Journal of Econometrics 248 (2025) 105728

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Identification robust inference for the risk premium in term structure models☆

Frank Kleibergen [a,*], Lingwei Kong [b]

[a] Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands
[b] Faculty of Economics and Business, University of Groningen, Nettelbosje 2, 9747 AE Groningen, The Netherlands

ARTICLE INFO

JEL classification:
C12
C58
G10
G12

Keywords:
Asset pricing
Risk premia
Robust inference
Weak identification

ABSTRACT

We propose identification robust statistics for testing hypotheses on the risk premia in dynamic affine term structure models. We do so using the moment equation specification proposed in Adrian et al. (2013). Statistical inference based on their three-stage estimator requires knowledge of the risk factors' quality and can be misleading when the β's are weak, which results when sampling errors are of comparable order of magnitude as the risk factor loadings. We extend the subset (factor) Anderson–Rubin test from Guggenberger et al. (2012) to models with multiple dynamic factors and time-varying risk prices. It provides a computationally tractable manner to conduct identification robust tests on a few risk premia when a larger number is present. We use it to analyze potential identification issues arising in the data from Adrian et al. (2013) for which we show that some factors, though potentially weak, may drive the time variation of risk prices, and weak identification issues are more prominent in multi-factor models.

1. Introduction

A variety of Dynamic affine term structure models (DATSMs) have been developed since the foundational work by Vasicek (1977) and Cox et al. (1985). They help to understand the movements of bond yields and to analyze financial markets. DATSMs are empirically appealing for their smooth tractability and simple characterization of how risks get priced. There are many studies employing this framework. To list a few: Cochrane and Piazzesi (2005) apply affine term structure models to study time variation in expected excess bond returns using a single explanatory factor; Wu and Xia (2016) use affine models to summarize the macroeconomic effects of unconventional monetary policy; Ang and Piazzesi (2003) investigate how macro economic variables affect bond prices and the dynamics of the yield curve, Buraschi and Jiltsov (2005) study the properties of the nominal and real risk premia of the term structure of interest rates and Goliński and Zaffaroni (2016) incorporate long memory state variables into the term structure model. We adopt the DATSMs setup developed by Adrian et al. (2013) which nests a general class of linear asset pricing models and can be regarded as a linear asset pricing model with time-varying risk premia and dynamic factors.

Many recent studies have developed approaches to estimate DATSMs. Most of them involve a time-consuming numerical optimization procedure which results from their high degree of non-linearity. Inference concerning the coefficients suffers similar challenges. An undesirable feature, as pointed out by, e.g., Hamilton and Wu (2012), is that identification can be problematic.

☆ We thank the editor, associate editor, two anonymous referees and participants of the conference at Cireq Montreal in honor of Eric Renault for many helpful comments and suggestions which considerably improved the manuscript. The research of Frank Kleibergen has been funded partly by the NWO (Dutch Research Council) grant 401.21.EB.002: "Double robust inference for structural economic models (DRISEM)".

* Corresponding author.
E-mail addresses: f.r.kleibergen@uva.nl (F. Kleibergen), l.kong@rug.nl (L. Kong).

https://doi.org/10.1016/j.jeconom.2024.105728
Received 31 October 2022; Received in revised form 22 March 2024; Accepted 24 March 2024
Available online 23 April 2024

Lack of identification, e.g., due to unspanned factors (Adrian et al., 2013), results in a relatively flat surface of the likelihood because of which traditional inference becomes unreliable, see e.g., Kan and Zhang (1999), Gospodinov et al. (2017), Kleibergen (2009), Hamilton and Wu (2012), Dovonon and Renault (2013), Beaulieu et al. (2016), Khalaf and Schaller (2016), Cattaneo et al. (2022).

Unspanned factors refers to factors that only affect the dynamics of bond prices under the historical measure but not the risk neutral one, see, e.g., Joslin et al. (2014). Empirical studies have shown their presence, see e.g., Ludvigson and Ng (2009), Adrian et al. (2013). They can lead to identification challenges since varying the parameters of the risk neutral pricing measure associated with these factors does not strongly influence bond prices. Adrian et al. (2013) therefore allow for the presence of unspanned factors using the prior knowledge of knowing which factors are unspanned. Their proposed estimation procedure then differs for cases with and without unspanned factors.

Recently Crump and Gospodinov (2022) have also argued that the persistence in bond yields raises inference issues when extracting principal components from them. Principal components extracted from excess returns and difference returns might therefore be preferable since they are much less persistent. Despite that these factors, including when constructed from yields, are by construction correlated with excess returns on bonds, these correlations might be of the same order of magnitude as the estimation error present in the DATSM which would make their risk premia weakly identified.

Because of these identification issues, traditional inference methods based on t-tests and Wald statistics can become unreliable for conducting inference on the risk prices in DATSMs, see e.g., Stock and Wright (2000), Kleibergen (2005), Antoine and Renault (2020), Andrews and Cheng (2012), Antoine and Lavergne (2022). We therefore propose easy-to-implement identification robust test procedures which are valid even when the model is not identified due to unspanned factors. The test procedures we provide can be used to study the time-varying risk-premia for linear asset pricing models. Our proposed inference procedures use the framework presented in Adrian et al. (2013), where the risk of bond prices is modeled as a linear functional in observed factors. The risk of bond prices can then be decomposed into two parts: a time-constant and a time-varying part. We propose statistics for testing hypotheses specified on all parameters of the time-varying components and on just subsets of them.

