UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Identification Robust Testing of Risk Premia in Finite Samples

Kleibergen, F.; Kong, L.; Zhan, Z.

DOI
10.1093/jjfinec/nbac010
Publication date
2023
Document Version
Final published version
Published in
Journal of Financial Econometrics

License
CC BY

Link to publication

Citation for published version (APA):
Kleibergen, F., Kong, L., & Zhan, Z. (2023). Identification Robust Testing of Risk Premia in Finite Samples. *Journal of Financial Econometrics*, *21*(2), 263–297. https://doi.org/10.1093/jjfinec/nbac010

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:15 Nov 2026

Journal of Financial Econometrics, 2023, Vol. 21, No. 2, 263–297
https://doi.org/10.1093/jjfinec/nbac010
Advance Access Publication Date: 28 April 2022
Halbert White Jr. Memorial JFEC Invited Paper

OXFORD

Identification Robust Testing of Risk Premia in Finite Samples

Frank Kleibergen [1], Lingwei Kong[2] and Zhaoguo Zhan[3]

[1]Amsterdam School of Economics, University of Amsterdam, The Netherlands, [2]University of Groningen, The Netherlands, and [3]Coles College of Business, Kennesaw State University, USA

Address correspondence to Frank Kleibergen, Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, Amsterdam 1018 WB, The Netherlands, or e-mail: f.r.kleibergen@uva.nl.

Received March 8, 2022; revised March 14, 2022; editorial decision March 14, 2022; accepted March 14, 2022

Abstract

The reliability of tests on the risk premia in linear factor models is threatened by limited sample sizes and weak identification of risk premia frequently encountered in applied work. We, therefore, propose novel tests on the risk premia that are robust to both limited sample sizes and the identification strength of the risk premia as reflected by the quality of the risk factors. These tests are appealing for empirically relevant settings, and lead to confidence sets of risk premia that can substantially differ from conventional ones. To show the latter, we revisit two high-profile empirical applications.

Key words: asset pricing, finite samples, identification robust inference, risk premia

JEL classification: G12

Over the past decades, it has become common practice to assess risk premia using the classic Fama–MacBeth (FM) two-pass approach, see for example, Fama and MacBeth (1973) and Shanken (1992). It is based on the, so-called, linear beta representation of expected asset returns (Cochrane 2001) and is one of the default approaches to evaluate asset pricing models in the financial economics literature, see for example, Fama and French (1992). The large variety of risk factors that this literature has proposed (see e.g., Lettau and Ludvigson 2001; Kroencke 2017) has, however, led to a growing concern that the reliability of the two-pass approach is at risk for many of them. This concern results from two common issues that threaten the two-pass approach: (i) the absence of large correlations between risk factors and asset returns needed to identify the risk premia and (ii) the rather small number of time-series observations relative to the number of assets.

A strand of the literature shows a failure in identifying risk premia when the quality of risk factors specified for the two-pass approach is unsatisfactory. Kan and Zhang (1999), for instance, show that if risk factors are useless (so betas are zero), their risk premia are

© The Author(s) 2022. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

unidentified and the commonly used FM t-statistic on risk premia tends to be spuriously significant. Many observed risk factors are, however, not completely useless but just minorly correlated with asset returns. For these factors, Kleibergen (2009) similarly warns that their risk premia are only weakly identified, which also makes the FM t-statistic unreliable. These findings lead to the emergence of the so-called identification robust tests on risk premia, which, unlike the FM t-test, remain trustworthy regardless of the quality of risk factors. In the ideal scenario where risk factors of good quality are used for the FM two-pass approach, identification robust tests and the FM t-test both provide valid and comparable inference on risk premia. In contrast, in empirical studies where the quality of risk factors is questionable, identification robust tests are more trustworthy than the FM t-test. See, for example, Kleibergen (2009), Beaulieu, Dufour and Khalaf (2013), Khalaf and Schaller (2016), and Kleibergen and Zhan (2018).

Another strand of literature is concerned with the challenges that result from the limited sample sizes in empirical studies. Macroeconomic risk factors, in particular, are commonly measured at quarterly or annual frequencies, so their number of time-series observations T is usually not much larger than the number of test assets N. Take Kroencke (2017), for example, the yearly consumption growth series from 1960 to 2014 has 55 time-series observations, $T = 55$, while 30 portfolios, sorted by size, value, and investment alongside the market portfolio, are used as test assets, so $N = 31$. This limited T versus large N setting jeopardizes asymptotic tests on the risk premia, whose validity typically relies on a much larger value of T compared with N (Newey and Windmeijer 2009). To conduct valid inference for limited sample sizes, researchers have, therefore, developed various finite-sample robust methods. Gibbons, Ross, and Shanken (GRS 1989), for example, propose the GRS test for portfolio efficiency, which circumvents the limit sample size problem by making distributional assumptions. Similar assumptions are made in Kan and Zhou (2004) for deriving the exact distribution of the Hansen–Jagannathan statistic. More recently, Kleibergen and Zhan (2020) extend the GRS test to develop a finite-sample robust test on the risk premia.

Both strands of literature above are empirically relevant, since sample sizes in empirical studies are often not sufficiently large to validate asymptotic tests, such as the FM t-test, while it has been well documented that a large number of risk factors are of questionable quality (Kleibergen and Zhan 2015). Despite the fact that the above two strands of literature are well established, tests on risk premia that account for both the limited sample size and the questionable quality of risk factors are lacking.

In light of the above, we propose novel tests on the risk premia, which are designed to be robust to both the sample size and the quality of risk factors. We do so by providing the exact finite-sample distributions of several identification robust test statistics, including those in, for example, Kleibergen (2009) and Beaulieu, Dufour, and Khalaf (2013). Put differently, while identification robust tests are asymptotically valid regardless of the quality of risk factors, we extend these tests to finite-sample settings by constructing finite-sample distributions of their test statistics. In line with Gibbons, Ross, and Shanken (1989), Kleibergen and Zhan (2020), and many others, these finite-sample distributions result from the (normality) assumptions of the error term in a linear factor model for asset returns.

We show that our proposed tests on risk premia work favorably in a simulation experiment calibrated to real data. As expected, these tests are found particularly appealing in simulation settings where the time-series dimension T is close to the cross-section dimension N. In such settings, the proposed tests lead to correct sizes for testing hypotheses on

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

risk premia, while their asymptotic counterparts, that is, the asymptotically valid tests, can suffer from size distortion. On the other hand, when T is much larger than N, the proposed tests become identical to the asymptotically valid identification robust tests, all of which are trustworthy regardless of the quality of risk factors. These numerical findings for the proposed tests are all in line with our theoretical results.

To illustrate the practical relevance of our proposed tests, we revisit the empirical studies on the consumption capital asset pricing model (CAPM) and the conditional consumption CAPM in Kroencke (2017) and Lettau and Ludvigson (2001), respectively.[1] Kroencke (2017) focuses on the classic consumption CAPM, where consumption growth is the single risk factor. The involved data set has $T = 55$ while $N = 31$ as aforementioned. Instead of the single factor model, Lettau and Ludvigson (2001) propose a multi-factor conditional consumption CAPM. We use their quarterly data from 1963Q3 to 1998Q3 with $T = 141$ and $N = 25$ portfolios sorted by size and book-to-market as test assets. In both applications, we find that inverting the proposed tests leads to 95% confidence sets of risk premia that are substantially different from those based on existing asymptotic tests. This should not be surprising, since both in Lettau and Ludvigson (2001) and Kroencke (2017), the limited T relative to N is unlikely to validate asymptotic tests. Moreover, the 95% confidence sets resulting from the proposed tests are found to be unbounded, reflecting that little information on risk premia is contained in the studied data sets. These empirical results thus cast doubt on the pricing performance of the proposed risk factors in Lettau and Ludvigson (2001) and Kroencke (2017).

Overall, our findings indicate that robustness to the quality of risk factors and robustness to the sample size should both be accounted for in asset pricing studies. Traditional asset pricing tests that are not robust to the strength of identification reflected by the quality of risk factors can lead to erroneous conclusions. Similarly, asymptotically valid tests that do not account for limited sample sizes can also induce spurious empirical findings. We simultaneously address these two empirically relevant issues.

The rest of the paper is organized as follows. Section 1 discusses the two issues that threaten reliable inference on risk premia: lack of identification due to poor quality risk factors and the limited sample size. It also introduces the two data sets from Kroencke (2017) and Lettau and Ludvigson (2001) that we use to emphasize the empirical relevance of our results. Section 2 presents the proposed test statistics and constructs their finite-sample distributions. Section 3 contains a simulation study as well as applications using the data from Lettau and Ludvigson (2001) and Kroencke (2017). Section 4 concludes the paper. Technical details are relegated to the Appendix.

1 Preliminaries

1.1 Linear Beta Representation and the Two-Pass Approach

Premia on risk factors are identified by the linear beta representation of expected asset returns (Cochrane 2001):

$$\mathbb{E}(R_t) = \iota_N \lambda_0 + \beta \lambda_F = (\iota_N, \beta) \begin{pmatrix} \lambda_0 \\ \lambda_F \end{pmatrix}, \tag{1}$$

1 We thank these authors for providing their data.

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

with R_t an $N \times 1$ vector of asset returns, ι_N the $N \times 1$ vector of ones, β an $N \times K$ matrix of factor loadings, λ_0 the zero-beta return, and λ_F the $K \times 1$ vector of risk premia.

The β matrix of factor loadings in Equation (1) is related to a linear factor model for asset returns:

$$R_t = c + \beta F_t + u_t, \quad t = 1, \ldots, T, \tag{2}$$

with F_t a $K \times 1$ vector of risk factors, c an $N \times 1$ vector of constant terms, and u_t an $N \times 1$ vector of errors.

A large portion of the asset pricing literature is centered around inference on the risk premia λ_F based on Equations (1) and (2). The conventional two-pass procedure estimates β in the first pass using the time-series regression in Equation (2), so

$$\widehat{\beta} = \left(\sum_{t=1}^{T} \overline{R}_t \overline{F}_t'\right)\left(\sum_{t=1}^{T} \overline{F}_t \overline{F}_t'\right)^{-1}, \tag{3}$$

with $\overline{R}_t = R_t - \overline{R}, \overline{R} = \frac{1}{T}\sum_{t=1}^{T} R_t, \overline{F}_t = F_t - \overline{F}, \overline{F} = \frac{1}{T}\sum_{t=1}^{T} F_t$; in the second pass, the risk premia are estimated using a cross-sectional regression of the average asset returns $\overline{R}$ on ι_N and $\widehat{\beta}$ estimated in the first pass:

$$\begin{pmatrix} \widehat{\lambda}_0 \\ \widehat{\lambda}_F \end{pmatrix} = \left[\left(\iota_N, \widehat{\beta}\right)'\left(\iota_N, \widehat{\beta}\right)\right]^{-1}\left(\iota_N, \widehat{\beta}\right)'\overline{R}. \tag{4}$$

1.2 Lack of Identification and Identification Robust Tests

The (ι_N, β) matrix plays a crucial role for inference on risk premia (Kleibergen and Zhan 2020). If (ι_N, β) has a full rank value, then λ_F is well defined by Equation (1) and its estimator $\widehat{\lambda}_F$ is consistent and asymptotically normally distributed under standard regularity conditions (Shanken 1992). It validates the common usage of the t-test on the risk premia. The full rank condition of (ι_N, β) is, however, plausibly violated in many empirical settings.

The pioneering study of Kan and Zhang (1999), for example, focuses on the useless factor setting with $\beta = 0$. If $\beta = 0$, then λ_F is unidentified in Equation (1), and Kan and Zhang (1999) show that the FM t-test on risk premia becomes unreliable, that is, its test statistic tends to be significant, which thus spuriously supports useless risk factors. Kleibergen (2009) further considers a so-called weak factor setting with $\beta \approx 0$, since many risk factors proposed in the literature exhibit just minor correlations with asset returns. Similar to the useless factor setting, Kleibergen (2009) finds that risk premia on weak factors are not well identified, which further induces a malfunction of the FM t-test. In addition, Kleibergen and Zhan (2020) show that β's are often close to be constant over the different assets so there is little cross-sectional variation in them. In this scenario, Kleibergen and Zhan (2020) warn that identification of the risk premia is then similarly jeopardized, so the FM t-test also breaks down. All these studies boil down to the full rank condition of (ι_N, β), which is jeopardized if $\beta = 0$, $\beta \approx 0$, or $\beta \propto \iota_N$.

To provide empirical evidence that the full rank condition is plausibly violated, we use the data from Kroencke (2017) and Lettau and Ludvigson (2001). The resulting estimates of β are presented in Tables 1 and 2, respectively.

Re-produced from Kleibergen and Zhan (2020), Table 1 presents the estimated βs using the 31 test assets from Kroencke (2017) and 5 different consumption growth measures

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

Table 1 β with 31 portfolio returns

	Reported		P-J		Q4-Q4		Garbage		Unfiltered	
	$\widehat{\beta}$	t-stat	$\widehat{\beta}$	t-stat	$\widehat{\beta}$	t-stat	$\widehat{\beta}$	t-stat	$\widehat{\beta}$	t-stat
(1)	−1.30	−0.41	1.19	0.93	4.18	1.51	4.54	3.33	5.40	3.81
(2)	−0.23	−0.09	1.14	1.06	4.42	1.93	4.06	3.52	4.70	3.98
(3)	0.14	0.06	0.99	1.06	4.15	2.09	3.74	3.86	3.80	3.64
(4)	−0.52	−0.25	0.73	0.86	3.46	1.92	3.66	4.30	3.33	3.50
(5)	−0.23	−0.11	0.73	0.85	3.74	2.06	3.62	4.30	3.33	3.44
(6)	−0.84	−0.43	0.58	0.72	3.01	1.75	3.42	4.39	3.10	3.43
(7)	−0.85	−0.44	0.49	0.63	2.73	1.63	3.51	4.90	2.95	3.32
(8)	−0.59	−0.33	0.39	0.53	2.69	1.69	3.06	4.55	2.85	3.41
(9)	−0.12	−0.07	0.30	0.43	2.75	1.87	2.61	3.98	2.48	3.16
(10)	0.34	0.20	0.64	0.92	2.97	2.01	3.14	4.53	2.60	3.29
(11)	−0.95	−0.46	0.21	0.25	2.25	1.22	3.58	4.29	2.85	2.91
(12)	−0.44	−0.25	0.60	0.81	2.77	1.77	3.15	4.22	2.59	3.09
(13)	−0.63	−0.38	0.52	0.78	2.49	1.73	2.95	4.38	2.22	2.84
(14)	−0.56	−0.31	0.07	0.10	2.26	1.39	2.51	3.10	2.04	2.28
(15)	0.88	0.53	0.46	0.67	3.14	2.19	2.75	3.89	2.25	2.85
(16)	0.73	0.45	0.58	0.88	3.04	2.19	2.86	4.49	2.62	3.56
(17)	0.96	0.56	0.71	1.03	3.50	2.40	2.93	4.53	2.81	3.61
(18)	1.21	0.65	1.05	1.40	4.27	2.72	3.17	4.35	3.33	4.02
(19)	1.23	0.67	1.09	1.47	3.97	2.56	2.86	3.65	2.96	3.53
(20)	1.77	0.84	1.32	1.56	5.22	2.98	2.93	3.26	4.21	4.67
(21)	−0.50	−0.22	0.58	0.61	3.57	1.78	3.85	4.30	3.88	3.73
(22)	−0.09	−0.05	0.66	0.88	3.14	1.98	3.00	4.10	3.15	3.82
(23)	0.37	0.23	0.54	0.81	2.84	2.02	2.94	4.49	2.27	2.96
(24)	0.44	0.31	0.54	0.92	2.96	2.40	2.84	5.30	2.25	3.38
(25)	−0.50	−0.32	0.31	0.49	2.11	1.57	2.69	5.05	2.22	3.10
(26)	1.04	0.62	0.79	1.16	3.70	2.60	3.09	4.44	2.54	3.25
(27)	1.10	0.66	0.98	1.46	3.71	2.63	2.90	4.13	2.74	3.60
(28)	0.03	0.01	0.71	0.98	3.18	2.07	3.15	4.32	3.07	3.83
(29)	−0.89	−0.41	0.17	0.19	2.44	1.26	3.54	3.82	2.70	2.59
(30)	−1.40	−0.58	0.03	0.03	2.10	0.97	3.92	3.95	2.92	2.52
(31)	0.23	0.13	0.72	1.00	3.32	2.21	3.21	4.74	2.95	3.72

Notes: The 31 test assets are taken from Kroencke (2017), that is, the 30 portfolios sorted by size, value, and investment, plus the market portfolio in 1960–2014 (yearly data). "Reported," "P-J," "Q4-Q4," "Garbage" and "Unfiltered" correspond to five consumption measures: (i) the consumption expenditure on nondurable goods and services reported by National Income and Product Accounts (NIPA) ("Reported"); (ii) the 3-year consumption measure from Parker and Julliard (2005) ("P-J"); (iii) the fourth-quarter to fourth-quarter consumption measure from Jagannathan and Wang (2007) ("Q4-Q4"); (iv) the garbage measure in Savov (2011) ("Garbage"); and (v) the unfiltered NIPA consumption measure in Kroencke (2017) ("Unfiltered"). Each of the five consumption growth rates is used as the single risk factor for asset returns to estimate β.

