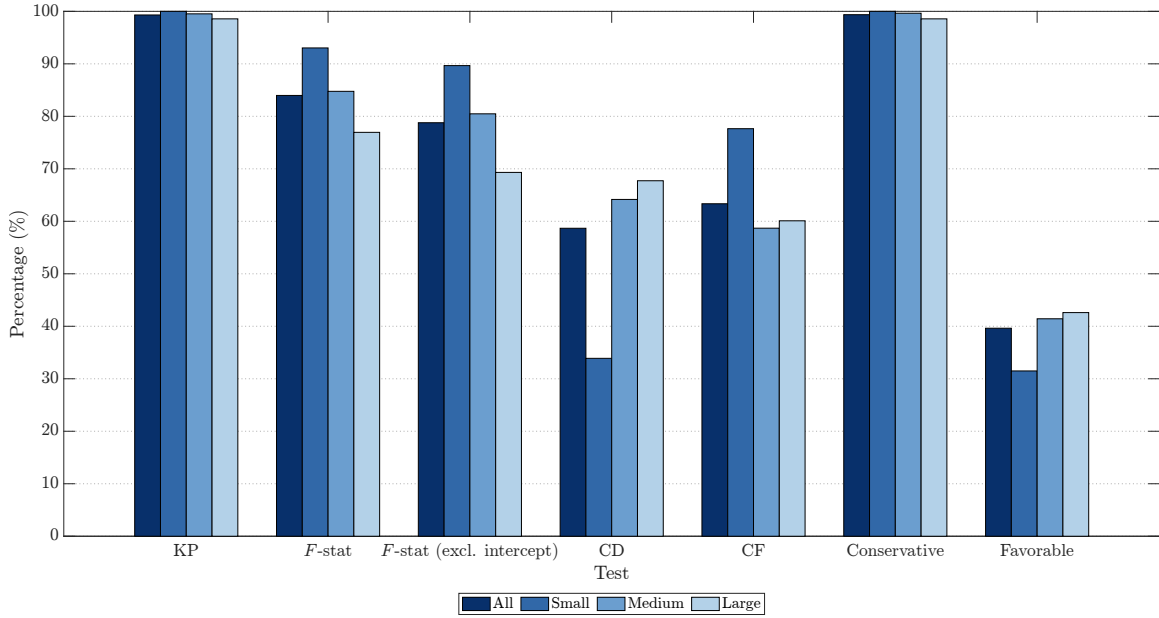


Figure IA.5: Identification failure of original model specifications after multiple-testing adjustment.

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 ($T > 400$).

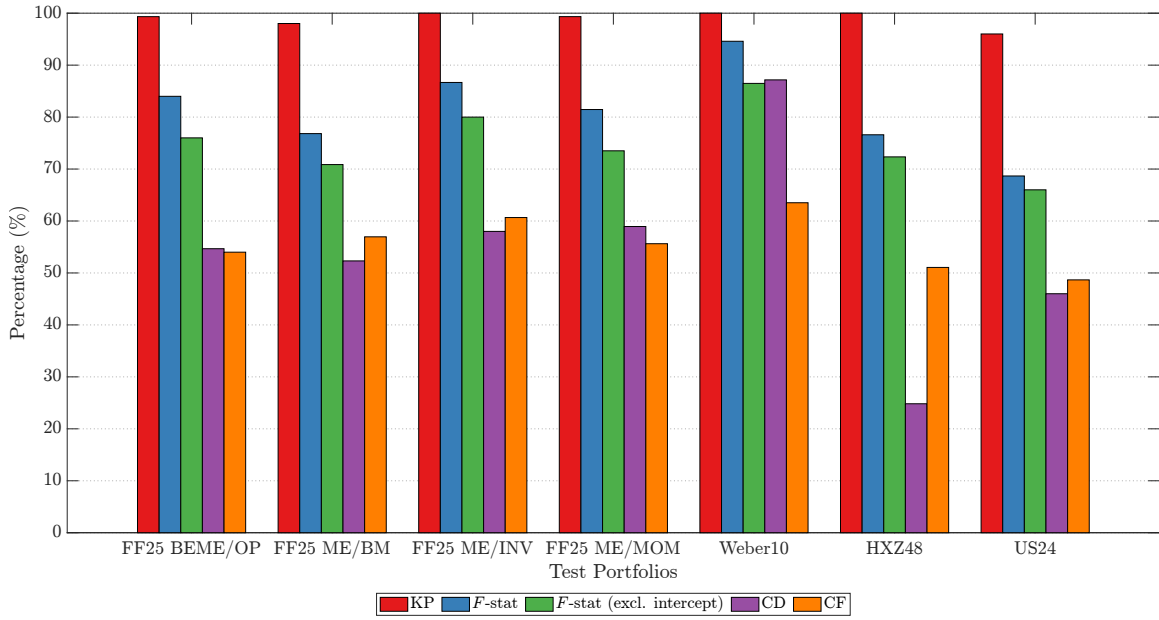


Figure IA.6: Identification failure of original model specifications with equity portfolios after multiple-testing adjustment.

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 \(2019\)](#).