The paper is organized as follows: Section 2 introduces the DATSM. Section 3 states the three-step estimation procedure from Adrian et al. (2013) and shows that it is sensitive to identification issues. Section 3 also shows the empirical relevance of these identification issues using the data from Adrian et al. (2013). Section A.1 of the Online Supplementary Appendix shows that these identification issues are similarly present when risk factors are extracted as principal components that result from other representations than yields, such as, for example, excess returns and difference returns. Section 4 introduces joint identification robust tests for time-varying risk premia. It conducts a small simulation experiment and applies them in empirical applications under a single-factor setting. A more detailed simulation experiment, including restrictions imposed by DATSM, is discussed in Section A.2 of the Online Supplementary Appendix. Section 5 introduces sub-vector identification robust tests for testing hypotheses on subsets of time-varying risk premia. It applies them in various multi-factor settings using data from Adrian et al. (2013). The concluding section wraps up the discussion.

We use the following notation throughout the paper: "$\otimes$" and "vec$(\cdot)$" represent respectively the Kronecker product and vectorization operator; "vecinv" refers to the inverse operation of "vec". The "vech" (vector half) operation transforms a symmetric matrix into a column vector by stacking only the lower triangular part of the matrix (including the diagonal) and "vechinv" is the inverse operation of "vech", which takes a vector that represents the lower triangular part of a symmetric matrix and reconstructs the original symmetric matrix from it. $\Sigma^{\frac{1}{2}}$ is the lower triangular Cholesky decomposition of the positive definite symmetric matrix Σ such that $\Sigma = \Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}'}$; for a $N \times K$ dimensional full rank matrix A : $P_A = A(A'A)^{-1}A'$ and $M_A = I - P_A$. The proofs of the main theorems, i.e. Theorems 2–4, are in the Appendix A while the proof of the more secondary Theorem 1 is in the Online Supplementary Appendix.

2. Dynamic affine term structure models

We briefly discuss the popular class of DATSMs with observed factors. Instead of working directly with the implied yields on an n-period bond as usually done in the term structure literature, we use the excess holding return of an n-period bond as in Adrian et al. (2013).

We first illustrate the model set-up following Adrian et al. (2013). For $P_{t,n}$, the price at time t of a zero-coupon bond maturing at time $t+n$, the pricing kernel, M_{t+1}, is such that

$$P_{t,n} = E_t(M_{t+1}P_{t+1,n-1}), \tag{1}$$

where $E_t(\cdot)$ represents the conditional expectation using all information up to time t. For r_t, the one-period short rate and λ_t the market price of risk, the pricing kernel is assumed exponential affine in innovation factors $v_t \sim_{i.i.d} N(0, \Sigma_v)$:

$$M_{t+1} = \exp\left(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\Sigma_v^{-\frac{1}{2}}v_{t+1}\right), \tag{2}$$

where the market price of risk λ_t is an affine function of the observed factors X_t :

$$\lambda_t = \Sigma_v^{-\frac{1}{2}}\left(\lambda_0 + \Lambda_1 X_t\right), \tag{3}$$

with λ_0 and Λ_1 resp. a K-dimensional vector and a $K \times K$ dimensional matrix, and the K-dimensional vector of state variables X_t results from a vector autoregressive model of order 1 (VAR(1)):

$$X_{t+1} = \mu + \Phi_1 X_t + v_{t+1}. \tag{4}$$

For the one-period (log) excess holding return of a n-period bond at time $t+1$:

$$r_{t+1,n} = \ln\left(P_{t+1,n}\right) - \ln\left(P_{t,n+1}\right) - r_t, \tag{5}$$

with $r_t = \ln P_{t,1}$, the structure of the pricing kernel implies that:

$$E_t\left[\exp\left(-r_{t+1,n} - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\Sigma_v^{-\frac{1}{2}} v_{t+1}\right)\right] = 1. \tag{6}$$

Assuming that $(r_{t+1,n},\ v_{t+1})$ are jointly normal, Adrian et al. (2013) show:

$$E_t\left(r_{t+1,n}\right) = \beta^{(n)\prime}\left(\lambda_0 + \Lambda_1 X_t\right) - \frac{1}{2} var(r_{t+1,n}), \tag{7}$$

with $\beta^{(n)} = \Sigma_v^{-1} cov(v_{t+1}, r_{t+1,n})$. When decomposing, $R_{t+1,n}$ into a component correlated with v_{t+1} and an uncorrelated component/prediction error $e_{t+1,n}$:

$$r_{t+1,n} - E_t\left(r_{t+1,n}\right) = \beta^{(n)\prime} v_{t+1} + e_{t+1,n}; \tag{8}$$

we then next have:

$$r_{t+1,n} = \beta^{(n)\prime}\left(\lambda_0 + \Lambda_1 X_t\right) + g^{(n)}(\beta, \Sigma_v, \Sigma_e) + \beta^{(n)\prime} v_{t+1} + e_{t+1,n}, \tag{9}$$

where $g^{(n)}(\beta, \Sigma_v, \Sigma_e) = -\frac{1}{2} var(r_{t+1,n})$, thus, for example, in case Σ_e =var$(e_{t+1,n}) = \sigma_e^2$, $g^{(n)}(\beta, \Sigma_v, \Sigma_e) = -\frac{1}{2}\left(\beta^{(n)\prime}\Sigma_v\beta^{(n)} + \sigma_e^2\right)$. Additional restrictions are often imposed on the parameters because of the cross-sectional term structure, but these restrictions are not used in Adrian et al. (2013)'s approach. Assumption 1 next summarizes the model setting.