("Reported," "P-J," "Q4-Q4," "Garbage," and "Unfiltered," see the notes to Table 1) from the asset pricing literature. Kroencke (2017) proposes "Unfiltered" consumption growth as a risk factor, which appears to outperform the other four consumption growth measures for consumption-based asset pricing.

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

Table 2 β with 25 portfolio returns

	$\triangle c$		cay		$\triangle c\times$ cay	
	$\widehat{\beta}_1$	t-stat	$\widehat{\beta}_2$	t-stat	$\widehat{\beta}_3$	t-stat
(1)	6.35	2.32	4.25	2.96	−4.56	−0.22
(2)	6.30	2.61	3.66	2.90	2.25	0.12
(3)	5.11	2.28	3.22	2.75	3.22	0.19
(4)	5.50	2.58	2.96	2.65	5.12	0.32
(5)	5.79	2.53	2.65	2.21	11.36	0.65
(6)	4.38	1.76	4.55	3.49	−15.82	−0.84
(7)	3.63	1.65	3.25	2.83	0.02	0.00
(8)	3.92	1.97	3.07	2.95	−1.08	−0.07
(9)	3.52	1.91	2.66	2.76	6.93	0.50
(10)	4.83	2.41	2.16	2.06	8.48	0.56
(11)	2.70	1.21	4.53	3.88	−22.20	−1.31
(12)	2.76	1.47	3.64	3.69	−3.35	−0.23
(13)	2.92	1.69	2.63	2.91	4.40	0.34
(14)	2.58	1.57	2.73	3.18	0.02	0.00
(15)	3.71	1.98	2.15	2.19	4.78	0.34
(16)	1.97	1.02	4.22	4.20	−20.28	−1.39
(17)	2.62	1.49	3.28	3.58	−10.24	−0.77
(18)	1.94	1.21	2.57	3.06	−4.98	−0.41
(19)	2.50	1.56	2.32	2.76	0.60	0.05
(20)	3.78	2.05	2.26	2.34	3.28	0.23
(21)	1.61	1.01	2.77	3.33	−21.22	−1.76
(22)	1.16	0.80	2.53	3.33	−4.20	−0.38
(23)	2.32	1.88	2.46	3.80	−8.27	−0.88
(24)	1.24	0.94	2.24	3.25	−10.49	−1.05
(25)	3.07	2.13	1.59	2.11	−1.25	−0.11

Notes: The 25 portfolios sorted by size and book-to-market in 1963Q3–1998Q3 (quarterly data) are used as test assets. The three risk factors taken from Lettau and Ludvigson (2001) are the quarterly consumption growth $\triangle c$, the (lagged) consumption-wealth ratio cay, and their interaction $\triangle c \times$ cay. The 25×3 dimensional β matrix is thus $\beta = (\beta_1, \beta_2, \beta_3)$, where β_i, $i = 1, 2, 3$, corresponds to the ith factor.

Table 1, however, suggests that the βs for all five consumption growth measures are problematic. For example, the estimated β-vector for "Reported" consumption growth is not statistically different from zero. On the other hand, although the βs for the "Garbage" and "Unfiltered" consumption growth measures all appear sizeable, they exhibit little cross-sectional variation so they are almost constant and proportional to a vector of ones. Kleibergen and Zhan (2020) further conduct a rank test on (ι_N, β) and find evidence that its full rank condition is likely violated under each of the five consumption growth measures. The reported estimates of the βs in Table 1 thus cast doubt on the empirical findings in Kroencke (2017).

Similar to Table 1, Table 2 reports the estimated βs using the 25 portfolios sorted by size and book-to-market for three risk factors (the quarterly consumption growth $\triangle c$, the lagged consumption-wealth ratio cay, and their interaction $\triangle c\times$ cay) taken from Lettau

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

and Ludvigson (2001). It is rather obvious in Table 2 that the last column of the β-matrix, which corresponds to the $\triangle c\times$ cay risk factor, is not statistically significant at the 5% level. We, therefore, cannot rule out that a column of the β-matrix for this multi-factor model equals zero, so the (ι_N, β) matrix does not have a full rank value. This is also consistent with the rank test results reported in Kleibergen and Zhan (2020). Consequently, Table 2 raises doubts on the credibility of the empirical findings based on the FM t-test in Lettau and Ludvigson (2001).

The βs of questionable quality, such as those reported in Tables 1 and 2, have motivated Kan and Zhang (1999), Kleibergen (2009), and thereafter Kleibergen and Zhan (2015, 2018, 2020), to analyze their impact on the risk premia. It is now well acknowledged that inference on the risk premia is sensitive to the strength of identification, as reflected by the full rank condition of (ι_N, β) or put differently, the quality of the risk factors that yield the factor loadings in β. This has led to the so-called identification robust tests on the risk premia, whose validity does not depend on the rank of (ι_N, β), that is, these tests remain trustworthy regardless of the quality of the risk factors. An example of such a test is the factor Anderson–Rubin (FAR) test proposed by Kleibergen (2009), which is an extension of the robust Anderson and Rubin (1949) test in the instrumental variables regression literature; see also the Hotelling (H) type test in Beaulieu, Dufour, and Khalaf (2013). Other examples of identification robust tests include the Lagrange multiplier (LM) and likelihood ratio (LR) tests.

One potential drawback of these identification robust tests is, however, that many of them are established using asymptotic arguments, so the validity of these tests often relies on a large sample size, see, for example, the FAR test in Kleibergen (2009). It is thus questionable whether these tests perform well in finite-sample applications, which we discuss next.

1.3 Limited T versus Large N and Finite-Sample Robust Tests

Other than lack of identification due to the quality of risk factors, limited sample sizes also impose challenges for inference on risk premia. The latter should come out naturally, since validity of asymptotic tests, such as the FM t-test and the FAR test, requires a large number of time-series observations T relative to the number of assets N. If T is small while N is large, it becomes doubtful whether we can still approximate the distributions of the FM t-statistic and the FAR statistic by normal and chi-squared distributions, respectively. The large number of assets N or in other words, many moment conditions, resembles the many instruments issue in instrumental variables regression (Bekker 1994; Bekker and Kleibergen 2003; Newey and Windmeijer 2009).

The limited T versus large N scenario is commonly encountered in empirical asset pricing studies. For instance, Savov (2011) uses the annual garbage growth series from 1960 to 2006, while the 25 Fama-French portfolios augmented by the 10 industry portfolios, as suggested by Lewellen, Nagel and Shanken (2010), are used as test assets. The resulting limited $T = 47$ and relatively large $N = 35$ thus cast doubt on the FM t-test used in Savov (2011). Similarly, the Kroencke (2017) data used for Table 1 use the yearly unfiltered consumption growth series for the time period 1960–2014, so $T = 55$, and 30 portfolios sorted by size, value, and investment alongside the market portfolio are used as test assets, so $N = 31$. The simulation experiment in Kleibergen and Zhan (2020), however, shows that neither the FM

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

t-test nor the FAR test is size-correct under the limited T versus large N setting. Given that many risk factors, like the garbage growth in Savov (2011) and unfiltered consumption growth in Kroencke (2017), are only available at low frequencies and/or within short time periods, the limited T versus large N problem similarly affects many other empirical studies.

In view of the above, we aim to develop tests on risk premia that are immune to the limited T versus large N problem and we label these tests as finite-sample robust tests. Meanwhile, we also want these tests to remain trustworthy when the quality of risk factors is potentially unsatisfactory. We therefore have a two-fold goal, that is, providing tests that are reliable when using risk factors of questionable quality as well as for limited sample sizes.

Kleibergen and Zhan (2020) develop two tests to meet this two-fold goal. First, they construct a finite-sample version of the rank test from Kleibergen and Paap (2006) so it can be used to test for a reduced rank value of (ι_N, β), implicating that the risk premia are not identified, when N is large compared with T. Second, they extend the GRS test statistic from Gibbons, Ross, and Shanken (1989) so that it can be used for testing hypotheses on the risk premia. They refer to it as the GRS-FAR statistic, to reflect that it also results from the FAR statistic in Kleibergen (2009) with an adjustment for the finite-sample setting. As in Gibbons, Ross, and Shanken (1989), Kleibergen and Zhan (2020) also assume that the error term in the linear factor model (2) is normally distributed, in order to derive the finite-sample F-distribution of the GRS-FAR statistic. Since this F-distribution does not depend on the quality of the risk factors, the finite-sample robust GRS-FAR test is also robust to the strength of identification.

In the next section, we propose several other tests, which, like the GRS-FAR test, are robust to both the sample size and the strength of identification reflected by the quality of the risk factors. Later on, we also show that these tests have comparable or even better power than the GRS-FAR test. To facilitate comparison and understanding, we also discuss the GRS-FAR test.

2 Analytical Results

2.1 Null Hypothesis and Assumptions

Our interest lies in testing the hypothesis on the risk premia: $\mathrm{H}_0 : \lambda_F = \lambda_{F,0}$, with $\lambda_{F,0}$ the hypothesized value of λ_F. We, therefore, delete λ_0 in Equation (1) by removing the Nth asset and taking all other asset returns in deviation from the return on the Nth asset. The moment condition (1) and the linear factor model (2) now become:

$$\mathbb{E}(\mathcal{R}_t) = \mathcal{B}\lambda_F \tag{5}$$

$$\mathcal{R}_t = \mathcal{C} + \mathcal{B}F_t + \mathcal{U}_t, \tag{6}$$

where $\mathcal{R}_t = J_N R_t$, $\mathcal{C} = J_N c$, $\mathcal{B} = J_N \beta$, $\mathcal{U}_t = J_N u_t$, with $J_N = (I_{N-1}, -\iota_{N-1})$, I_{N-1} is the $(N-1) \times (N-1)$ identity matrix, and ι_{N-1} is the $(N-1)$-dimensional vector of ones.

Remark 1. *Our proposed test statistics are invariant to which asset return is to be subtracted. Put differently, the choice of the Nth asset does not affect the resulting values of our statistics. See the Online Appendix of Kleibergen and Zhan (2020).*

Remark 2. *In tests with excess returns, $\lambda_0 = 0$ is sometimes imposed. See*, for example, *Savov (2011). With this zero-restriction, we do not have to remove the Nth asset, and*

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

Equations (5) and (6) are almost identical to Equations (1) and (2) except for notation. Our tests can thus be extended straightforwardly to incorporate the $\lambda_0 = 0$ *restriction.*

Remark 3. *The full rank condition of* (ι_N, β)*, which we discussed previously for identifying risk premia, is equivalent to the full rank condition of* $\mathcal{B}$*. Since we allow for risk factors of questionable quality, we do not need to assume that* $\mathcal{B}$ *has a full rank value.*

We make the following two assumptions for the linear factor model, both of which have been used in the previous literature. Further explanations are provided under each assumption.

Assumption 1. $\mathcal{U}_t$ *is i.i.d. normally distributed with covariance* Σ.

The normality assumption gives rise to the finite-sample distributions of the test statistics proposed later on. This assumption is commonly adopted in the existing literature to facilitate finite-sample analysis, see, for example, Gibbons, Ross, and Shanken (1989) and Kan and Zhou (2004). We can as well replace normality with other distributional assumptions, and if we do so, the finite-sample distributions of our proposed statistics can be derived in a similar fashion or using simulation.

Assumption 2. *Let* F *be the* $T \times K$ *matrix whose* t*-th row is* $\overline{F}_t'$ *and we assume that* (ι_T, F) *is a deterministic matrix of full rank.*

The full rank condition of (ι_T, F), where ι_T is the T-dimensional vector of ones, rules out multi-collinearity of risk factors, so β and thus $\mathcal{B}$ are estimable using time-series regression. Here, we assume that risk factors are fixed, instead of imposing restrictions on their limiting behaviors, in order to facilitate the finite-sample analysis of test statistics. The assumption of fixed factors can be similarly found in, for example, Gibbons, Ross, and Shanken (1989). We could allow for randomness of risk factors by regarding the proposed test statistics as conditional random variables given the risk factors.

2.2 Test Statistics and Their Finite-Sample Distributions

Under the null hypothesis $H_0 : \lambda_F = \lambda_{F,0}$ and the moment condition (5), we have $\mathbb{E}(\mathcal{R}_t) = \mathcal{B}\lambda_{F,0}$, so the average returns $\overline{\mathcal{R}} = \frac{1}{T}\sum_{t=1}^{T} \mathcal{R}_t$ should be centered around $\mathcal{B}\lambda_{F,0}$. In order to test $H_0 : \lambda_F = \lambda_{F,0}$, we can examine whether $\overline{\mathcal{R}} - \mathcal{B}\lambda_{F,0}$ is centered around zero. This motivates the test statistics we present next.

Put differently, our proposed test statistics do not rely on an estimator of λ_F. Instead, we just examine whether the factor pricing condition (5) holds at the hypothesized risk premia $\lambda_{F,0}$. A rejection of $H_0 : \lambda_F = \lambda_{F,0}$ could result if Equation (5) holds at a value other than $\lambda_{F,0}$ or if Equation (5) does not hold at any value of the risk premia, that is, there exists model mis-specification.[2] Since our test statistics do not involve risk premia estimation,

2 If $H_0 : \lambda_F = \lambda_{F,0}$ is rejected for every hypothesized $\lambda_{F,0}$, then it signals that the adopted risk factors are mis-specified for the factor pricing condition (5). In this case, mis-specification robust inference methods can be adopted; see, for example, Kan, Robotti, and Shanken (2013) and Kleibergen and Zhan (2021).

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

they remain trustworthy regardless of the strength of risk premia identification, as reflected by the quality of risk factors or $\mathcal{B}$.

For later use, we distinguish the following two estimators for $\mathcal{B}$: a restricted least squares estimator under the null,

$$\widetilde{\mathcal{B}} = \left(\sum_{t=1}^{T} \mathcal{R}_t \left(\overline{F}_t + \lambda_{F,0}\right)'\right)\left(\sum_{t=1}^{T} \left(\overline{F}_t + \lambda_{F,0}\right)\left(\overline{F}_t + \lambda_{F,0}\right)'\right)^{-1},$$

and an unrestricted least squares estimator,

$$\widehat{\mathcal{B}} = \left(\sum_{t=1}^{T} \overline{\mathcal{R}}_t \overline{F}_t'\right)\left(\sum_{t=1}^{T} \overline{F}_t \overline{F}_t'\right)^{-1},$$

where $\overline{\mathcal{R}}_t = \mathcal{R}_t - \overline{\mathcal{R}}$, $\overline{F}_t$ is the demeaned factor similarly defined under Equation (3).