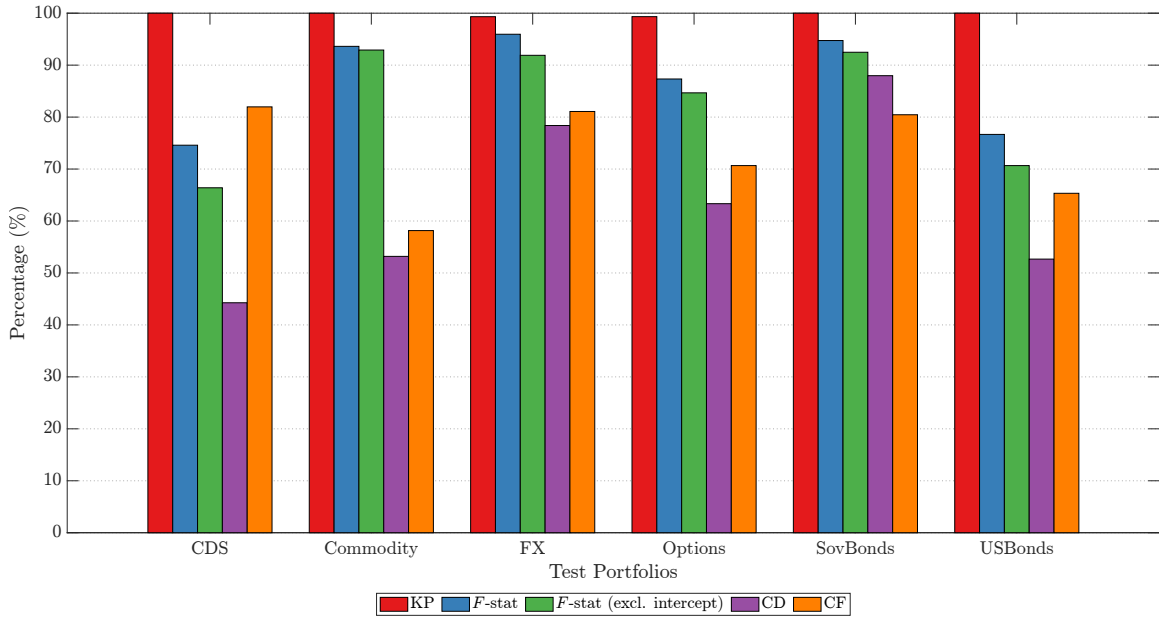


Figure IA.7: Identification failure of original model specifications with non-equity portfolios after multiple-testing adjustment.

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 \(2019\)](#).

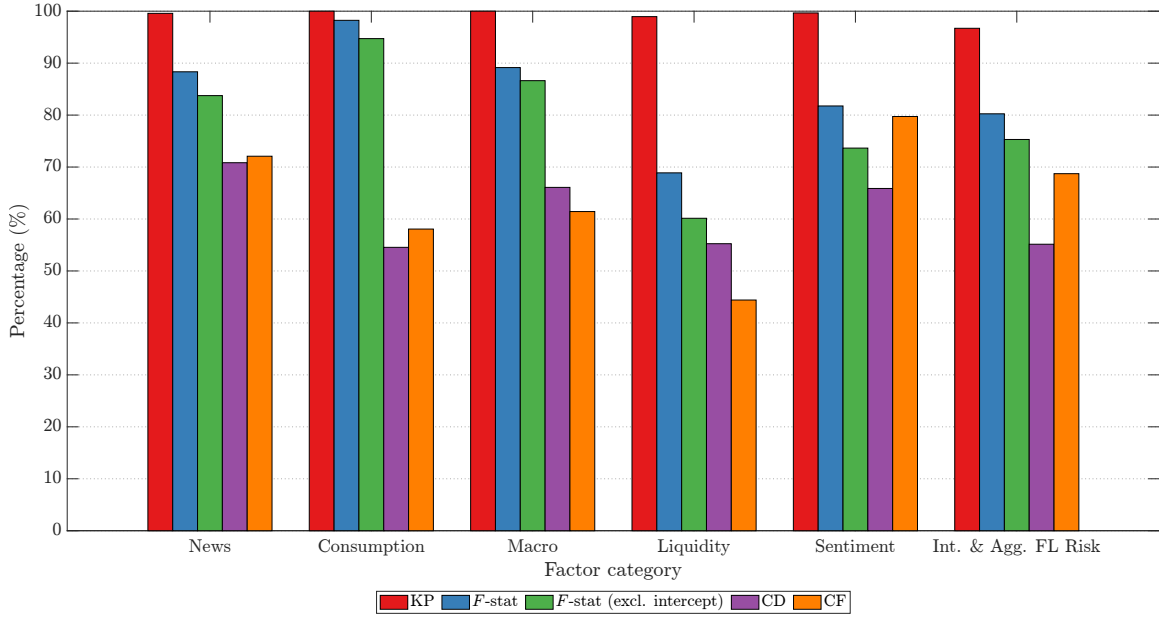


Figure IA.8: Identification failure of original model specifications by theme after multiple-testing adjustment.

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.