Assumption 1 (a)**.** Consider a $K \times 1$ vector of factors X_t, $t = 0, \ldots, T$ that results from a stationary VAR(1):

$$X_{t+1} = \mu + \Phi_1 X_t + v_{t+1}, \tag{10}$$

where v_t are the innovation shocks (or innovation factors). The log excess holding return $r_{t+1,n-1}$ satisfies:

$$r_{t+1,n-1} = \beta^{(n-1)'}\left(\lambda_0 + \Lambda_1 X_t\right) + g^{(n-1)}\left(\beta, \Sigma_v, \Sigma_e\right) + \beta^{(n-1)'} v_{t+1} + e_{t+1,n-1} \tag{11}$$

with $g^{(n-1)}(\cdot)$ a parametric function. Furthermore:

$$(v_{t+1}', e_{t+1}')' \,|\, \left\{X_s\right\}_{s=0}^{t} \sim i.i.d.N\left(0, \mathrm{diag}(\Sigma_v, \Sigma_e)\right). \tag{12}$$

(b) For $\Sigma_e = \sigma_e^2$, $g^{(n-1)}(\beta, \Sigma_v, \Sigma_e) = \frac{1}{2}(\beta^{(n-1)'}\Sigma_v\beta^{(n-1)} + \sigma_e^2)$.

3. Regression estimator and Wald based inference

To estimate the price of risk, Adrian et al. (2013) propose a three-step procedure akin to the two-pass procedure from Fama and MacBeth (1973):

1. Estimate the VAR(1):

$$X_{t+1} = \mu + \Phi_1 X_t + v_{t+1}, \tag{13}$$

by least squares to obtain $\hat{\mu}$, $\hat{\Phi}$, $\hat{v}_t = X_t - \hat{\mu} - \hat{\Phi}_1 X_{t-1}$, $t = 1, \ldots, T$ and $\hat{\Sigma}_v = \frac{1}{T}\sum_{t=1}^{T} \hat{v}_t\hat{v}_t'$.

2. Estimate:

$$r_{t+1,n} = a^{(n)} + d^{(n)\prime} X_t + \beta^{(n)\prime}\hat{v}_{t+1} + e_{t+1,n}, \tag{14}$$

by least squares to obtain $\hat{a}^{(n)}$, $\hat{d}^{(n)}$ and $\hat{\beta}^{(n)}$, $n = 1, \ldots, N$.

3. Construct $\hat{a} = (\hat{a}^{(1)} \ldots \hat{a}^{(N)})'$, $\hat{\beta} = (\hat{\beta}^{(1)} \ldots \hat{\beta}^{(N)})'$, $\hat{d} = (\hat{d}^{(1)} \ldots d^{(n)})'$, $\hat{g} = (\hat{g}^{(1)} \ldots \hat{g}^{(n)})$ for $\hat{g}^{(n)} = g^{(n)}(\hat{\beta}, \hat{\Sigma}_v, \hat{\Sigma}_e)$, $n = 1, \ldots, N$, and estimate λ_0 and Λ_1 using:

$$\begin{aligned} \hat{\lambda}_0 &= (\hat{\beta}'\hat{\beta})'^{-1}\hat{\beta}'(\hat{a} + \hat{g} + \hat{\beta}\hat{\mu}) \\ \hat{\Lambda}_1 &= (\hat{\beta}'\hat{\beta})'^{-1}\hat{\beta}'(\hat{d} + \hat{\beta}\hat{\Phi}). \end{aligned} \tag{15}$$

The three-step procedure regresses transformed returns on estimated β's. Because of the inversion of $\hat{\beta}'\hat{\beta}$ in the estimation of the risk premia, their reliability crucially depends on the quality of the β's. If the β's are relative small or the β matrix is close to a reduced rank value, the estimation error resulting from the error terms can be of the same order of magnitude as β and make the risk premia estimates unreliable. This is the setting of so-called weak factors, see e.g. Kleibergen (2009) and Anatolyev and Mikusheva (2022). Tests for a reduced rank value of β are then typically just barely significant if at all.

Table 1
Least squares estimates of the β's associated with the excess returns for bonds with 11 different maturities of $6, 12, 18 \cdots, 60$ and 84, 120 months over the sample period 1987:01-2011:12. The factors are the first five principal components generated using the cross-section of bond yields for maturities $3, \cdots, 120$ months (data is from Adrian et al. (2013)). The t-statistics of the β estimates are in round brackets while p-values of F-tests of $H_0 : \beta_j = 0$ are in square brackets below each column. The Kleibergen–Paap rank statistic (Kleibergen and Paap, 2006) testing $H_0 :$ rank$(\beta) = 4$, equals 1.6561 [0.9764], and to avoid near-singularity of the covariance matrix, it uses a diagonal covariance matrix for e_t.

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
(1)	−0.0094	0.0031	−0.0008	0.0002	0.0000
	(−3.6027)	(2.5058)	(−1.4886)	(0.4344)	(0.0347)
(2)	−0.0213	0.0057	−0.0007	−0.0003	0.0002
	(−8.2107)	(4.5416)	(−1.2228)	(−0.7341)	(0.5816)
(3)	−0.0446	0.0070	0.0010	−0.0005	−0.0001
	(−17.1745)	(5.6040)	(1.8889)	(−1.3535)	(−0.4551)
(4)	−0.0656	0.0048	0.0024	0.0000	−0.0003
	(−25.2648)	(3.8500)	(4.5279)	(0.0386)	(−0.9643)
(5)	−0.0843	0.0003	0.0028	0.0007	−0.0001
	(−32.4670)	(0.2023)	(5.2857)	(1.7192)	(−0.1732)
(6)	−0.1011	−0.0059	0.0022	0.0010	0.0003
	(−38.9474)	(−4.6910)	(4.1584)	(2.6824)	(1.0713)
(7)	−0.1164	−0.0130	0.0008	0.0010	0.0006
	(−44.8511)	(−10.3464)	(1.5602)	(2.5469)	(1.9102)
(8)	−0.1305	−0.0206	−0.0011	0.0005	0.0005
	(−50.2742)	(−16.4072)	(−2.0105)	(1.2971)	(1.8297)
(9)	−0.1435	−0.0284	−0.0033	−0.0003	0.0002
	(−55.2826)	(−22.6118)	(−6.1006)	(−0.8978)	(0.6354)
(10)	−0.1556	−0.0361	−0.0056	−0.0015	−0.0005
	(−59.9307)	(−28.7667)	(−10.3483)	(−3.7935)	(−1.6457)
(11)	−0.1669	−0.0436	−0.0078	−0.0028	−0.0014
	(−64.2706)	(−34.7284)	(−14.4913)	(−7.1319)	(−4.8558)
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0001]