Since $\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}$ and $\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}$ are natural choices to help examine whether $\mathbb{E}(\mathcal{R}_t) = \mathcal{B}\lambda_{F,0}$ holds, we use them to construct test statistics. Under $H_0 : \lambda_F = \lambda_{F,0}$ imposed on factor pricing and Assumptions 1 and 2, $\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}$ and $\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}$ are both normally distributed $(N-1)$-dimensional vectors and the distribution of one of these two vectors is just a scaled version of the distribution of the other[3]:

$$\sqrt{T}\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right) \sim N\left(0, \left(1 - \lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}(\lambda_{F,0})^{-1}\lambda_{F,0}\right) \otimes \Sigma\right), \tag{7}$$

$$\sqrt{T}\left(\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}\right) \sim N\left(0, \left(1 + \lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}\right) \otimes \Sigma\right), \tag{8}$$

where $\widehat{\mathcal{Q}}_{FF}(\lambda_{F,0}) = \frac{1}{T}\sum_{t=1}^{T}\left(\overline{F}_t + \lambda_{F,0}\right)\left(\overline{F}_t + \lambda_{F,0}\right)'$, $\widehat{\mathcal{Q}} = \frac{1}{T}\sum_{t=1}^{T}\overline{F}_t\overline{F}_t'$, and Σ is the covariance of the error $\mathcal{U}_t$ in Equation (6).

2.2.1 Far and *H* statistics

The FAR statistic in Kleibergen (2009) is based on the normal approximation of the $\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}$ vector. Similarly, the H type test statistic (see Beaulieu, Dufour, and Khalaf 2013) is based on the normal approximation of $\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}$. Both statistics can then be viewed as the standardized lengths of these vectors, which explain their χ^2 limiting distributions:

$$\mathrm{FAR}(\lambda_{F,0}) = \frac{T}{1 - \lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\lambda_{F,0}}\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)'\widehat{\Sigma}^{-1}\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right) \tag{9}$$

$$H(\lambda_{F,0}) = \frac{T}{1 + \lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}}\left(\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}\right)'\widehat{\Sigma}^{-1}\left(\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}\right) \tag{10}$$

with $\widehat{\Sigma}$ a consistent estimator for the covariance matrix Σ. Unless specified otherwise, we use

$$\widehat{\Sigma} = \frac{1}{T - K - 1}\sum_{t=1}^{T}\left(\overline{\mathcal{R}}_t - \widehat{\mathcal{B}}\overline{F}_t\right)\left(\overline{\mathcal{R}}_t - \widehat{\mathcal{B}}\overline{F}_t\right)'$$

as the covariance matrix estimator of Σ for the proposed test statistics.

3 See Lemmas 1 and 2 in the Appendix. For ease of exposition, all proofs are provided in the Appendix, so we can focus on our prime results in the main text.

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

Theorem 1. *If the same covariance estimator* $\widehat{\Sigma}$ *is used for the FAR and H statistics, then* $\mathrm{FAR}(\lambda_{F,0}) = H(\lambda_{F,0})$. *Under* $H_0 : \lambda_F = \lambda_{F,0}$ *and Assumptions 1 and 2, the FAR and H statistics converge to a* χ^2_{N-1} *distributed random variable, as* $T \to \infty$ *while* N *is fixed.*

Proof: See Appendix A.[4]

Kleibergen and Zhan (2020) provide another interpretation of the FAR statistic by relating it to the well-known GRS statistic. They show that $\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}$ can be viewed as an estimator of the constant terms (alphas) in a set of appropriately specified regressions of asset returns on risk factors. While alphas are zero under the null, the FAR statistic, like the GRS statistic, just tests whether alphas are zero. Therefore, we could also label the FAR statistic as the GRS-FAR statistic, as in Kleibergen and Zhan (2020).

The asymptotic result in Theorem 1, however, is at risk in a finite-sample setting where the time-series dimension T may not be much larger than the cross-section dimension N. We, therefore, adjust the FAR statistic (or equivalently, the H statistic) and derive its finite-sample distribution, which is provided in Theorem 2.

Theorem 2. *Under* $H_0 : \lambda_F = \lambda_{F,0}$ *and Assumptions 1 and 2,*

$$\frac{T-K-N+1}{(T-K-1)(N-1)} \times \mathrm{FAR}(\lambda_{F,0}) \sim F_{N-1,T-K-N+1}, \tag{11}$$

where $F_{N-1,T-K-N+1}$ *is the F-distribution with* $N-1$, $T-K-N+1$ *degrees of freedom.*

Proof: See Appendix A.

The finite-sample F-distribution in Theorem 2 indicates the importance of the time-series dimension T relative to the cross-section dimension N. If T is much larger than N, then it is straightforward to verify that the finite-sample result in Theorem 2 is equivalent to the asymptotic result in Theorem 1. On the other hand, if T is close to N, which is not uncommon in practice, then inference on risk premia based on Theorem 2 is expected to be substantially different from that based on Theorem 1. The related simulation evidence can be found in Kleibergen and Zhan (2020).

2.2.2 GLS-LM and JGLS statistics.

As stated in Theorem 1, asymptotically the FAR statistic has a χ^2 distribution with $N-1$ degrees of freedom. Therefore, when we use a large number of test assets, so the degrees of freedom increase, the FAR test becomes less powerful. This issue has been similarly discussed for the Anderson and Rubin (1949) statistic in the instrumental variables regression literature (Kleibergen 2002).

To reduce the degrees of freedom and thus improve the power of tests on the risk premia, Kleibergen (2009) decomposes the FAR statistic into two parts, the GLS-LM statistic and the JGLS statistic, both can be used for testing $\mathrm{H}_0 : \lambda_F = \lambda_{F,0}$:

4 Theorem 1, as well as other asymptotic results presented later on, also holds under weaker conditions; see Kleibergen (2009).

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

$$\mathrm{FAR}(\lambda_{F,0}) = \mathrm{GLS\text{-}LM}(\lambda_{F,0}) + \mathrm{JGLS}(\lambda_{F,0}).$$

The GLS-LM statistic results from projecting $\widehat{\Sigma}^{-\frac{1}{2}}\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)$ on the column space spanned by $\widehat{\Sigma}^{-\frac{1}{2}}\widetilde{\mathcal{B}}$, so

$$\begin{aligned}
&\mathrm{GLS\text{-}LM}(\lambda_{F,0}) \\
&= \frac{T}{1 - \lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\lambda_{F,0}} \left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)' \widehat{\Sigma}^{-\frac{1}{2}'} P_{\widehat{\Sigma}^{-\frac{1}{2}}\widetilde{\mathcal{B}}} \widehat{\Sigma}^{-\frac{1}{2}} \left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right) \\
&= \frac{T}{1 - \lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\lambda_{F,0}} \left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)' \widehat{\Sigma}^{-1}\widetilde{\mathcal{B}} \left(\widetilde{\mathcal{B}}'\widehat{\Sigma}^{-1}\widetilde{\mathcal{B}}\right)^{-1} \widetilde{\mathcal{B}}'\widehat{\Sigma}^{-1} \left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)
\end{aligned}, \tag{12}$$

where $P_{\widehat{\Sigma}^{-\frac{1}{2}}\widetilde{\mathcal{B}}} = \widehat{\Sigma}^{-\frac{1}{2}}\widetilde{\mathcal{B}}\left(\widetilde{\mathcal{B}}'\widehat{\Sigma}^{-1}\widetilde{\mathcal{B}}\right)^{-1}\widetilde{\mathcal{B}}'\widehat{\Sigma}^{-\frac{1}{2}'}$, $\widehat{\Sigma}^{-1} = \widehat{\Sigma}^{-\frac{1}{2}'}\widehat{\Sigma}^{-\frac{1}{2}}$. Moreover, the GLS-LM statistic can also be viewed as a quadratic form of the derivative of the FAR statistic (Kleibergen 2009).

Similarly, the JGLS statistic results from projecting $\widehat{\Sigma}^{-\frac{1}{2}}\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)$ on the space orthogonal to the one spanned by $\widehat{\Sigma}^{-\frac{1}{2}}\widetilde{\mathcal{B}}$, so it equals $\mathrm{FAR}(\lambda_{F,0}) - \mathrm{GLS\text{-}LM}(\lambda_{F,0})$:

$$\begin{aligned}
&\mathrm{JGLS}(\lambda_{F,0}) \\
&= \frac{T}{1 - \lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\lambda_{F,0}} \left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)' \widehat{\Sigma}^{-\frac{1}{2}'} \left(I_{N-1} - P_{\widehat{\Sigma}^{-\frac{1}{2}}\widetilde{\mathcal{B}}}\right) \widehat{\Sigma}^{-\frac{1}{2}} \left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right).
\end{aligned} \tag{13}$$

By decomposing the FAR statistic as described above, the distributions of the resulting GLS-LM and JGLS statistics have fewer degrees of freedom, as shown in Theorem 3.

Theorem 3. *Under $H_0 : \lambda_F = \lambda_{F,0}$ and Assumptions 1 and 2, the GLS-LM and JGLS statistics converge to independent χ^2_K and χ^2_{N-K-1} distributed random variables, respectively, as $T \to \infty$ while N is fixed.*

Proof: See Appendix A.

To make GLS-LM and JGLS tests applicable for the limited T versus large N setting, we similarly derive the finite-sample distributions of their (re-scaled) statistics. These are provided in Theorem 4.

Theorem 4. *Under $H_0 : \lambda_F = \lambda_{F,0}$ and Assumptions 1 and 2,*

$$\mathrm{GLS} - \mathrm{LM}(\lambda_{F,0}) \sim \grave{\psi}\mathcal{W}^{-1}\psi - \grave{\psi}C(\grave{C}\mathcal{W}C)^{-1}\grave{C}\psi \tag{14}$$

$$\frac{T - N + 1}{(T - K - 1)(N - K - 1)} \times \mathrm{JGLS}(\lambda_{F,0}) \sim F_{N-K-1,T-N+1}, \tag{15}$$

where $\psi \sim N(0, I_{N-1})$, $\mathcal{W} \sim \frac{\mathcal{W}_{N-1}(T-K-1,I_{N-1})}{T-K-1}$ (scaled Wishart distribution) independent of ψ, and C is an arbitrary $(N-1) \times (N-K-1)$ matrix of full rank.

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

Proof: See Appendix A.

The finite-sample distribution of the GLS-LM statistic is non-standard, but, since it is invariant to the specification of C, it does not depend on any further nuisance parameters so its critical values can easily be obtained by simulation.[5] The F-distribution of the JGLS statistic has degrees of freedom different from those for the finite-sample FAR statistic, although both statistics take the limited T versus large N problem into account.

2.2.3 FM-LM and JFM statistics.

Next, we decompose the FAR statistic in an alternative manner, which is similar to the above decomposition resulting in the GLS-LM and JGLS statistics. This decomposition leads to yet two other test statistics, denoted by FM-LM and JFM, respectively, in Kleibergen (2009). As shown later, it yields a test whose finite-sample distribution does not depend on N.

In particular, we project $\widehat{\Sigma}^{-\frac{1}{2}}\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)$ on the column space spanned by $\widehat{\Sigma}^{\frac{1}{2}}\widetilde{\mathcal{B}}$ and the associated orthogonal space, to construct the FM-LM and JFM statistics, respectively. These two statistics can thus be written as:

$$\begin{aligned}
&\text{FM-LM}(\lambda_{F,0}) \\
&= \frac{T}{1 - \lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\lambda_{F,0}}\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)'\widehat{\Sigma}^{-\frac{1'}{2}}P_{\widehat{\Sigma}^{-\frac{1}{2}}\widetilde{\mathcal{B}}}\widehat{\Sigma}^{-\frac{1}{2}}\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right) \\
&= \frac{T}{1 - \lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\lambda_{F,0}}\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)'\widetilde{\mathcal{B}}\left(\widetilde{\mathcal{B}}'\widehat{\Sigma}\widetilde{\mathcal{B}}\right)^{-1}\widetilde{\mathcal{B}}'\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)
\end{aligned} \tag{16}$$

$$\begin{aligned}
&\text{JFM}(\lambda_{F,0}) \\
&= \text{FAR}(\lambda_{F,0}) - \text{FM-LM}(\lambda_{F,0}) \\
&= \frac{T}{1 - \lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\lambda_{F,0}}\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)'\widehat{\Sigma}^{-\frac{1'}{2}}\left(I_{N-1} - P_{\widehat{\Sigma}^{-\frac{1}{2}}\widetilde{\mathcal{B}}}\right)\widehat{\Sigma}^{-\frac{1}{2}}\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)
\end{aligned} \tag{17}$$

with $P_{\widehat{\Sigma}^{-\frac{1}{2}}\widetilde{\mathcal{B}}} = \widehat{\Sigma}^{\frac{1'}{2}}\widetilde{\mathcal{B}}\left(\widetilde{\mathcal{B}}'\widehat{\Sigma}\widetilde{\mathcal{B}}\right)^{-1}\widetilde{\mathcal{B}}'\widehat{\Sigma}^{\frac{1}{2}}$, $\widehat{\Sigma} = \widehat{\Sigma}^{\frac{1}{2}}\widehat{\Sigma}^{\frac{1'}{2}}$.

Theorems 5 and 6 provide the asymptotic and finite-sample distributions for these statistics, respectively.

Theorem 5. *Under $H_0 : \lambda_F = \lambda_{F,0}$ and Assumptions 1 and 2, the FM-LM and JFM statistics converge to independent χ^2_K and χ^2_{N-K-1} distributed random variables, respectively, as $T \to \infty$ while N is fixed.*

5 To get the finite-sample critical values for GLS-LM, we can independently simulate ψ, $\mathcal{W}$ from normal and Wishart distributions, respectively, and use an $(N-1)\times(N-K-1)$ full rank matrix for C. The quantiles resulting from the simulated values of the GLS-LM statistic provide the critical values.

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

Proof: See Appendix A.

Theorem 6. *Under* $H_0 : \lambda_F = \lambda_{F,0}$ *and Assumptions 1 and 2,*

$$\frac{T-2K}{(T-K-1)K} \times \text{FM-LM}(\lambda_{F,0}) \sim F_{K,T-2K} \tag{18}$$

$$\text{JFM}(\lambda_{F,0}) \sim \psi' \mathcal{W}^{-1} \psi - \psi' \mathbf{C} (\mathbf{C}' \mathcal{W} \mathbf{C})^{-1} \mathbf{C}' \psi, \tag{19}$$

where ψ *and* $\mathcal{W}$ *are as in Theorem 4, while* $\mathbf{C}$ *is an arbitrary* $(N-1) \times K$ *matrix of full rank.*

Proof: See Appendix A.

It is worth noting that the finite-sample F-distribution of the FM-LM statistic in Theorem 6 has a nice feature: its degrees of freedom do not depend on the number of assets N. This statistic can thus be used to overcome the deficiency due to the presence of many assets, for which we provide simulation evidence later on.

2.2.4 Split-sample CLR statistic.

The conditional LR (CLR) test for $H_0 : \lambda_F = \lambda_{F,0}$ resembles the CLR test in Moreira (2003) for linear instrumental variables regression. In the single factor case, the CLR statistic has a closed form expression:

$$\begin{aligned} &\text{CLR}(\lambda_{F,0}) \\ &= \frac{1}{2}\left\{ \text{FAR}(\lambda_{F,0}) - r(\lambda_{F,0}) + \sqrt{\left(\text{FAR}(\lambda_{F,0}) + r(\lambda_{F,0})\right)^2 - 4r(\lambda_{F,0})\text{JGLS}(\lambda_{F,0})} \right\}, \end{aligned} \tag{20}$$

where $r(\lambda_{F,0}) = \widehat{\mathcal{Q}}_{FF}(\lambda_{F,0}) \widetilde{\mathcal{B}}' \widehat{\Sigma}^{-1} \widetilde{\mathcal{B}}$ corresponds to a rank test statistic for testing $\text{rank}(\mathcal{B}) = 0$ provided $\lambda_F = \lambda_{F,0}$ and it measures the strength of the risk premia identification. When $r(\lambda_{F,0})$ equals zero, so there is no identification, the CLR statistic is identical to the FAR statistic. On the other hand, for large values of $r(\lambda_{F,0})$, the CLR statistic is close to the GLS-LM statistic. Therefore, the CLR statistic can also be regarded as a data-dependent weighted average of the FAR and GLS-LM statistics.