When modeling unspanned factors that are present in the historical measure but not the risk neutral one, Adrian et al. (2013) show that the entries in the β's corresponding to unspanned factors are zero, hence, they influence the dynamics of bond prices through the state variables present under the historical measure but have zero factor loadings and are therefore not present under the risk neutral pricing measure. Including such unspanned factors or weak factors that are just minorly correlated with returns can then further aggravate identification issues, see e.g., Kleibergen (2009), Beaulieu et al. (2013), Kleibergen and Zhan (2015).

We illustrate the empirical relevance of weak factors and the quality of the β's using data from Adrian et al. (2013), i.e., the zero coupon yield data constructed by Gürkaynak et al. (2007). Table 1 therefore shows the factor loading estimates and p-values of significance tests for a zero value for each of the columns of the β matrix. The factors are the first five principal component (PCA) factors from Adrian et al. (2013). Table 1 shows that many elements of β are small and not statistically different from zero at the 5% significance level. While the F-tests on the columns of β are significant, the rank test of the β -matrix indicates potential identification issues because it does not reject a lower rank value of the β matrix at the 5% significance level.

To further illustrate the quality of the β's, Fig. 1 shows the (log) singular values of $(\hat{d} \ \vdots \ \hat{\beta})$, which results from the second of the three-step estimation procedure (14), and the percentage of the variation that is explained by the principal components. When the model is correctly specified, $(\hat{d} \ \vdots \ \hat{\beta})$ provides an estimate of $\beta(\Lambda_1 \ \vdots \ I_K)$ so its smallest K singular values should be close to zero. Panel (a) of Fig. 1 uses the largest five principal components as risk factors and shows that the smallest six singular values are close to zero and not just the smallest five. It indicates a weak/unspanned factor problem and explains why the rank test does not reject rank$(\beta) = 4$ in Table 1 at the 5% significance level.

It has previously been noted that, see e.g., Kleibergen and Zhan (2020), Kleibergen et al. (2023), weak identification issues are often present when macro factors are used. Following Ang and Piazzesi (2003), we construct a macro factor, i.e. the real activity measure, which equals the first principal component resulting from four variables that capture real US macro activity: the "Help Wanted Advertising in Newspapers (HELP)"[1] index, unemployment (UE), the growth rate of employment (EMPLOY), and the growth rate of industrial production (IP). Table 2 shows that the macro factor is much less correlated with returns and thus is more likely to result in an identification issue. This is further reflected in Panel (b) of Fig. 1. It shows that while there are six factors, the smallest seven singular values are close to zero. Table 2 also shows a tiny value for the rank test which provides another indication of a weak/unspanned factor problem.

Crump and Gospodinov (2022) raise caution with respect to the persistence present in bond yields, which carries over to principal components extracted from them. We, therefore, repeated the analyzes in Tables 1 and 2 when using principal components extracted from excess returns and difference returns, as Crump and Gospodinov (2022) advocate. Figures A.1-A.5 in Section A.1 of the Online Supplementary Appendix contain the results when using principal components extracted from excess returns and difference returns

[1] We use the HELP-Wanted index from Barnichon (2010) to match the time periods of the excess returns.

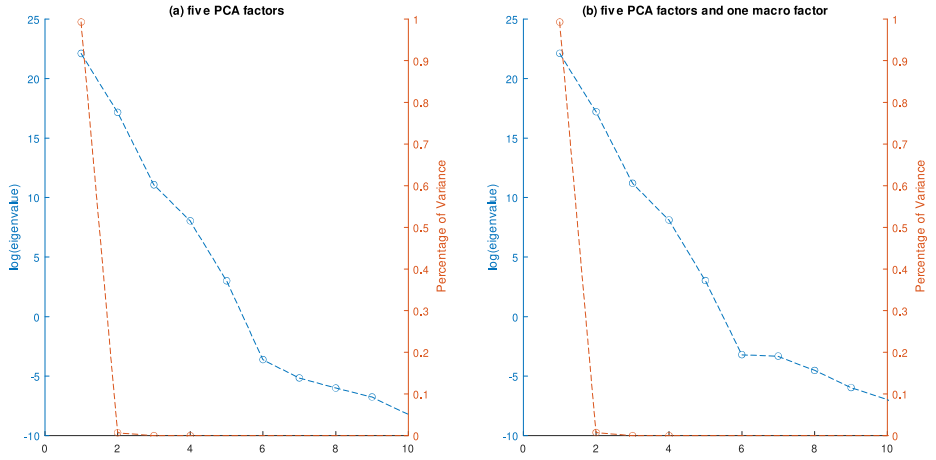

Fig. 1. (log) singular values (blue) of $(\hat{d} \vdots \hat{\beta})$ from step 2 in the three-step estimation procedure and the percentage of the variance explained (red) by the principal components. R_t uses the demeaned excess returns of the same bonds as in Table 1, $\bar{X}_t$ uses the first five principal components as factors in Panel (a) and uses the same five PCA factors with one additional macro factor in Panel (b). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Least squares estimates of the β's associated with the excess returns for bonds with 11 different maturities of $6, 12, 18 \cdots, 60$ and 84, 120 months over the sample period 1987:01-2011:12. The factors are the first five principal components generated using the cross-section of bond yields for maturities $3, \cdots, 120$ months (data is from Adrian et al. (2013)) and a macro factor (real activity) that is constructed following Ang and Piazzesi (2003). The t-statistics of the β estimates are in round brackets while p -values of F-tests of $H_0 : \beta_j = 0$ are in square brackets below each column. The Kleibergen–Paap rank statistic (Kleibergen and Paap, 2006) testing $H_0 : \text{rank}(\beta) = 5$, equals 0.0025 [1.000], and to avoid near-singularity of the covariance matrix, it uses a diagonal covariance matrix for e_t.

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_{macro}$
(1)	−0.0094	0.0032	−0.0008	0.0002	0.0000	0.0001
	(−3.4930)	(2.4109)	(−1.4749)	(0.4253)	(0.0309)	(0.0516)
(2)	−0.0213	0.0057	−0.0007	−0.0003	0.0002	0.0000
	(−7.9353)	(4.3501)	(−1.2115)	(−0.7259)	(0.5717)	(0.0254)
(3)	−0.0446	0.0070	0.0010	−0.0005	−0.0001	−0.0000
	(−16.5873)	(5.3631)	(1.8715)	(−1.3372)	(−0.4497)	(−0.0037)
(4)	−0.0656	0.0049	0.0024	0.0000	−0.0003	0.0001
	(−24.4149)	(3.6980)	(4.4871)	(0.0346)	(−0.9529)	(0.0655)
(5)	−0.0843	0.0003	0.0028	0.0007	−0.0001	0.0001
	(−31.3724)	(0.2100)	(5.2383)	(1.6937)	(−0.1732)	(0.0801)
(6)	−0.1011	−0.0059	0.0022	0.0010	0.0003	0.0000
	(−37.6206)	(−4.4799)	(4.1208)	(2.6462)	(1.0539)	(0.0333)
(7)	−0.1164	−0.0130	0.0008	0.0010	0.0006	−0.0000
	(−43.3100)	(−9.8999)	(1.5456)	(2.5142)	(1.8813)	(−0.0252)
(8)	−0.1305	−0.0206	−0.0011	0.0005	0.0005	−0.0001
	(−48.5424)	(−15.7013)	(−1.9930)	(1.2811)	(1.8021)	(−0.0512)
(9)	−0.1435	−0.0284	−0.0033	−0.0003	0.0002	−0.0000
	(−53.3870)	(−21.6288)	(−6.0457)	(−0.8872)	(0.6243)	(−0.0204)
(10)	−0.1557	−0.0361	−0.0056	−0.0015	−0.0005	0.0001
	(−57.8990)	(−27.4937)	(−10.2541)	(−3.7506)	(−1.6271)	(0.0745)
(11)	−0.1670	−0.0435	−0.0078	−0.0028	−0.0014	0.0002
	(−62.1305)	(−33.1560)	(−14.3589)	(−7.0557)	(−4.7981)	(0.2297)
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0002]	[1.0000]

as risk factors compared to principal components extracted from bond yields. These results similarly show that weak identification remains an issue when extracting principal components from other representations than yields. Because of the similarity of these results, we, for reasons of brevity, only report them in the Online Supplementary Appendix.

The potential identification issues arising from the quality of the β's revealed in Fig. 1 and Tables 1–2 affect the estimators' validity and the reliability of traditional inference procedures. Adrian et al. (2013) suggest that knowing the unspanned factors helps identify the zero rows in β and mitigate the identification problem. We remain agnostic about this and propose test procedures related to Λ_1 that are identification robust without requiring prior knowledge of the unspanned factors.

Before introducing our novel test procedures, we first show that traditional Wald-type inference becomes unreliable when unspanned factors are weak. The notion of weak unspanned factors is formalized in Assumption 2. It assumes that (part of) the β matrix drifts to zero at rate $\sqrt{T}$. Assumption 2 should not be taken too literally. It is a device which leads to the smallish rank

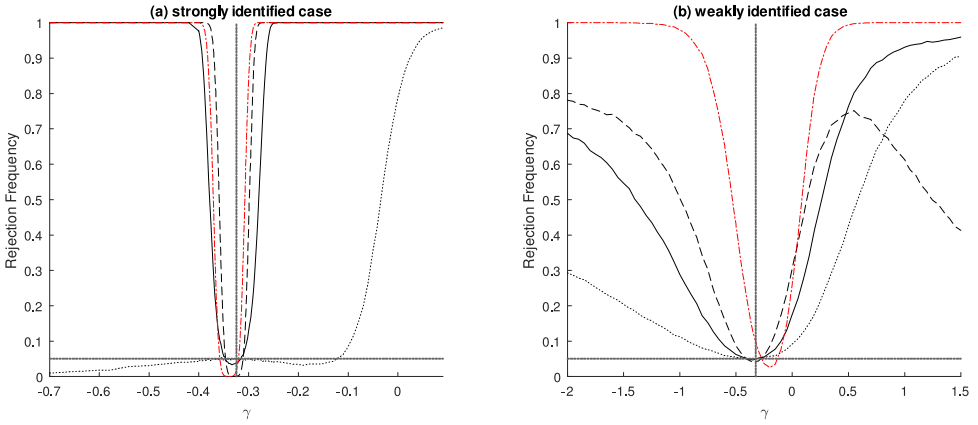

Fig. 2. Left panel (a) is the strong identified case, while the right panel (b) is for weak identification. Power curves (rejection frequencies) of Wald (dash-dotted, red), FAR (solid), KLM (dashed), and JKLM (dotted) testing Λ_1 (scalar) with its values on the horizontal axis; horizontal dashed line at 5% and vertical one at the true calibrated value of Λ_1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tests on the β's reported in Tables 1 and 2. Under the usual strong factor assumption, these rank tests should be proportional to the sample size which they are clearly not. Because the outcome of the rank tests is at odds with their implied value under the strong factor assumption, it is unclear if traditional inference procedures remain unaffected by small values of the rank tests. We therefore establish the limiting distribution of the β and risk premia estimator resulting from the three-step procedure under the weak unspanned factor assumption that accommodates small values of the rank tests.