Asymptotically, $r(\lambda_{F,0})$ is independent of $\text{FAR}(\lambda_{F,0})$ and $\text{JGLS}(\lambda_{F,0})$, so the conditional limiting distribution of the CLR statistic given the realized value of $r(\lambda_{F,0})$ results from combining the independent limiting distributions of the GLS-LM and JGLS statistics whose sum makes up the FAR statistic. The conditional critical values of the CLR statistic can then straightforwardly be simulated for any given value of $r(\lambda_{F,0})$ in the manner of Moreira (2003); see also Kleibergen (2009). In the finite-sample setting, however, $\text{FAR}(\lambda_{F,0})$ and $\text{JGLS}(\lambda_{F,0})$ are no longer independently distributed of $r(\lambda_{F,0})$, which complicates the derivation of conditional finite-sample critical values for the CLR statistic. To address this issue, we propose to conduct a CLR test by usage of the split-sample CLR (sCLR) statistic.

Specifically, the full sample with $t = 1, \ldots, T$, is divided into two parts: $t = 1, \ldots, T_1$ for the first sub-sample, while $t = T_1 + 1, \ldots, T$ for the second. We use the first sub-sample to estimate Σ, which is only used for calculating $r(\lambda_{F,0})$. All the other parts of the CLR statistic are computed using the second sub-sample. The resulting CLR statistic is

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

labeled as the sCLR statistic. The $r(\lambda_{F,0})$ statistic used for the sCLR statistic is then independent of the FAR$(\lambda_{F,0})$ and JGLS$(\lambda_{F,0})$ statistics. We can thus provide the finite-sample distribution of the sCLR statistic, as in the next theorem.

Theorem 7. *Under* $H_0 : \lambda_F = \lambda_{F,0}$ *and Assumptions 1 and 2 with* $K = 1$, *the* sCLR$(\lambda_{F,0})$ *statistic conditional on the rank statistic r has the distribution:*

$$\frac{1}{2}\left(\psi'\mathcal{W}^{-1}\psi - r + \sqrt{\left(\psi'\mathcal{W}^{-1}\psi + r\right)^2 - 4r\psi'C(C'\mathcal{W}C)^{-1}C'\psi}\right),$$

where ψ, $\mathcal{W}$, and C *are as in Theorem 4, except that* $T - T_1$ *replaces* T *for the (second) sample size, and r is the rank statistic whose* $\widehat{\Sigma}$ *results from the first sample with* $t = 1, \ldots, T_1$.

Proof: See Appendix A.

The sCLR test is, however, not always feasible. If T is too close to N, the two split-sample sizes, T_1 and $T - T_1$, may not exceed N. Yet we require the time-series dimension to at least exceed the cross-section dimension for our proposed tests. For example, the degrees of freedom in Theorem 2 implicitly implies that we need $T - K - N + 1 > 0$. In the single factor model with $K = 1$, $T - K - N + 1 > 0$ reduces to $T > N$. For the sCLR statistic, both T_1 and $T - T_1$ thus have to exceed N so T has to exceed $2N$. For this reason, we focus on the FAR test and its components, the GLS-LM, JGLS, FM-LM, and JFM tests, rather than the CLR test.

2.3 Subset FAR Statistic

So far we have focused on testing a joint hypothesis specified on all risk premia: $H_0 : \lambda_F = \lambda_{F,0}$. To construct a confidence set of each individual risk premium in multi-factor models, we can just project out the joint confidence set which can be computed using tests of the joint hypothesis specified on all risk premia. Alternatively, Kleibergen (2009) and Kleibergen and Zhan (2020) use a subset FAR (sFAR) statistic. We provide the asymptotic and finite-sample behaviors of the sFAR statistic in Theorems 9 and 10, respectively. We refrain from constructing the asymptotic and finite-sample behaviors of the other statistics since these, unlike for the sFAR statistic, need the risk premia left unrestricted by the hypothesis of interest to be well identified.

For the $K \times 1$ dimensional vector of risk premia $\lambda_F = (\lambda_{F,1}, \lambda_{F,2}, \ldots, \lambda_{F,K})'$, consider its partition such that $\lambda_F = (\lambda_{F,1}, \lambda_2')'$ with $\lambda_2 = (\lambda_{F,2}, \ldots, \lambda_{F,K})'$. Our objective is to test the null hypothesis $H_0 : \lambda_{F,1} = \lambda_{F,1}^0$.[6] The corresponding sFAR statistic, denoted by sFAR$(\lambda_{F,1}^0)$, results from minimizing the FAR statistic under the hypothesized $\lambda_{F,1}^0$:

$$\text{sFAR}\left(\lambda_{F,1}^0\right) = \min_{\lambda_2} \text{FAR}\left(\left(\lambda_{F,1}^0, \lambda_2'\right)'\right).$$

6 Our results can easily extend to cases where $\lambda_{F,1}$ is a vector instead of a scalar, and for simplicity and for the fact that scalar tests are popular among empirical studies, we focus on the scalar case.

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

Consider the following characteristic polynomial:

$$|\mu \begin{pmatrix} 1+\lambda_{F,1}'\widehat{\mathcal{Q}}_{FF,11}^{-1}\lambda_{F,1} & -\lambda_{F,1}'\widehat{\mathcal{Q}}_{FF,12}^{-1} \\ -\widehat{\mathcal{Q}}_{FF,21}^{-1}\lambda_{F,1} & \widehat{\mathcal{Q}}_{FF,22}^{-1} \end{pmatrix} - \left(\overline{\mathcal{R}} - \widehat{\mathcal{B}}_1\lambda_{F,1}, \widehat{\mathcal{B}}_2\right)'\widehat{\Sigma}^{-1}\left(\overline{\mathcal{R}} - \widehat{\mathcal{B}}_1\lambda_{F,1}, \widehat{\mathcal{B}}_2\right)| = 0,$$

with $\widehat{\mathcal{Q}}_{FF}^{-1} = \begin{pmatrix} \widehat{\mathcal{Q}}_{FF,11}^{-1} & \widehat{\mathcal{Q}}_{FF,12}^{-1} \\ \widehat{\mathcal{Q}}_{FF,21}^{-1} & \widehat{\mathcal{Q}}_{FF,22}^{-1} \end{pmatrix}$, $\widehat{\mathcal{B}} = \left(\widehat{\mathcal{B}}_1, \widehat{\mathcal{B}}_2\right)$. The $\text{sFAR}\left(\lambda_{F,1}^0\right)$ statistic equals T times the smallest root of this characteristic polynomial, which shows that the $\text{sFAR}\left(\lambda_{F,1}^0\right)$ statistic is also a rank test statistic testing the null hypothesis that the rank of $\left(\mathbb{E}(\mathcal{R}_t) - \mathcal{B}_1\lambda_{F,1}, \mathcal{B}_2\right)$ is smaller than $K-1$; see Kleibergen and Paap (2006).

Proposition 8. *The sFAR statistic is bounded according to*

$$\text{sFAR}\left(\lambda_{F,1}^0\right) \leq \widetilde{\text{JGLS}}\left(\left(\lambda_{F,1}^0, \lambda_2^{0\prime}\right)'\right) \tag{21}$$

with $\widetilde{\text{JGLS}}((\lambda_{F,1}^0, \lambda_2^{0\prime})') = \text{FAR}((\lambda_{F,1}^0, \lambda_2^{0\prime})') - \widetilde{\text{GLS-LM}}((\lambda_{F,1}^0, \lambda_2^{0\prime})')$, *where* $\widetilde{\text{GLS-LM}}((\lambda_{F,1}^0, \lambda_2^{0\prime})')$ *only uses the score with respect to the* λ_2.

Proof: See Appendix A.

Proposition 8 directly implies that the distribution of a JGLS statistic can serve as an upper bound for the associated sFAR statistic, as shown in Theorems 9 and 10.

Theorem 9. *Under* $H_0 : \lambda_{F,1} = \lambda_{F,1}^0$ *and Assumptions 1 and 2, as* $T \to \infty$ *while* N *is fixed, the limiting distribution of the* $\text{sFAR}(\lambda_{F,1}^0)$ *statistic is bounded by the* χ^2_{N-K} *distribution, that is,*

$$\text{sFAR}(\lambda_{F,1}^0) \preceq \chi^2_{N-K}. \tag{22}$$

Proof: See Appendix A.

Theorem 9 is consistent with the sharp asymptotic sFAR upper bound provided by Guggenberger et al. (2012), while we derive the result via a different approach, that is, we derive the upper bound by showing that the sFAR statistic is bounded by a JGLS statistic as stated in Proposition 8. Theorem 4 provides an exact finite-sample distribution of the JGLS statistic, similarly we can derive the exact finite-sample distribution of the $\widetilde{\text{JGLS}}((\lambda_{F,1}^0, \lambda_2^{0\prime})')$ statistic and thus provide a finite-sample upper bound for the sFAR statistic in Theorem 10.

Theorem 10. *Under* $H_0 : \lambda_{F,1} = \lambda_{F,1}^0$, *Assumptions 1 and 2, the scaled* $\text{sFAR}(\lambda_{F,1}^0)$ *statistic,* $\frac{T-N}{(T-K-1)(N-K)}\text{sFAR}(\lambda_{F,1}^0)$, *is in finite samples bounded by the* $F_{N-K,T-N}$ *distribution, that is,*

$$\frac{T-N}{(T-K-1)(N-K)} \times \text{sFAR}(\lambda_{F,1}^0) \preceq F_{N-K,T-N}. \tag{23}$$

Proof: See Appendix A.

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

Theorem 10 is also consistent with the asymptotic χ^2-upper bound from Theorem 9 as the scaled F-distribution will converge to the same χ^2-limit as T increases to infinity. The sFAR statistic, as previously discussed, is also a rank test statistic. Theorem 11 extends the finite-sample bound of the sFAR statistic to the generalized reduced rank test statistic (Kleibergen and Paap 2006) with a Kronecker structure, which coincides with the canonical correlation rank statistic of Anderson (1951).

Theorem 11. *Under Assumptions 1 and 2, the canonical correlation rank test which tests for a lower rank value of* $\mathcal{B}$ *(from Equation (6))*, $rk = \min_\lambda T \frac{(\widehat{\mathcal{B}}\lambda)'\widehat{\Sigma}^{-1}(\widehat{\mathcal{B}}\lambda)}{\lambda'\widehat{\mathcal{Q}}_{FF}^{-1}\lambda}$, *is in finite samples bounded by the* $F_{N-K,T-N}$ *distribution, that is,*

$$\frac{T-N}{(T-K-1)(N-K)} \times rk \preceq F_{N-K,T-N}. \tag{24}$$

Proof: See Appendix A.

Theorem 11 provides a finite-sample upper bound for the canonical correlation rank test of $\mathcal{B}$, which commonly serves as a pre-test testing the identification strength of linear asset pricing models. Provided that Equation (6) is simply a linear model, Theorem 11 easily extends to the canonical correlation rank test of slope coefficients in linear models with independent Gaussian noise and thus provides a finite-sample adaption for the rank test with a Kronecker structure (Kleibergen and Paap 2006).

3 Numerical and Empirical Results

3.1 Simulation Evidence

To illustrate the performance of our proposed tests, we conduct a simple simulation study. For the data generating process (DGP), we consider a single factor model:

$$R_t = c + \beta F_t + u_t, \text{ with } F_t \sim \text{NID}\big(0, V_f\big), \; u_t \sim \text{NID}(0, \Omega),$$

where c, β, V_f, and Ω are calibrated to data from Kroencke (2017). We set $c = \iota_N \lambda_0 + \beta\lambda_F$, so the moment condition (1) holds.

3.1.1 Size.

Using the simulated data of F_t and R_t from the DGP, we test $H_0 : \lambda_F = 2$ at the 5% significance level for generating the sizes presented in Figure 1. In particular, we fix the time-series sample size at $T = 55$ to mimic the Kroencke (2017) study, while we consider a sequence of values for the number of test assets N, up to $N = 31$. The FAR, GLS-LM, JGLS, FM-LM, and JFM tests described in the last section are implemented, using both their finite-sample and asymptotic critical values. The simulated sizes result from 1,000 Monte Carlo replications.

Figure 1 shows that all finite-sample robust tests (dashed curves) have actual sizes close to the nominal 5%, regardless of the value of N. This is in sharp contrast with the asymptotically valid tests (solid curves), whose sizes tend to exceed the nominal 5% as N increases. When N is small, so it is far below T, we observe in Figure 1 that asymptotic tests

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

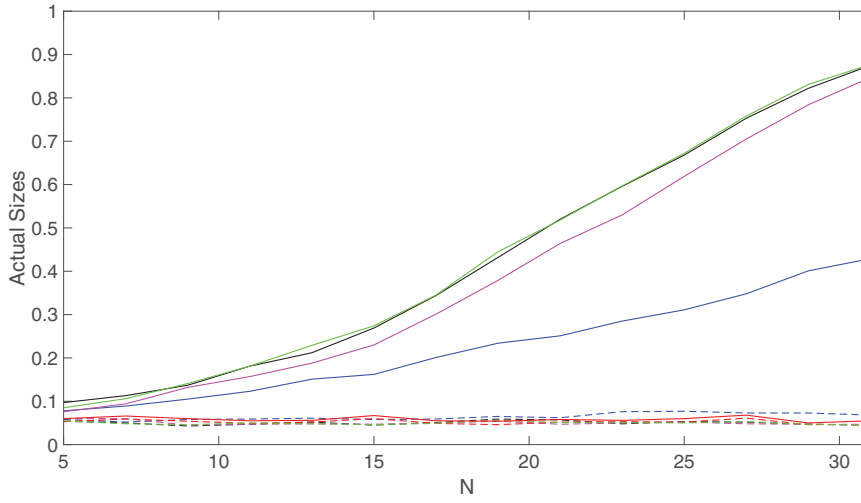

Figure 1 Simulated sizes as a function of the number of test assets N with $T = 55$. Dashed curves for finite-sample robust tests: FAR, dashed black; GLS-LM, dashed blue; JGLS, dashed magenta; FM-LM, dashed red; JFM, dashed green. Solid curves for asymptotically valid tests: FAR, solid black; GLS-LM, solid blue; JGLS, solid magenta; FM-LM, solid red; and JFM, solid green. The null hypothesis is $H_0 : \lambda_F = 2$ in a single factor CAPM calibrated to Kroencke (2017). The significance level is 5%.

are close to be size-correct. On the other hand, when N is large, so the limited T versus large N problem becomes severe, asymptotic tests (except FM-LM) exhibit size distortion while finite-sample robust tests remain trustworthy. All these findings are in line with our earlier analysis.

It is worth emphasizing that the FM-LM test performs well using either finite-sample (dashed red) or asymptotic (solid red) critical values. This is explained by Theorem 6, where N does not appear in the degrees of freedom of the finite-sample distribution of the FM-LM statistic. Put differently, while the limited T versus large N problem affects many asymptotically valid tests, it is less so for the FM-LM test. This feature makes the FM-LM test more appealing than the FAR, GLS-LM, JGLS, and JFM tests for limited T versus large N.

Figure 1 is to be compared with Figure 2, for which we increase T from 55 to 500. Under the large $T = 500$, Figure 2 shows that both the asymptotically valid tests (solid) and finite-sample robust tests (dashed) have actual sizes near the nominal 5%. This further demonstrates the importance of the T versus N issue. If researchers adopt a large number of test assets, such as $N = 31$ assets in Kroencke (2017), or $N = 35$ as in Savov (2011), then a much larger T is generally needed to validate asymptotic tests. This argument also applies to the identification robust FAR, GLS-LM, JGLS, and JFM tests when using asymptotic critical values. Kleibergen and Zhan (2020) provide similar simulation evidence for the FM t-test, whose size not only suffers from the limited T versus large N problem, but also relies on risk factors of satisfactory quality.