Assumption 2. Potential existence of (nearly) unspanned factors: $\beta' = (B, C)$, B represents the spanned factors and is of full rank $K_B \leq K$ while $C = O(1/\sqrt{T})$ reflects the (nearly) unspanned factors. If $K_B = K$, there are no (nearly) unspanned factors.

Assumption 2 captures weak factors by having certain rows of β to be inversely proportional to the square root of the sample size so they are of the same order of magnitude as the estimation error. These rows of small values of the β's correspond to the unspanned factors. Adrian et al. (2013) assume that their location is known, and adapt their three-step procedure accordingly. It therefore excludes the unspanned factors in the second step and solely incorporates spanned factors. The third step would otherwise encounter identification issues in line with the classical multicollinearity problem because a zero value of the β's leads to an unidentified value of λ. When we instead just have small β's, which are comparable in magnitude to the estimation error as stated in Assumption 2, we similarly encounter such an identification problem, see e.g., Kleibergen (2009), Antoine and Renault (2009), Antoine and Renault (2012) Kleibergen and Zhan (2020), Kleibergen et al. (2023).

Theorem 1. *Under Assumption 1.(a), denote $\Lambda = [\lambda_0, \Lambda_1]$:*

(a) If β is of full rank:

$$\sqrt{T}\begin{bmatrix} vec\left(\hat{\beta}' - \beta'\right) \\ vec\left(\hat{\Lambda} - \Lambda\right) \end{bmatrix} \rightarrow_d N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathcal{V}_\beta & C'_{\Lambda,\beta} \\ C_{\Lambda,\beta} & \mathcal{V}_\Lambda \end{bmatrix}\right), \tag{16}$$

where $\mathcal{V}_\beta, C_{\Lambda,\beta}, \mathcal{V}_\Lambda$ are specified in the proof in the Online Supplementary Appendix.

(b) If Assumption 2 holds with $K_B < K$, then $\hat{\Lambda} \rightarrow_d \Lambda + \epsilon$, where ϵ follows a non-standard distribution, so $\hat{\Lambda}$ no longer converges to the true value Λ at rate $\sqrt{T}$.

Proof. See the Online Supplementary Appendix. □

Theorem 1 states some well-known results from the weak identification literature. It shows that the risk premia estimator $\hat{\Lambda}$ becomes inconsistent in the presence of weak factors because it converges to a random variable with a non-standard distribution. This results since varying the value of the risk premia associated with the weak factors does not change the excess returns much. The asymptotic distribution of the conventional Wald statistic for testing the null hypothesis $H_0 : \Lambda_1 = \Lambda_1^0$ then no longer converges to a χ^2-distribution, and the same holds for subset tests based on this estimator. Therefore, the conventional test statistics can be misleading in the presence of unspanned/weak factors, as depicted in Fig. 2 in Section 4.1 showing a case where the Wald test is size distorted under weak identification. Theorem 1 does not rely on the normality assumption in Assumption 1 whose violations, however, make the model misspecified.

4. Identification robust tests of time-varying risk premia

We focus on inference concerning the time-varying component of risk prices Λ_1. We, therefore, demean the one-period (log) excess holding returns by subtracting its time-series average:

$$\bar{r}_{t+1,n} = \beta^{(n)\prime}\left(\Lambda_1 \bar{X}_t\right) + \beta^{(n)\prime}\bar{v}_{t+1} + \bar{e}_{t+1,n}, \tag{17}$$

with $\bar{z}_{t+1,n} = z_{t+1,n} - \bar{z}$ and $\bar{z} = \frac{1}{T}\sum_{t=1}^{T} z_t$ for $z = r$, X, v and e resp. When stacking the equations for N different maturities:

$$R_{t+1} = \beta\left(\Lambda_1 \bar{X}_t\right) + \beta\bar{v}_{t+1} + e_{t+1}, \tag{18}$$

where $R_{t+1} = (\bar{r}_{t+1,1} \dots \bar{r}_{t+1,N})'$, $\beta = (\beta^{(1)} \dots \beta^{(n)})'$, $e_t = (\bar{e}_{t,1} \dots \bar{e}_{t,N})'$, the pricing equation closely resembles the beta-pricing model for the return on (portfolios of) assets:

$$\mathrm{r}_{t+1} = \beta\lambda + \beta\mathrm{F}_{t+1} + u_{t+1}, \tag{19}$$

with r_t an N-dimensional vector with the returns on N assets, β an $N\times K$-dimensional factor loading matrix and F_t a K -dimensional vector of risk factors. A further important similarity that both models imply is the reduced rank structure that becomes obvious using a slight respecification:

$$R_{t+1} = \beta\left(\Lambda_1 \ \vdots \ I_K\right)\begin{pmatrix} \bar{X}_t \\ \bar{v}_{t+1} \end{pmatrix} + e_t \tag{20}$$

and

$$\mathrm{r}_{t+1} = \beta\left(\lambda \ \vdots \ I_K\right)\begin{pmatrix} 1 \\ F_{t+1} \end{pmatrix} + u_t, \tag{21}$$

where the $N \times 2K$ and $N \times (K+1)$ dimensional matrices $\beta(\Lambda_1 \ \vdots \ I_K)$ and $\beta(\lambda \ \vdots \ I_K)$ are each at most of rank K, so except for the largest K singular values, all, K and 1 resp., smallest singular values of these matrices are zero. Further for Λ_1 and λ to be well defined, β should be of full rank. When β is near a reduced rank value, or in other words, if some factors are weak/unspanned, we encounter an identification issue which is also reflected by more than just K or one resp. singular values of the above matrices being equal or close to zero.

We provide identification robust tests based on sample moment equations in a generalized methods of moments (GMM) setting. The sample moments result from the model in (9). We subtract from it the return specific time-series average to obtain (17). Our sample moment vector next imputes estimates of $\bar{v}_{t+1}$, $\hat{v}_{t+1}$, and β, $\hat{\beta}$, which result from steps 1 and 2 of the three-step estimation procedure, in (17) and assumes that the resulting estimation error is uncorrelated with $\bar{X}_t$. For $\hat{Q}_{XX} = \frac{1}{T}\sum_{t=1}^{T}\bar{X}_{t-1}\bar{X}_{t-1}'$, our sample moment vector for Λ_1 then becomes:

$$\begin{aligned} f_T(\Lambda_1, X) = & \ \frac{1}{T}\sum_{t=1}^{T}\left(\bar{X}_{t-1} \otimes \left[(R_t - \hat{\beta}\hat{V}_t) - \hat{\beta}\Lambda_1\bar{X}_{t-1}\right]\right) \\ = & \ \frac{1}{T}\sum_{t=1}^{T}(\bar{X}_{t-1} \otimes (R_t - \hat{\beta}\hat{V}_t)) - \left(\hat{Q}_{XX} \otimes \hat{\beta}\right)\mathrm{vec}(\Lambda_1), \end{aligned} \tag{22}$$

and its derivative with respect to $\mathrm{vec}(\Lambda_1)$ is

$$q_T(X) = -\left(\hat{Q}_{XX} \otimes \hat{\beta}\right). \tag{23}$$

We next make an assumption regarding the large sample behavior of the sample moment vector and its derivative.

Assumption 3. Under $H_0 : \Lambda_1 = \Lambda_1^0$,

$$\sqrt{T}\begin{pmatrix} f_T(\Lambda_1^0, X) \\ \mathrm{vec}(q_T(X) - J) \end{pmatrix} \underset{d}{\to} \begin{pmatrix} \psi_f \\ \psi_q \end{pmatrix}, \tag{24}$$

where the Jacobian $J = -(Q_{XX} \otimes \beta)$, $\hat{Q}_{XX} \underset{p}{\to} Q_{XX}$, and ψ_f and ψ_q are NK- and NK^3-dimensional random vectors:

$$\begin{pmatrix} \psi_f \\ \psi_q \end{pmatrix} \sim N\left(0, V(\Lambda_1^0)\right), \tag{25}$$

with

$$V(\Lambda_1^0) = \begin{pmatrix} V_{ff}(\Lambda_1^0) & V_{qf}(\Lambda_1^0)' \\ V_{qf}(\Lambda_1^0) & V_{qq}(\Lambda_1^0) \end{pmatrix}, \tag{26}$$

where $V_{ff}(\Lambda_1^0)$, $V_{qf}(\Lambda_1^0)$ and $V_{qq}(\Lambda_1^0)$ are $NK \times NK$, $NK^3 \times NK$ and $NK^3 \times NK^3$ dimensional matrices.

Assumption 3 is a high-level assumption which resembles Assumption 1 in Kleibergen (2005) and holds under mild conditions. Assumption 3 holds true irrespective of Assumption 2. Assumption 1 is sufficient for Assumption 3, but our proposed test statistics can be applied for more general cases than the model implied in Assumption 1. For our setting, we specifically have that:

$$\psi_f = \psi_{\bar{f}} + \Psi_q \mathrm{vec}(\Lambda_1^0), \tag{27}$$

for Ψ_q =vecinv(ψ_q) and

$$\sqrt{T}\left(\tfrac{1}{T}\textstyle\sum_{t=1}^{T}(\bar{X}_{t-1}\otimes(\bar{R}_t-\hat{\beta}\hat{V}_t))-\left(Q_{XX}\otimes\beta\right)\mathrm{vec}(\Lambda_1^0)\right)\underset{d}{\to}\psi_{\tilde{f}}. \tag{28}$$

We also have

$$\psi_q=\mathrm{vec}((Q_{XX}\otimes\Psi_\beta)+(\Psi_{XX}\otimes\beta)), \tag{29}$$

where

$$\begin{aligned}\sqrt{T}\mathrm{vech}(\tfrac{1}{T}\textstyle\sum_{t=1}^{T}\bar{X}_t\bar{X}_t'-Q_{XX})\underset{d}{\to}\psi_{XX},\quad & \Psi_{XX}=\ \mathrm{vechinv}(\psi_{XX}),\\ \sqrt{T}\mathrm{vec}(\hat{\beta}-\beta)\underset{d}{\to}\psi_\beta,\quad & \Psi_\beta=\mathrm{vecinv}(\psi_\beta),\end{aligned} \tag{30}$$

with vech(A) containing the unique elements of a symmetric matrix A. Since ψ_q has NK^3 elements while the number of unique elements in Ψ_β and Ψ_{XX} equals $NK+\frac{1}{2}K(K+1)$, the joint normal distribution of (ψ_f,ψ_q) is further allowed to be degenerate.