3.1.2 Power.

Next, we briefly analyze the power performance of our proposed tests. To do so, we vary λ_F imposed in the DGP described above and keep testing $H_0 : \lambda_F = 2$, which is close to the

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

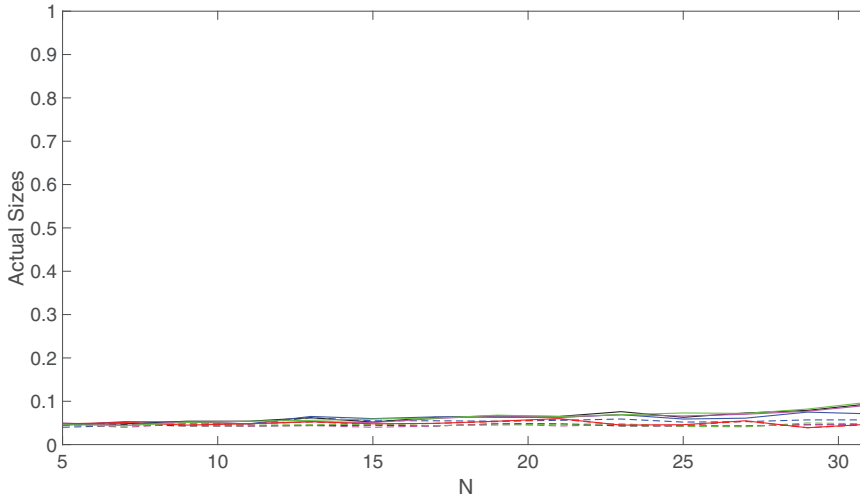

Figure 2 This figure is generated in the same manner as Figure 1, except that we fix T at $T = 500$.

point estimate of the consumption risk premium reported in Kroencke (2017). To facilitate comparison, we consider two scenarios: (i) $N = 31$ and $T = 55$, as in Kroencke (2017) and (ii) $N = 31$ and $T = 500$. The simulated power curves for these two scenarios are presented in Figures 3 and 4, respectively.

The left-hand side panel of Figure 3 shows that finite-sample robust tests are all size-correct in our simulation study calibrated to Kroencke (2017), and these tests have good power. The power curves of the GLS-LM and FM-LM tests show a sudden decline in power which results since they are (functions of) quadratic forms of the derivative of the FAR statistic with respect to the risk premium. The power decline therefore occurs from this derivative not only being zero at the minimizer of the FAR statistic. Andrews, Moreira, and Stock (2006) show that the CLR statistic is optimal for testing in the homoskedastic linear instrumental variables regression model with one structural parameter, since it optimally combines the GLS-LM and JGLS statistic using the conditioning statistic. Since the conditioning statistic is not independently distributed from the GLS-LM and JGLS statistics, we cannot use it here.

In contrast, the right hand side panel of Figure 3 shows that among the examined asymptotically valid tests, only the FM-LM test appears to be size-correct for the $N = 31$ and $T = 55$ setting. This is consistent with Figure 1 as well as Theorem 6.

Figure 3 is to be compared with Figure 4, for which we consider $T = 500$. The two panels of Figure 4 now appear similar to each other, since both the finite-sample robust tests and the asymptotic tests are expected to be valid when T is sufficiently large. Similar to the left-hand side panel of Figure 3, Figure 4 also suggests that the power performances of the examined tests are comparable to each other.

Kleibergen (2009) discusses how the power decline of GLS-LM and FM-LM tests, as shown in Figure 3, can be avoided by sequentially testing the risk premia using either the independent GLS-LM and JGLS tests or the FM-LM and JFM tests. For illustration, here we

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

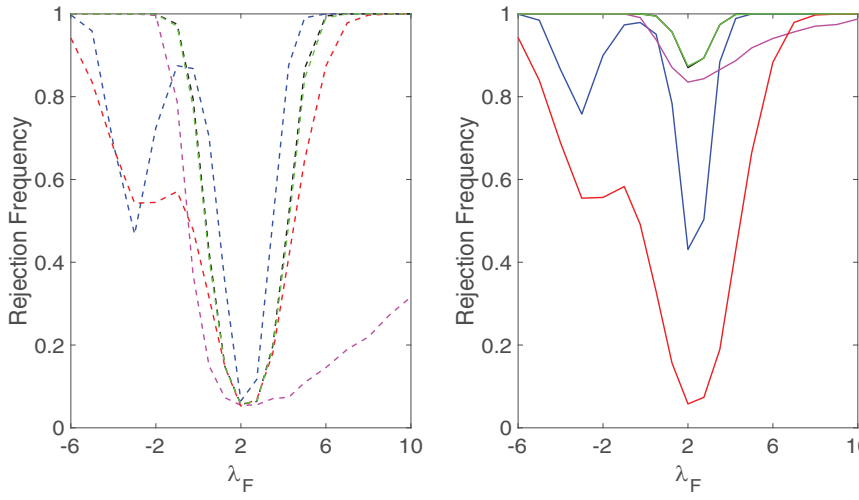

Figure 3 Simulated power curves with $N = 31$ and $T = 55$. Dashed curves for finite-sample robust tests in the left panel: FAR, dashed black; GLS-LM, dashed blue; JGLS, dashed magenta; FM-LM, dashed red; and JFM, dashed green. Solid curves for asymptotically valid tests in the right panel: FAR, solid black; GLS-LM, solid blue; JGLS, solid magenta; FM-LM, solid red; and JFM, solid green. The null hypothesis is $H_0 : \lambda_F = 2$ in a single factor CAPM calibrated to Kroencke (2017). The significance level is 5%.

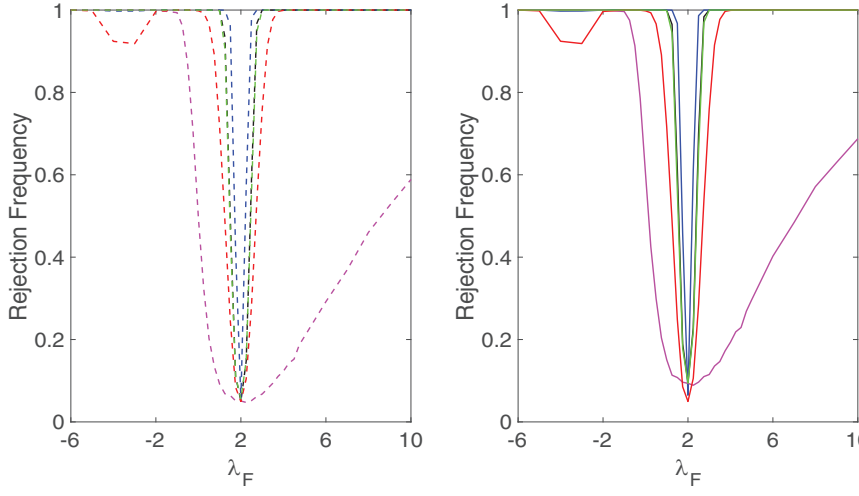

Figure 4 This figure is generated in the same manner as Figure 3, except that we fix T at $T = 500$.

consider a combination such that a 96% critical value is applied to GLS-LM and a 99% critical value to JGLS, so the overall size equals 5%. Similarly, we also consider the combined test using FM-LM (96%) and JFM (99%).

Figure 5 presents the power curves of the GLS-LM and JGLS combined test and the FM-LM and JFM combined test in the $N = 31$ and $T = 55$ setting. For comparison, we also

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

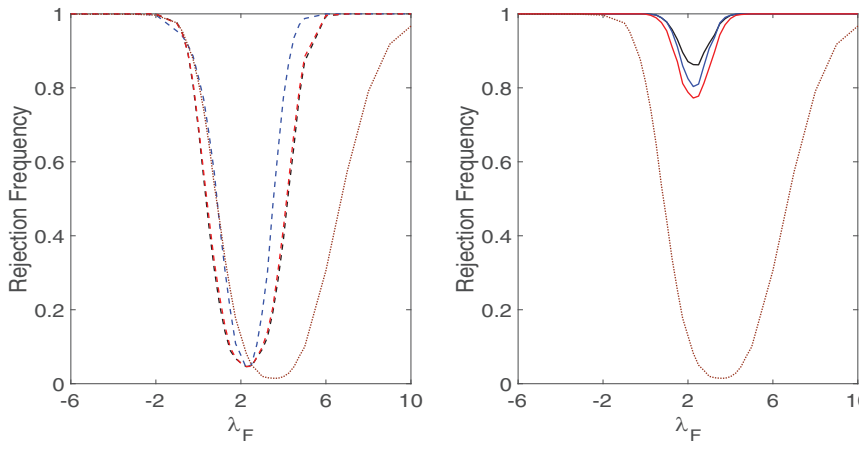

Figure 5 Simulated power curves with $N = 31$ and $T = 55$. Dashed curves for finite-sample robust tests in the left panel: FAR, dashed black; a combined test of GLS-LM (96%) and JGLS (99%), dashed blue; a combined test of FM-LM (96%) and JFM (99%), dashed red. Solid curves for asymptotically valid tests in the right panel: FAR, solid black; a combination of JGLS and GLS-LM, solid blue; a combination of JFM and FM-LM, solid red. In both panels, the dotted brown power curve is for the FM *t*-test with the Shanken (1992) correction, using critical values from the normal distribution. The null hypothesis is $H_0 : \lambda_F = 2$ in a single factor CAPM calibrated to Kroencke (2017). The significance level is 5%.

present the power curves of the FAR test and the FM *t*-test with the Shanken (1992) correction.

The left hand side panel of Figure 5 shows that the power decline in Figure 3 is no longer in place, once we combine tests in the manner suggested by Kleibergen (2009). Moreover, the GLS-LM and JGLS combined test (dashed blue) appears more powerful than the FM-LM and JFM combined test (dashed red) and the FAR test (dashed black). On the other hand, the FM *t*-test (dotted brown) exhibits size distortion, since it over-rejects $H_0 : \lambda_F = 2$. The right hand side panel of Figure 5 again shows that the asymptotic counterparts of these tests do not function well in the $N = 31$ and $T = 55$ setting.

Figure 5 is to be compared with Figure 6, where T is increased to 500. Under this large value of T, the two panels in Figure 6 are, as expected, almost indistinguishable. Figure 6 also shows that the FM *t*-test starts to perform well under the large $T(= 500)$. However, its power curve lies below those of robust tests, that is, it appears to be less powerful.

3.2 Risk Premia in the Consumption CAPM

In this section, we use our proposed tests to analyze risk premia in the consumption CAPM and conditional consumption CAPM. The data sets from Kroencke (2017) and Lettau and Ludvigson (2001) are adopted to facilitate comparison. The yearly data taken from Kroencke (2017) has $T = 55$ and $N = 31$, while the quarterly data taken from Lettau and Ludvigson (2001) has $T = 141$ and $N = 25$. The detailed description of the involved test assets and risk factors is provided for Tables 1 and 2, where the estimated βs are reported.

For ease of exposition, we present the *p*-value plots resulting from testing risk premia using our proposed tests. A *p*-value larger than, say, 5%, implies that we could not reject the null at the 5% significance level. Thus, if we find that the *p*-value curves are all above

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

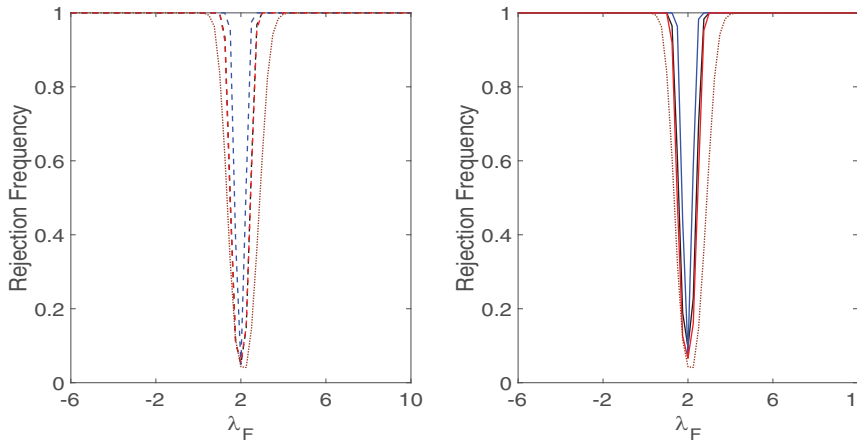

Figure 6 This figure is generated in the same manner as Figure 5, except that we fix T at $T = 500$.

the 5% line, then we cannot reject any hypothesized value of the risk premium, that is, the 95% confidence sets of the risk premium is unbounded. In contrast, if the p-value curves cross the 5% line, then we can reject those hypothesized values of the risk premium whose corresponding p-values are below 5%. Finally, if the p-value curves lie fully below the 5% line, then we reject every hypothesized value of the risk premium.

Figure 7 provides the p-value curves resulting from testing the consumption risk premium using the data from Kroencke (2017). We use five of our proposed tests: FAR, GLS-LM, JGLS, FM-LM, and JFM.[7] Overall, the p-value curves based on the finite-sample robust tests (dashed black) imply unbounded 95% confidence sets of the consumption risk premium. In particular, the p-values from the finite-sample FAR, GLS-LM, JGLS, and JFM tests are all above 5%, so the resulting confidence sets are unbounded. Only the FM-LM test has some of its p-values below 5%, for some hypothesized risk premia near zero. Therefore, only these hypothesized values can be rejected at the 5% level, and the 95% confidence set resulting from the FM-LM test is thus unbounded but disjoint. The unboundedness of the 95% confidence sets reflects both the questionable quality of the consumption growth data and the limited sample size, and casts doubt on the findings in Kroencke (2017) based on the FM t-test, which we re-produce in Panel A of Table 3. Dufour (1997) and Kleibergen and Zhan (2020) provide further detailed explanations for why confidence sets resulting from robust tests can be unbounded. In particular, Kleibergen and Zhan (2020) show that unbounded confidence sets commonly occur, if risk premia are unidentified due to poor quality risk factors, and/or the involved time series is short, so little information on risk premia can be extracted from the data.

Figure 7 also shows the large discrepancy between finite-sample robust tests (dashed black) and asymptotically valid tests (solid red). Most striking is that the asymptotic FAR, JGLS, and JFM tests reject every hypothesized risk premium, while their finite-sample counterparts do not. The FM-LM test distinguishes itself from the other tests, since its

7 Since $T = 55$ while $N = 31$, we do not employ the sCLR test which needs $T > 2N$.

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

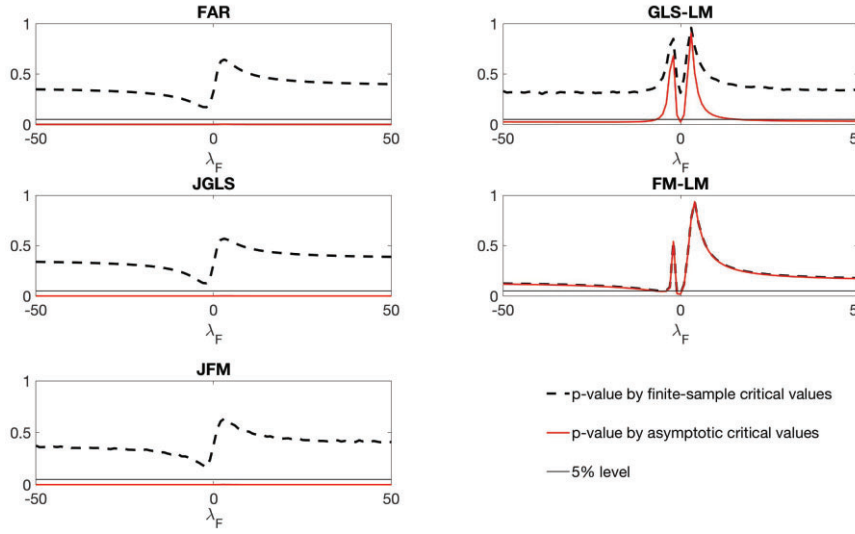

Figure 7 *p*-values of the FAR, GLS-LM, JGLS, FM-LM, and JFM tests for testing the values of the consumption risk premium on the horizontal axis. Dashed black curves result from the finite-sample critical values of these test statistics, while solid red curves of *p*-values result from the asymptotic critical values. The test assets (30 portfolios sorted by size, value, and investment, plus the market portfolio) and the single risk factor (unfiltered consumption growth) in 1960–2014 are taken from Kroencke (2017). The 5% benchmark line is also plotted.

Table 3 Conventional 95% confidence intervals (CIs) of risk premia

	A. Kroencke (2017)	B. Lettau and Ludvigson (2001)		
	$\triangle c$	$\triangle c$	cay	$\triangle c \times$cay
Estimate of λ_F	2.04	0.02	−0.13	0.06
FM t	2.18	0.20	−0.43	3.12
95% CI by FM t	(0.20, 3.88)	(−0.20, 0.25)	(−0.70, 0.45)	(0.02, 0.09)
Shanken t	1.75	0.15	−0.31	2.25
95% CI by Shanken t	(−0.24, 4.32)	(−0.29, 0.34)	(−0.93, 0.68)	(0.01, 0.11)

Notes: The estimate of λ_F and the FM t-statistic results from the FM (1973) two-pass procedure. The t-statistic with the Shanken (1992) correction is also presented. 95% CIs are constructed by inverting the t-tests. For the consumption CAPM, we replicate the Kroencke (2017) study: the yearly unfiltered consumption growth in 1960–2014 is used for $\triangle c$, while 30 portfolios sorted by size, value, and investment, plus the market portfolio, are used as test assets. For the conditional consumption CAPM, we replicate the Lettau and Ludvigson (2001) study: the quarterly consumption growth $\triangle c$, the (lagged) consumption-wealth ratio cay, and their interaction in 1963Q3–1998Q3 are used as three risk factors, while 25 portfolios sorted by size and book-to-market are used as test assets.

finite-sample and asymptotic versions lead to almost identical *p*-values. This should not be surprising, since Theorem 6 shows that the FM-LM test does not suffer from the limited T versus large N problem. As long as T exceeds $2K$ by a large enough margin, the finite-sample adjusted FM-LM test is almost equivalent to its asymptotic counterpart.

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

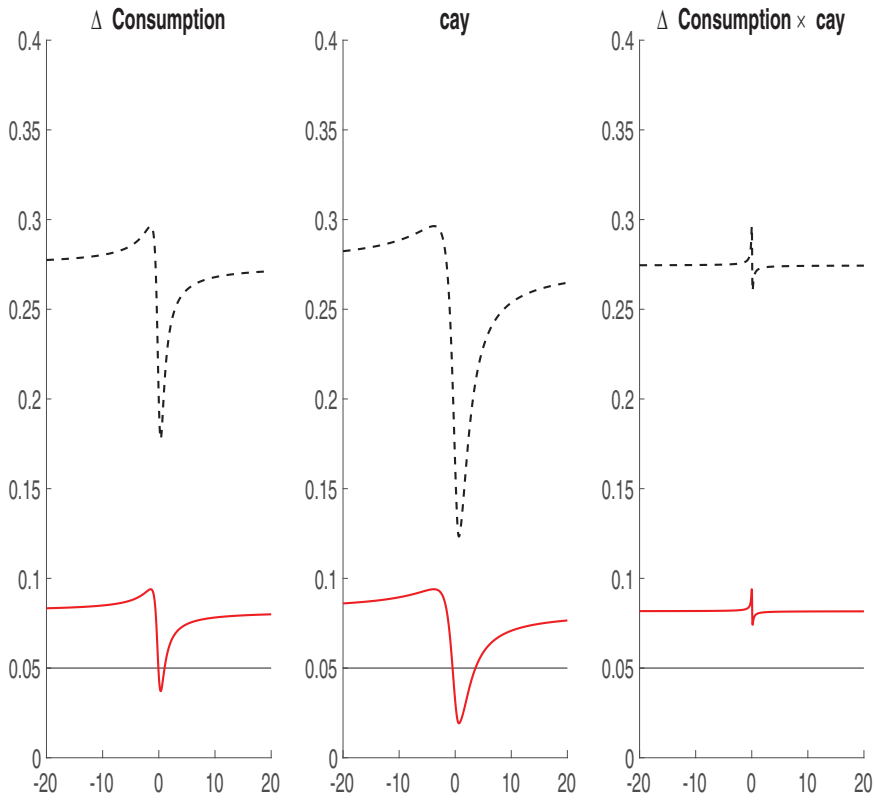

Figure 8 *p*-values of the sFAR test for testing the values of risk premia on the horizontal axis. Dashed black curves result from the sFAR test using the finite-sample *F*-critical values, while solid red curves of *p*-values result from the asymptotic χ^2-critical values. The test assets (25 portfolios sorted by size and book-to-market ratio) and three risk factors (Δc, *cay*, $\Delta c \times cay$) in 1963Q3–1998Q3 are taken from Lettau and Ludvigson (2001). The 5% benchmark line is also plotted.

Similar to Figure 7, Figure 8 provides *p*-value plots based on the sFAR test for each of the three risk premia in the conditional consumption CAPM in Lettau and Ludvigson (2001). They are used to construct the 95% confidence sets of each of the risk premia, to compare with the conventional ones presented in Panel B of Table 3.[8]

Figure 8 shows that the finite-sample adjusted sFAR test (dashed black) leads to unbounded 95% confidence sets of the risk premium on each of the three risk factors in the model ($\triangle c$: consumption growth; cay: the lagged consumption–wealth ratio; $\triangle c\times$ cay: their interaction). In contrast, 95% confidence sets based on the asymptotic sFAR test (solid red) are unbounded and disjoint for $\triangle c$ and cay, since only a bounded set of hypothesized risk

8 For our other proposed tests, the 95% confidence sets for the individual risk premium result from projecting the 95% joint confidence sets of all three risk premia on the separate axes reflecting each risk premium. We, therefore, refrain from using them to construct the 95% confidence sets for the individual risk premium.

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

premia are associated with p-values below 5%. Furthermore, for the interaction term $\triangle c \times$ cay, both the finite-sample and the asymptotic p-value curves lie above 5%, so they suggest an unbounded 95% confidence set. Lettau and Ludvigson (2001), on the other hand, report a significant FM t-statistic on the interaction term to help establish their conditional consumption CAPM, which we re-produce in Panel B of Table 3. In contrast with Lettau and Ludvigson (2001), the unbounded confidence sets implied by Figure 8 thus put the significant risk premium from the FM t-test under doubt; see also Kleibergen (2009) and Kleibergen and Zhan (2020).

Finally, it is worth emphasizing that the unbounded confidence sets of risk premia in Figures 7 and 8 are in line with the corresponding β estimates in Tables 1 and 2. As indicated by Tables 1 and 2, we cannot rule out that the full rank condition of (ι_N, β) is violated in a statistical sense. If the full rank condition does not hold, then the risk premia λ_F in Equation (1) are unidentified. Any real value can thus become possible for the unidentified λ_F, which consequently leads to the unbounded confidence sets in Figures 7 and 8.

4 Conclusions

We provide finite-sample distributions for the identification robust statistics testing risk premia in the beta representation of expected asset returns. By doing so, we make these tests suitable for empirical settings where sample sizes are often limited and the quality of risk factors is questionable. The size and power performances in Monte Carlo simulations show that these tests work favorably, and their empirical usage is illustrated using two well-known applications.

Acknowledgement

We thank Paolo Zaffaroni and Francisco Peñaranda for their discussion of the Hall White Memorial Lecture during the SoFiE 2019 Annual Meeting at Fudan University in Shanghai, which greatly improved the manuscript.

Supplementary Data

Supplemental data are available at https://www.datahostingsite.com

Appendix A: Proofs

In this Appendix, we provide the proofs for the analytical results stated in the main text. These proofs build on Lemmas 1 and 2 below.

Lemma 1. *For any given* $\lambda_{F,0}$,

$$\widetilde{\mathcal{B}} = \widehat{\mathcal{B}} - \left(\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}\right) \frac{\begin{pmatrix} 1 \\ -\lambda_{F,0} \end{pmatrix}' \begin{pmatrix} 1 & 0 \\ 0 & \widehat{\mathcal{Q}} \end{pmatrix} \begin{pmatrix} 0 \\ I_K \end{pmatrix}}{\begin{pmatrix} 1 \\ -\lambda_{F,0} \end{pmatrix}' \begin{pmatrix} 1 & 0 \\ 0 & \widehat{\mathcal{Q}} \end{pmatrix} \begin{pmatrix} 1 \\ -\lambda_{F,0} \end{pmatrix}}, \tag{25}$$

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

$$\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0} = \frac{1}{1+\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}}\left(\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}\right), \tag{26}$$

where $\overline{\mathcal{R}}$, $\widetilde{\mathcal{B}}$, $\widehat{\mathcal{B}}$, and $\widehat{\mathcal{Q}}$ are defined in the main text.

Proof of Lemma 1:

$$\widehat{\mathcal{Q}} = \left(\sum_t \overline{F}_t\overline{F}_t'\right)/T \tag{27}$$

$$\widehat{\mathcal{Q}}_{FF}(\lambda_{F,0}) = \left(\sum_t \left(\overline{F}_t + \lambda_{F,0}\right)\left(\overline{F}_t + \lambda_{F,0}\right)'\right)/T = \widehat{\mathcal{Q}} + \lambda_{F,0}\lambda_{F,0}', \tag{28}$$

where the last equality results from the fact that $\overline{F}_t$ are demeaned factors. Using $(A+aa\prime)^{-1} = A^{-1} - \frac{1}{1+a'A^{-1}a}A^{-1}aa'A^{-1}$, we have

$$\widehat{\mathcal{Q}}_{FF}(\lambda_{F,0})^{-1} = \left(\widehat{\mathcal{Q}} + \lambda_{F,0}\lambda_{F,0}'\right)^{-1} = \widehat{\mathcal{Q}}^{-1} - \frac{1}{1+\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}}\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}, \tag{29}$$

which implies

$$1 - \lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}(\lambda_{F,0})^{-1}\lambda_{F,0} = \frac{1}{1+\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}}. \tag{30}$$

Plugging Equation (29) into $\widetilde{\mathcal{B}}$, which can be rewritten as $\widetilde{\mathcal{B}} = \frac{1}{T}\mathcal{R}'\left(\iota_T\lambda_{F,0}' + F\right)\widehat{\mathcal{Q}}_{FF}(\lambda_{F,0})^{-1}$ with $F = \left(\overline{F}_1, \ldots, \overline{F}_T\right)'$ and $\mathcal{R} = (\mathcal{R}_1, \ldots, \mathcal{R}_T)'$, we get:

$$\begin{aligned}
\widetilde{\mathcal{B}} &= \frac{1}{T}\mathcal{R}'\left(\iota_T\lambda_{F,0}' + F\right)\widehat{\mathcal{Q}}_{FF}(\lambda_{F,0})^{-1} \\
&= \frac{1}{T}\mathcal{R}'F\left(\widehat{\mathcal{Q}}^{-1} - \frac{1}{1+\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}}\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\right) + \overline{\mathcal{R}}\lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}(\lambda_{F,0})^{-1} \\
&= \widehat{\mathcal{B}} - \frac{1}{1+\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}}\widehat{\mathcal{B}}\lambda_{F,0}\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1} + \overline{\mathcal{R}}\lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}(\lambda_{F,0})^{-1} \\
&= \widehat{\mathcal{B}} - \frac{1}{1+\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}}\widehat{\mathcal{B}}\lambda_{F,0}\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1} + \overline{\mathcal{R}}\lambda_{F,0}'\left(\widehat{\mathcal{Q}}^{-1} - \frac{1}{1+\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}}\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\right), \\
&= \widehat{\mathcal{B}} - \frac{1}{1+\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}}\widehat{\mathcal{B}}\lambda_{F,0}\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1} + \frac{1}{1+\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}}\overline{\mathcal{R}}\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1} \\
&= \widehat{\mathcal{B}} - \left(\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}\right)\frac{-\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}}{1+\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}}
\end{aligned} \tag{31}$$

which then implies the following Equations (32) and (33):

$$\widetilde{\mathcal{B}} = \widehat{\mathcal{B}} - \left(\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}\right)\frac{\begin{pmatrix}1\\-\lambda_{F,0}\end{pmatrix}'\begin{pmatrix}1 & 0\\0 & \widehat{\mathcal{Q}}^{-1}\end{pmatrix}\begin{pmatrix}0\\I_K\end{pmatrix}}{\begin{pmatrix}1\\-\lambda_{F,0}\end{pmatrix}'\begin{pmatrix}1 & 0\\0 & \widehat{\mathcal{Q}}^{-1}\end{pmatrix}\begin{pmatrix}1\\-\lambda_{F,0}\end{pmatrix}} \tag{32}$$

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

$$\widetilde{\mathcal{B}}\lambda_{F,0} = \widehat{\mathcal{B}}\lambda_{F,0} - \left(\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}\right)\frac{-\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}}{1+\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}}. \tag{33}$$

From Equation (33), it results that

$$\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0} = \frac{1}{1+\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}}\left(\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}\right). \tag{34}$$

Q.E.D.

Remarks. Lemma 1 also implies

$$\widetilde{\mathcal{B}} - \left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1} = \widehat{\mathcal{B}}. \tag{35}$$

If the model is correctly specified, the term $\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)$ is asymptotically zero, so the two estimators $\widetilde{\mathcal{B}}$ and $\widehat{\mathcal{B}}$ are asymptotically equivalent.

Lemma 2. *Under* $H_0 : \lambda_F = \lambda_{F,0}$ *and Assumptions 1 and 2,*

$$\sqrt{T}\left(\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}\right) \sim N\left(0, \left(1+\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}\right)\otimes\Sigma\right), \tag{36}$$

$$\sqrt{T}\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right) \sim N\left(0, \left(1-\lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}(\lambda_{F,0})^{-1}\lambda_{F,0}\right)\otimes\Sigma\right), \tag{37}$$

$$\left(\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}\right)\perp\widetilde{\mathcal{B}}, \tag{38}$$

$$\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)\perp\widetilde{\mathcal{B}}, \tag{39}$$

$$\left(\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}\right)\perp\widehat{\Sigma}, \tag{40}$$

$$\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)\perp\widehat{\Sigma}, \tag{41}$$

$$\widehat{\Sigma} = \frac{1}{T-K-1}\sum_t\left(\overline{\mathcal{R}}_t - \widehat{B}\overline{F}_t\right)\left(\overline{\mathcal{R}}_t - \widehat{B}\overline{F}_t\right)' \sim \frac{\mathcal{W}_{N-1}(T-K-1,\Sigma)}{T-K-1}, \tag{42}$$

$$\overline{\mathcal{R}}\perp\widehat{\mathcal{B}},\ \overline{\mathcal{R}}\perp\widehat{\Sigma},\ \widehat{\mathcal{B}}\perp\widehat{\Sigma},\ \widetilde{\mathcal{B}}\perp\widehat{\Sigma}. \tag{43}$$

Proof of Lemma 2: We only need to show Equations (36), (39), (40), and (42). The proof of these results, together with Lemma 1, implies the rest of Lemma 2.

(1) Show $\sqrt{T}\left(\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}\right) \sim N\left(0, \left(1+\lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}\right)\otimes\Sigma\right)$:

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

$$\begin{aligned}\sqrt{T}\left(\overline{\mathcal{R}}-\widehat{\mathcal{B}}\lambda_{F,0}\right) &= \sqrt{T}\left(\overline{\mathcal{R}}-\mathcal{B}\lambda_{F,0}-\left(\widehat{\mathcal{B}}-\mathcal{B}\right)\lambda_{F,0}\right)\\ &= \sqrt{T}\left(\overline{\mathcal{U}}-\left(\widehat{\mathcal{B}}-\mathcal{B}\right)\lambda_{F,0}\right)\\ &= \sqrt{T}\left(\mathcal{U}'\iota_T\left(\iota_T'\iota_T\right)^{-1}-\mathcal{U}'F\left(F'F\right)^{-1}\lambda_{F,0}\right),\\ &= \sqrt{T}\mathcal{U}'X\left(X'X\right)^{-1}\begin{pmatrix}1 & -\lambda_{F,0}'\end{pmatrix}'\end{aligned} \tag{44}$$

where $\mathcal{U}=(\mathcal{U}_1,\ldots,\mathcal{U}_T)'$, $F=\left(\overline{F}_1,\ldots,\overline{F}_T\right)'$, and $X=(\iota_T,F)$. It then follows that

$$\begin{aligned}\sqrt{T}\left(\overline{\mathcal{R}}-\widehat{\mathcal{B}}\lambda_{F,0}\right) &= \sqrt{T}\mathcal{U}'X\left(X'X\right)^{-1}\begin{pmatrix}1 & -\lambda_{F,0}'\end{pmatrix}'\\ &= \left(\begin{pmatrix}1 & -\lambda_{F,0}'\end{pmatrix}\left(X'X/T\right)^{-1}\otimes I_{N-1}\right)\left(\frac{1}{\sqrt{T}}\sum_t X_t\otimes\mathcal{U}_t\right)\end{aligned} \tag{45}$$

and we thus get Equation (36) by using

$$\frac{1}{\sqrt{T}}\sum_t X_t\otimes\mathcal{U}_t\sim N\left(0,\left(X'X\right)/T\otimes\Sigma\right). \tag{46}$$

(2) Show $\left(\overline{\mathcal{R}}-\widetilde{\mathcal{B}}\lambda_{F,0}\right)\perp\widetilde{\mathcal{B}}$, where $\perp$ stands for independence:

We prove the claim by showing

$$\begin{pmatrix}\overline{\mathcal{R}}-\widetilde{\mathcal{B}}\lambda_{F,0}\\ \operatorname{vec}\left(\widetilde{\mathcal{B}}-\mathcal{B}\right)\end{pmatrix}\sim\frac{1}{\sqrt{T}}\begin{pmatrix}\psi_{\mathcal{R}}\\ \psi_{\mathcal{B}}\end{pmatrix}, \tag{47}$$

where $\psi_{\mathcal{R}}\perp\psi_{\mathcal{B}}$ and they are $(N-1)\times 1$ and $(N-1)K\times 1$ normally distributed random vectors with mean zero and covariance matrices $\left(1-\lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\lambda_{F,0}\right)\otimes\Sigma$ and $\widehat{\mathcal{Q}}_{FF}(\lambda_{F,0})^{-1}\otimes\Sigma$, respectively. To show it, we start with

$$\begin{pmatrix}\overline{\mathcal{R}}-\widetilde{\mathcal{B}}\lambda_{F,0}\\ \operatorname{vec}\left(\widetilde{\mathcal{B}}-\mathcal{B}\right)\end{pmatrix}=\begin{pmatrix}\overline{\mathcal{R}}-\mathcal{B}\lambda_{F,0}-\left(\lambda_{F,0}'\otimes I_{N-1}\right)\operatorname{vec}\left(\widetilde{\mathcal{B}}-\mathcal{B}\right)\\ \operatorname{vec}\left(\widetilde{\mathcal{B}}-\mathcal{B}\right)\end{pmatrix}, \tag{48}$$

where

$$\overline{\mathcal{R}}-\mathcal{B}\lambda_{F,0}=\overline{\mathcal{U}}=\mathcal{U}'\iota_T\left(\iota_T'\iota_T\right)^{-1}=\left(\left(\iota_T'\iota_T\right)^{-1}\iota_T'\otimes I_{N-1}\right)\operatorname{vec}\left(\mathcal{U}'\right) \tag{49}$$

$$\operatorname{vec}\left(\widetilde{\mathcal{B}}-\mathcal{B}\right)=\frac{1}{T}\left(\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\left(F'+\lambda_{F,0}\iota_T'\right)\otimes I_{N-1}\right)\operatorname{vec}\left(\mathcal{U}'\right). \tag{50}$$

We therefore have

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

$$\begin{pmatrix} \overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0} \\ \mathrm{vec}(\widetilde{\mathcal{B}} - \mathcal{B}) \end{pmatrix} \sim \begin{pmatrix} I_{N-1} & -\left(\lambda_{F,0}' \otimes I_{N-1}\right) \\ 0 & I_{K(N-1)} \end{pmatrix} \begin{pmatrix} \left(\iota_T'\iota_T\right)^{-1}\iota_T' \otimes I_{N-1} \\ \frac{1}{T}\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\left(F + \iota_T\lambda_{F,0}'\right)' \otimes I_{N-1} \end{pmatrix} \mathrm{vec}(\mathcal{U}')$$
$$\sim \begin{pmatrix} \left(\left(\iota_T'\iota_T\right)^{-1}\iota_T' - \frac{1}{T}\lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\left(F + \iota_T\lambda_{F,0}'\right)'\right) \otimes I_{N-1} \\ \frac{1}{T}\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\left(F + \iota_T\lambda_{F,0}'\right)' \otimes I_{N-1} \end{pmatrix} \mathrm{vec}(\mathcal{U}')$$
$$\sim \frac{1}{T}\begin{pmatrix} \left(\iota_T' - \lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\left(F + \iota_T\lambda_{F,0}'\right)'\right) \otimes I_{N-1} \\ \widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\left(F + \iota_T\lambda_{F,0}'\right)' \otimes I_{N-1} \end{pmatrix} \mathrm{vec}(\mathcal{U}') \tag{51}$$

Note that $(\iota_T' - \lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})(F + \iota_T\lambda_{F,0}')') \times (\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})(F + \iota_T\lambda_{F,0}')')' = 0$, which implies $\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0} \perp \widetilde{\mathcal{B}}$.

(3) Show $(\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}) \perp \widehat{\Sigma}$:

Note that $\widehat{\Sigma} = \mathcal{U}'M_X\mathcal{U}/(T - K - 1)$ with $M_X = I_T - X(X\dot{X})^{-1}X$, while $(\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}) = \mathcal{U}'X(X'X)^{-1}(1 \quad -\lambda_{F,0}')'$ as shown in Equation (44). The independence thus results from $\mathcal{U}'X \perp \mathcal{U}'M_X$, since

$$\mathbb{E}(\mathrm{vec}(\mathcal{U}'X)'\mathrm{vec}(\mathcal{U}'M_X)) = \mathbb{E}(((X' \otimes I_{N-1})\mathrm{vec}(\mathcal{U}'))'(M_X \otimes I_{N-1})\mathrm{vec}(\mathcal{U}')) = 0. \tag{52}$$

(4) Show $\widehat{\Sigma} \sim \frac{\mathcal{W}_{N-1}(T-K-1,\Sigma)}{T-K-1}$:

$\widehat{\Sigma} = \mathcal{U}'M_X\mathcal{U}/(T - K - 1)$, where $M_X = I_T - X(X\dot{X})^{-1}X$ is the projection matrix with rank $T - K - 1$, $\mathcal{U} = (\mathcal{U}_1, \ldots, \mathcal{U}_T)'$ with $\mathcal{U}_t \sim N(0_{N-1}, \Sigma)$. Using Theorem 2 of Mathew and Nordstrom (1997), it follows that

$$\widehat{\Sigma} = \mathcal{U}'M_X\mathcal{U}/(T - K - 1) \sim \mathcal{W}_{N-1}(T - K - 1, \Sigma)/(T - K - 1). \tag{53}$$

Q.E.D.

Proof of Theorem 1: Equivalence of FAR and H statistics results from Equations (26) and (30):

$$\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0} = \frac{1}{1 + \lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}}(\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}) \tag{54}$$

$$1 - \lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}(\lambda_{F,0})^{-1}\lambda_{F,0} = \frac{1}{1 + \lambda_{F,0}'\widehat{\mathcal{Q}}^{-1}\lambda_{F,0}}. \tag{55}$$

Assumptions 1 and 2 imply that Assumption 1 in Kleibergen (2009) holds. Theorem 1 thus results from Theorem 5 in Kleibergen (2009).

Q.E.D.

Proof of Theorem 2: From Lemma 2, we know:

$$\sqrt{\frac{T}{1 - \lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}(\lambda_{F,0})^{-1}\lambda_{F,0}}}(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}) \sim N(0, \Sigma) \tag{56}$$

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

$$(\overline{\mathcal{R}} - \widehat{\mathcal{B}}\lambda_{F,0}) \perp \widehat{\Sigma} \tag{57}$$

$$\widehat{\Sigma} = \frac{1}{T-K-1}\sum_t (\overline{\mathcal{R}}_t - \widehat{B}\overline{F}_t)(\overline{\mathcal{R}}_t - \widehat{B}\overline{F}_t)' \sim \frac{\mathcal{W}_{N-1}(T-K-1,\Sigma)}{T-K-1}. \tag{58}$$

With these results, the $F_{N-1,T-K-N+1}$ distribution for the finite-sample FAR statistic results from Theorem 3.2.13 from Muirhead (2009).

Q.E.D.

Proof of Theorem 3: Similar to the proof of Theorems 1 and 3 results from Theorem 6 in Kleibergen (2009).

Q.E.D.

Proof of Theorem 4: Let $\widetilde{\mathcal{B}}_\perp$ be an arbitrary $(N-1)\times(N-K-1)$ matrix such that $(\widetilde{\mathcal{B}}, \widetilde{\mathcal{B}}_\perp)$ is of full rank and $\widetilde{\mathcal{B}}_\perp'\widetilde{\mathcal{B}} = 0$. The projection matrix $M_{\widehat{\Sigma}^{-\frac{1}{2}}\widetilde{\mathcal{B}}} = I_{N-1} - P_{\widehat{\Sigma}^{-\frac{1}{2}}\widetilde{\mathcal{B}}}$ can be written as $\widehat{\Sigma}^{-\frac{1'}{2}}\widetilde{\mathcal{B}}_\perp(\widetilde{\mathcal{B}}_\perp'\widehat{\Sigma}\widetilde{\mathcal{B}}_\perp)^{-1}\widetilde{\mathcal{B}}_\perp'\widehat{\Sigma}^{-\frac{1}{2}}$. The JGLS($\lambda_{F,0}$) statistic can thus be written as

$$\begin{aligned}
&\text{JGLS}(\lambda_{F,0}) \\
&= \frac{T}{1-\lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\lambda_{F,0}}(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0})'\widehat{\Sigma}^{-\frac{1'}{2}}(I_{N-1} - P_{\widehat{\Sigma}^{-\frac{1}{2}}\widetilde{\mathcal{B}}})\widehat{\Sigma}^{-\frac{1}{2}}(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}) \\
&= \frac{T}{1-\lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\lambda_{F,0}}(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0})'\widetilde{\mathcal{B}}_\perp(\widetilde{\mathcal{B}}_\perp'\widehat{\Sigma}\widetilde{\mathcal{B}}_\perp)^{-1}\widetilde{\mathcal{B}}_\perp'(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0})
\end{aligned} \tag{59}$$

Conditional on $\widetilde{\mathcal{B}}$, Lemma 2 implies:

$$\sqrt{\frac{T}{1-\lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\lambda_{F,0}}}\widetilde{\mathcal{B}}_\perp'(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}) \sim N(0, \widetilde{\mathcal{B}}_\perp'\Sigma\widetilde{\mathcal{B}}_\perp) \tag{60}$$

$$\widetilde{\mathcal{B}}_\perp'\widehat{\Sigma}\widetilde{\mathcal{B}}_\perp \sim \frac{\mathcal{W}_{N-K-1}(T-K-1, (\widetilde{\mathcal{B}}_\perp)'\Sigma(\widetilde{\mathcal{B}}_\perp))}{T-K-1} \tag{61}$$

$$\widetilde{\mathcal{B}}_\perp'\widehat{\Sigma}\widetilde{\mathcal{B}}_\perp \perp \widetilde{\mathcal{B}}_\perp'(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}) \tag{62}$$

since $\widetilde{\mathcal{B}}_\perp$ only depends on $\widetilde{\mathcal{B}}$ by construction. These results imply that $\frac{T-N+1}{(T-K-1)(N-K-1)} \times \text{JGLS}(\lambda_{F,0}) \sim F_{N-K-1,T-N+1}$. Note that GLS-LM($\lambda_{F,0}$) = FAR($\lambda_{F,0}$) − JGLS($\lambda_{F,0}$), where FAR($\lambda_{F,0}$) and JGLS($\lambda_{F,0}$), once properly scaled, follow *F*-distributions with different degrees of freedom. The finite distribution of GLS-LM($\lambda_{F,0}$) is thus the difference of the finite-sample distributions of the FAR($\lambda_{F,0}$) and JGLS($\lambda_{F,0}$) statistics. From the proof of Lemma 2, we know

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

$$\sqrt{T}\begin{pmatrix}\overline{\mathcal{R}}-\widetilde{\mathcal{B}}\lambda_{F,0}\\ \\ \mathrm{vec}\left(\widetilde{\mathcal{B}}-\mathcal{B}\right)\end{pmatrix}\sim\begin{pmatrix}\psi_{\mathcal{R}}\\ \\ \psi_{\mathcal{B}}\end{pmatrix}$$
$$\sim N\left(0,\begin{pmatrix}\left(1-\lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\lambda_{F,0}\right)\otimes\Sigma & 0\\ \\ 0 & \widehat{\mathcal{Q}}_{FF}(\lambda_{F,0})^{-1}\otimes\Sigma\end{pmatrix}\right) \tag{63}$$

and $\widehat{\Sigma}\sim\frac{\mathcal{W}_{N-1}(T-K-1,\Sigma)}{T-K-1}$, $\overline{\mathcal{R}}-\widetilde{\mathcal{B}}\lambda_{F,0}\perp\widehat{\Sigma}$. Therefore, we have $\mathrm{FAR}(\lambda_{F,0})\sim\psi'\mathcal{W}^{-1}\psi$, where $\psi\sim N(0,I_{N-1})$, $\mathcal{W}\sim\frac{\mathcal{W}_{N-1}(T-K-1,I_{N-1})}{T-K-1}$. As $\overline{\mathcal{R}}-\widetilde{\mathcal{B}}\lambda_{F,0},\widehat{\Sigma}$ are independent from $\widetilde{\mathcal{B}}_{\perp}$, we have $\mathrm{JGLS}(\lambda_{F,0})\sim\psi'C(C'\mathcal{W}C)^{-1}C'\psi$ independent of C from Equation (59).*Q.E.D.*

Proof of Theorem 5: Similar to the proof of Theorems 1 and 5 results from Theorem 7 in Kleibergen (2009).

Q.E.D.

Proof of Theorem 6: Conditional on $\widetilde{\mathcal{B}}$, Lemma 2 implies:

$$\sqrt{\frac{T}{1-\lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\lambda_{F,0}}}\widetilde{\mathcal{B}}'\left(\overline{\mathcal{R}}-\widetilde{\mathcal{B}}\lambda_{F,0}\right)\sim N\left(0,\widetilde{\mathcal{B}}'\Sigma\widetilde{\mathcal{B}}\right) \tag{64}$$

$$\widetilde{\mathcal{B}}'\widehat{\Sigma}\widetilde{\mathcal{B}}\sim\frac{\mathcal{W}_K\left(T-K-1,\widetilde{\mathcal{B}}'\Sigma\widetilde{\mathcal{B}}\right)}{T-K-1} \tag{65}$$

$$\widetilde{\mathcal{B}}'\widehat{\Sigma}\widetilde{\mathcal{B}}\perp\widetilde{\mathcal{B}}'\left(\overline{\mathcal{R}}-\widetilde{\mathcal{B}}\lambda_{F,0}\right). \tag{66}$$

These results imply that $\frac{T-2K}{(T-K-1)K}\times\text{FM-LM}(\lambda_{F,0})\sim\mathrm{F_{K,T-2K}}$, and this distribution does not depend on $\widetilde{\mathcal{B}}$. Given $\mathrm{JFM}(\lambda_{F,0})=\mathrm{FAR}(\lambda_{F,0})-\text{FM-LM}(\lambda_{F,0})$, the finite distribution of $\mathrm{JFM}(\lambda_{F,0})$ is thus the difference of the finite-sample distributions of the $\mathrm{FAR}(\lambda_{F,0})$ and $\text{FM-LM}(\lambda_{F,0})$ statistics: $\mathrm{JFM}(\lambda_{F,0})\sim\psi'\mathcal{W}^{-1}\psi-\psi'\mathbf{C}(\mathbf{C}'\mathcal{W}\mathbf{C})^{-1}\mathbf{C}'\psi$. The proof is similar to that for GLS-LM in Theorem 4, except for that $\mathbf{C}$ is an arbitrary $(N-1)\times K$ matrix of full rank.
Q.E.D.

Proof of Theorem 7: The i.i.d. condition in Assumption 1 implies that r in the sCLR statistic is independent from the other parts. Theorem 7 thus results from Lemma 2 and Theorem 4.

Q.E.D.

Proof of Proposition 8: Proposition 1 in Online Appendix B shows that

$$\frac{\partial\mathrm{FAR}(\lambda_F)}{\partial\lambda_F'}=-2T\left(\overline{\mathcal{R}}-\widetilde{\mathcal{B}}\lambda_F\right)'\widehat{\Sigma}^{-1}\widetilde{\mathcal{B}} \tag{67}$$

and thus

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

$$\text{GLS-LM}(\lambda_F) = \frac{1+\lambda_F'\widehat{\mathcal{Q}}^{-1}\lambda_F}{4T}\frac{\partial \text{FAR}(\lambda_F)}{\partial \lambda_F'}\left(\widetilde{\mathcal{B}}'\widehat{\Sigma}^{-1}\widetilde{\mathcal{B}}\right)^{-1}\left(\frac{\partial \text{FAR}(\lambda_F)}{\partial \lambda_F'}\right)'. \tag{68}$$

Similarly, we have

$$\frac{\partial \text{FAR}(\lambda_F)}{\partial \lambda_2'} = -2T(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_F)'\widehat{\Sigma}^{-1}\widetilde{\mathcal{B}}_2 \tag{69}$$

which then leads to the construction of $\widetilde{\text{GLS-LM}}\left(\left(\lambda_{F,1}^{0\prime}, \lambda_2^{0\prime}\right)'\right)$:

$$\begin{aligned}
&\widetilde{\text{GLS-LM}}\left(\left(\lambda_{F,1}^{0\prime}, \lambda_2^{0\prime}\right)'\right)\\
&= \frac{1+\lambda_F'\widehat{\mathcal{Q}}^{-1}\lambda_F}{4T}\frac{\partial \text{FAR}(\lambda_F)}{\partial \lambda_2'}|_{\lambda_F=\lambda_F^0}\left(\widetilde{\mathcal{B}}_2'\widehat{\Sigma}^{-1}\widetilde{\mathcal{B}}_2\right)^{-1}\left(\frac{\partial \text{FAR}(\lambda_F)}{\partial \lambda_2'}|_{\lambda_F=\lambda_F^0}\right)'\\
&= \frac{T}{1-\lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\lambda_{F,0}}\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)'\widehat{\Sigma}^{-\frac{1}{2}'}P_{\widehat{\Sigma}^{-\frac{1}{2}}\widetilde{\mathcal{B}}_2}\widehat{\Sigma}^{-\frac{1}{2}}\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right).
\end{aligned}$$

For $\widetilde{\text{JGLS}}((\lambda_{F,1}^{0\prime}, \lambda_2^{0\prime})')$:

$$\begin{aligned}
&\widetilde{\text{JGLS}}\left(\left(\lambda_{F,1}^{0\prime}, \lambda_2^{0\prime}\right)'\right) = \text{FAR}\left(\left(\lambda_{F,1}^{0\prime}, \lambda_2^{0\prime}\right)'\right) - \widetilde{\text{GLS-LM}}\left(\left(\lambda_{F,1}^{0\prime}, \lambda_2^{0\prime}\right)'\right)\\
&= \frac{T}{1-\lambda_{F,0}'\widehat{\mathcal{Q}}_{FF}^{-1}(\lambda_{F,0})\lambda_{F,0}}\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right)'\widehat{\Sigma}^{-\frac{1}{2}'}\left(I_{N-1} - P_{\widehat{\Sigma}^{-\frac{1}{2}}\widetilde{\mathcal{B}}_2}\right)\widehat{\Sigma}^{-\frac{1}{2}}\left(\overline{\mathcal{R}} - \widetilde{\mathcal{B}}\lambda_{F,0}\right).
\end{aligned}$$

Denote $\widehat{\lambda}_2(\lambda_{F,1}^0) = \text{argmin}_{\lambda_2}\text{FAR}((\lambda_{F,1}^0, \lambda_2')')$. Thus, $\widetilde{\text{GLS-LM}}((\lambda_{F,1}^0, \widehat{\lambda}_2'(\lambda_{F,1}^0))') = 0$ by construction. As a result, once we establish $\widetilde{\text{JGLS}}((\lambda_{F,1}^0, \widehat{\lambda}_2'(\lambda_{F,1}^0))') \leq \widetilde{\text{JGLS}}((\lambda_{F,1}^0, \lambda_2^{0\prime})')$, it then implies

$$\text{sFAR}\left(\lambda_{F,1}^0\right) = \widetilde{\text{JGLS}}\left(\left(\lambda_{F,1}^0, \widehat{\lambda}_2'\left(\lambda_{F,1}^0\right)\right)'\right) \leq \widetilde{\text{JGLS}}\left(\left(\lambda_{F,1}^0, \lambda_2^{0\prime}\right)'\right).$$

Proposition 3 in Online Appendix B shows that

$$\frac{\partial \text{JGLS}}{\partial \lambda_F'}(\lambda_F) = \frac{1}{T}\text{JGLS}(\lambda_F)\frac{\partial \text{FAR}}{\partial \lambda_F'}(\lambda_F)\left(\widetilde{\mathcal{B}}'\widehat{\Sigma}^{-1}\widetilde{\mathcal{B}}\right)^{-1}\left(\widehat{\mathcal{Q}}^{-1} - \frac{\widehat{\mathcal{Q}}^{-1}\lambda_F\lambda_F'\widehat{\mathcal{Q}}^{-1}}{1+\lambda_F'\widehat{\mathcal{Q}}^{-1}\lambda_F}\right)$$

and in the similar fashion, we have

$$\frac{\partial \widetilde{\text{JGLS}}}{\partial \lambda_2^{0\prime}}\left(\left(\lambda_{F,1}^0, \lambda_2^{0\prime}\right)'\right) = \frac{1}{T}\text{JGLS}\left(\lambda_F^{0\prime}\right)\frac{\partial \text{FAR}}{\partial \lambda_2^{0\prime}}\left(\lambda_F^{0\prime}\right)\left(\widetilde{\mathcal{B}}_2'\widehat{\Sigma}^{-1}\widetilde{\mathcal{B}}_2\right)^{-1}C'\left(\widehat{\mathcal{Q}}^{-1} - \frac{\widehat{\mathcal{Q}}^{-1}\lambda_F\lambda_F'\widehat{\mathcal{Q}}^{-1}}{1+\lambda_F'\widehat{\mathcal{Q}}^{-1}\lambda_F}\right)C$$

with $C = (0 \vdots I_{K-1})$. In practice, we deal with $N > K+1$ and thus without loss of generality we may safely assume that almost surely, zeros of $\frac{\partial \widetilde{\text{JGLS}}}{\partial \lambda_2'}((\lambda_{F,1}^0, \lambda_2')')$ (as a function of λ_2) are zeros of $\frac{\partial \text{FAR}}{\partial \lambda_2'}((\lambda_{F,1}^0, \lambda_2')')$ (as a function of λ_2). Given the facts that the stationary points of $\text{FAR}((\lambda_{F,1}^0, \lambda_2')')$ and $\widetilde{\text{JGLS}}((\lambda_{F,1}^0, \lambda_2')')$ (as functions of λ_2) coincide, they achieve the minimum at the same point. Therefore, the inequality $\widetilde{\text{JGLS}}((\lambda_{F,1}^0, \widehat{\lambda}_2'(\lambda_{F,1}^0))') \leq \widetilde{\text{JGLS}}((\lambda_{F,1}^0, \lambda_2^{0\prime})')$ holds.

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

Q.E.D.

Proof of Theorem 9: This is a direct result of Proposition 8 and Theorem 3. See also Guggenberger et al. (2012), Gospodinov, Kan, and Robotti (2017).

Q.E.D.

Proof of Theorem 10: This is a direct result of Proposition 8 and the fact that

$$\frac{T-N}{(T-K-1)(N-K)} \times \widetilde{\text{JGLS}}\left(\left(\lambda_{F,1}^{0}, \lambda_{2}^{0\prime}\right)'\right) \sim F_{N-K,T-N},$$

which is indicated by the proof of Theorem 4.*Q.E.D.*

Proof of Theorem 11: We know that $\widehat{\mathcal{B}} \sim N\left(B, \mathcal{Q}_{FF}^{-1} \otimes \Sigma\right)$. If B is of lower rank than K, then the null is equivalent to that there exists non-zero λ such that $B\lambda = 0$. If the first entry in λ is non-zero, there exists λ_0 such that $B\left(1, \lambda_0'\right)' = 0$. Denote

$$\widetilde{\mathcal{B}}_2 = \widehat{\mathcal{B}}_2 + \widehat{\mathcal{B}}\left(1, -\lambda_0'\right)' A$$

$$rk(\lambda) = \frac{T}{\lambda' Q \lambda} (\widehat{\mathcal{B}}\lambda)' \widehat{\Sigma}^{-1} (\widehat{\mathcal{B}}\lambda)$$

$$rk_F(\lambda_0) = \frac{T}{\left(1, -\lambda_0'\right) Q \left(1, -\lambda_0'\right)'} \left(\widehat{\mathcal{B}}\left(1, -\lambda_0'\right) I\right)' \widehat{\Sigma}^{-1} \left(\widehat{\mathcal{B}}\left(1, -\lambda_0'\right)'\right)$$

$$\underset{G}{\widetilde{rk}}(\lambda_0) = \frac{T}{\left(1, -\lambda_0'\right) Q \left(1, -\lambda_0'\right)'} \left(\widehat{\mathcal{B}}\left(1, -\lambda_0'\right) I\right)' \widehat{\Sigma}^{-\frac{1}{2}'} P_{\widehat{\Sigma}^{-\frac{1}{2}} \widetilde{\mathcal{B}}_2} \widehat{\Sigma}^{-\frac{1}{2}} \left(\widehat{\mathcal{B}}\left(1, -\lambda_0'\right)'\right)$$

$$\underset{J}{\widetilde{rk}}(\lambda_0) = rk_F(\lambda_0) - \underset{G}{\widetilde{rk}}(\lambda_0)$$

with $Q = \mathcal{Q}_{FF}^{-1}$, $A = \dfrac{\begin{pmatrix} 1 \\ -\lambda_0 \end{pmatrix}' Q \begin{pmatrix} 0 \\ I_{K-1} \end{pmatrix}}{\begin{pmatrix} 1 \\ -\lambda_0 \end{pmatrix}' Q \begin{pmatrix} 1 \\ -\lambda_0 \end{pmatrix}}$. By construction, $rk = \min_{\lambda} rk(\lambda) \leq \min_{\lambda_0} rk_F(\lambda_0)$. Then the result follows the proof of Theorem 10 by noticing $\min_{\lambda_0} rk_F(\lambda_0) \leq \underset{J}{\widetilde{rk}}(\lambda_0)$ and $\min_{\lambda} rk(\lambda)$ equals T times the minimal root of the following polynomial, which is the canonical correlation rank test

$$|\mu Q - \widehat{\mathcal{B}}' \widehat{\Sigma}^{-1} \widehat{\mathcal{B}}| = 0.$$

Q.E.D.

References

Anderson, T. W. 1951. Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions. *The Annals of Mathematical Statistics* 22: 327–351.

Anderson, T. W., and H. Rubin. 1949. Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. *The Annals of Mathematical Statistics* 20: 46–63.

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

Andrews, D. W., M. J. Moreira, and J. H. Stock. 2006. Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression. *Econometrica* 74: 715–752.

Beaulieu, M. C., J. M. Dufour, and L. Khalaf. 2013. Identification-Robust Estimation and Testing of the Zero-Beta CAPM. *The Review of Economic Studies* 80: 892–924.

Bekker, P. 1994. Alternative Approximations to the Distributions of Instrumental Variable Estimators. *Econometrica* 62: 657–681.

Bekker, P., and F. Kleibergen. 2003. Finite-Sample Instrumental Variables Inference Using an Asymptotically Pivotal Statistic. *Econometric Theory* 19: 744–753.

Cochrane, J. H. 2001. *Asset Pricing*. Princeton, NJ: Princeton University Press.

Dufour, J. M. 1997. Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models. *Econometrica* 65: 1365–1387.

Fama, E. F., and K. French. 1992. The Cross-Section of Expected Stock Returns. *The Journal of Finance* 47: 427–465.

Fama, E. F., and J. D. MacBeth. 1973. Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy* 81: 607–636.

Gibbons, M. R., S. A. Ross, and J. Shanken. 1989. A Test of the Efficiency of a Given Portfolio. *Econometrica* 57: 1121–1152.

Gospodinov, N., R. Kan, and C. Robotti. 2017. Spurious Inference in Reduced-Rank Asset-Pricing Models. *Econometrica* 85: 1613–1628.

Guggenberger, P., F. Kleibergen, S. Mavroeidis, and L. Chen. 2012. On the Asymptotic Sizes of Subset Anderson–Rubin and Lagrange Multiplier Tests in Linear Instrumental Variables Regression. *Econometrica* 80: 2649–2666.

Jagannathan, R., and Y. Wang. 2007. Lazy Investors, Discretionary Consumption, and the Cross-Section of Stock Returns. *The Journal of Finance* 62: 1623–1661.

Kan, R., C. Robotti, and J. Shanken. 2013. Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology. *The Journal of Finance* 68: 2617–2649.

Kan, R., and C. Zhang. 1999. Two-Pass Tests of Asset Pricing Models with Useless Factors. *The Journal of Finance* 54: 203–235.

Kan, R., and G. Zhou. 2004. Hansen–Jagannathan Distance: Geometry and Exact Distribution. Working paper.

Khalaf, L., and H. Schaller. 2016. Identification and Inference in Two-Pass Asset Pricing Models. *Journal of Economic Dynamics and Control* 70: 165–177.

Kleibergen, F. 2002. Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression. *Econometrica* 70: 1781–1803.

Kleibergen, F. 2009. Tests of Risk Premia in Linear Factor Models. *Journal of Econometrics* 149: 149–173.

Kleibergen, F., and R. Paap. 2006. Generalized Reduced Rank Tests Using the Singular Value Decomposition. *Journal of Econometrics* 133: 97–126.

Kleibergen, F., and Z. Zhan. 2015. Unexplained Factors and Their Effects on Second Pass R-Squared's. *Journal of Econometrics* 189: 101–116.

Kleibergen, F., and Z. Zhan. 2018. Identification-Robust Inference on Risk Premia of Mimicking Portfolios of Non-Traded Factors. *Journal of Financial Econometrics* 16: 155–190.

Kleibergen, F., and Z. Zhan. 2020. Robust Inference for Consumption-Based Asset Pricing. *The Journal of Finance* 75: 507–550.

Kleibergen, F., and Z. Zhan. 2021. Double Robust Inference for Continuous Updating GMM. Working paper. *arXiv preprint arXiv:2105.08345*.

Kroencke, T. A. 2017. Asset Pricing without Garbage. *The Journal of Finance* 72: 47–98.

Lettau, M., and S. Ludvigson. 2001. Resurrecting the (C) CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying. *Journal of Political Economy* 109: 1238–1287.

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023

Lewellen, J., S. Nagel, and J. Shanken. 2010. A Skeptical Appraisal of Asset Pricing Tests. *Journal of Financial Economics* 96: 175–194.

Mathew, T., and K. Nordström. 1997. Wishart and Chi-Square Distributions Associated with Matrix Quadratic Forms. *Journal of Multivariate Analysis* 61: 129–143.

Moreira, M. J. 2003. A Conditional Likelihood Ratio Test for Structural Models. *Econometrica* 71: 1027–1048.

Muirhead, R. J. 2009. *Aspects of Multivariate Statistical Theory*, Vol. 197. Hoboken, NJ: John Wiley & Sons.

Newey, W. K., and F. Windmeijer. 2009. GMM with Many Weak Moment Conditions. *Econometrica* 77: 687–719.

Parker, J. A., and C. Julliard. 2005. Consumption Risk and the Cross Section of Expected Returns. *Journal of Political Economy* 113: 185–222.

Savov, A. 2011. Asset Pricing with Garbage. *The Journal of Finance* 66: 177–201.

Shanken, J. 1992. On the Estimation of Beta-Pricing Models. *Review of Financial Studies* 5: 1–33.

Downloaded from https://academic.oup.com/jfec/article/21/2/263/6575317 by guest on 18 June 2023