The identification robust statistics use an estimator of the (recentered) Jacobian whose limit behavior under $\mathrm{H}_0:\Lambda_1=\Lambda_1^0$ is independent of the limit behavior of the sample moment, see Kleibergen (2005):

$$\begin{aligned}\hat{D}_T(\Lambda_1,X)=\ & (\hat{D}_{1,T}(\Lambda_1,X)\dots\hat{D}_{1,T}(\Lambda_k,X))\\ \mathrm{vec}(\hat{D}_T(\Lambda_1,X))=\ & \mathrm{vec}(q_T(X))-\hat{V}_{qf}(\Lambda_1)\hat{V}_{ff}(\Lambda_1)^{-1}f_T(\Lambda_1,X)\\ \sqrt{T}\mathrm{vec}(\hat{D}_T(\Lambda_1^0,X)-J)\underset{d}{\to}\ & \psi_{q.f}\sim N(0,V_{qq.f}(\Lambda_1^0))\end{aligned} \tag{31}$$

with $V_{qq.f}(\Lambda_1)=V_{qq}-V_{qf}(\Lambda_1)V_{ff}(\Lambda_1)^{-1}V_{qf}(\Lambda_1)'$, $\hat{V}_{qf}(\Lambda_1)$ and $\hat{V}_{ff}(\Lambda_1)$ consistent estimators of $V_{qf}(\Lambda_1^0)$ and $V_{ff}(\Lambda_1^0)$, and $\psi_{q.f}$ independent of ψ_f.

We can next define the identification robust Factor Anderson–Rubin (FAR), (Kleibergen) Lagrange multiplier (KLM) and JKLM statistics for testing $\mathrm{H}_0:\Lambda_1=\Lambda_1^0$:

$$\begin{aligned}\mathrm{FAR}(\Lambda_1^0)=\ & T\times f_T(\Lambda_1^0,X)'\hat{V}_{ff}(\Lambda_1^0)^{-1}f_T(\Lambda_1^0,X)\underset{d}{\to}\chi^2(KN)\\ \mathrm{KLM}(\Lambda_1^0)=\ & T\times f_T(\Lambda_1^0,X)'\hat{V}_{ff}(\Lambda_1^0)^{-\frac{1}{2}}P_{\hat{V}_{ff}(\Lambda_1^0)^{-\frac{1}{2}}\hat{D}_T(\Lambda_1^0,X)}\hat{V}_{ff}(\Lambda_1^0)^{-\frac{1}{2}}f_T(\Lambda_1^0,X)\underset{d}{\to}\chi^2(K^2)\\ \mathrm{JKLM}(\Lambda_1^0)=\ & T\times f_T(\Lambda_1^0,X)'\hat{V}_{ff}(\Lambda_1^0)^{-\frac{1}{2}}M_{\hat{V}_{ff}(\Lambda_1^0)^{-\frac{1}{2}}\hat{D}_T(\Lambda_1^0,X)}\hat{V}_{ff}(\Lambda_1^0)^{-\frac{1}{2}}f_T(\Lambda_1^0,X)\underset{d}{\to}\chi^2(K(N-K)).\end{aligned} \tag{32}$$

The limiting distributions are a direct result of Assumption 3 and do not depend on the rank of the Jacobian or β, so the limiting distributions hold regardless of Assumption 2.

4.1. Illustrative simulation and empirical results

We conduct a single-factor model simulation study to illustrate the performance of the proposed identification robust tests. For the data generating process (DGP), we consider

$$R_t=c+\beta\left(\Lambda_1X_{t-1}+v_t\right)+e_t,$$

where the parameters are calibrated to the data as in Table 1. We set the sample size to be $T=300$ and keep all parameters fixed except β's. We calibrate β's with the second PCA factor from Adrian et al. (2013) to mimic the strong identification case and the fifth one for the weak identification setting. Section A.2. of the Online Supplementary Appendix has another more extensive simulation study which mimics the data being generated from a DATSM and subsequently employs the identification robust tests. Since these results are qualitatively similar to those we report next and are considerably more elaborate, we, for expository purposes, only report them in the Online Supplementary Appendix.

Fig. 2 shows power curves of 5% significance tests using the conventional t-statistic and the identification robust test statistics for both strong and weak identification. It shows that FAR, KLM and JKLM tests are all size correct, while the Wald test is size distorted under weak identification and size correct but biased for strong identification. For weak identification, the KLM test has some power loss away from the hypothesized value because of which it is preferred to combine it in a conditional or unconditional manner with the J-test to improve power, see Moreira (2003) and Kleibergen (2005).

We use the identification robust tests to analyze the time-varying component of the risk premium. A detailed description of the involved excess returns and risk factors has been discussed previously for Tables 1–2. Fig. 3 shows the p-values of testing the risk premium associated with all six factors in a single-factor model using the different tests. A p-value larger than, say, 5%, implies that we cannot reject the null hypothesis at the 5% significance level.

Fig. 3 shows that for all factors, the FAR and KLM tests provide bounded 95% confidence sets since only for bounded regions the p-values are above the 95% line, even for those potentially weak factors such as the fifth PCA factor and the macro factor. It implies that in a single factor setting, all these risk premia are identified. For the high-order PCA factors, the robust tests, however, result in 95% confidence sets that differ from those resulting from the Wald test. Most striking is that a zero value for the risk premium is not rejected for strong factors such as the first and second PCA factors but rejected for potentially weak factors. For example, the null hypothesis that $\Lambda_1=0$ is rejected by the FAR and KLM test for both the fifth factor and the macro factor. This is partly in line with Adrian et al. (2013), which highlight the role of the higher-order principal components as the time variation may be largely

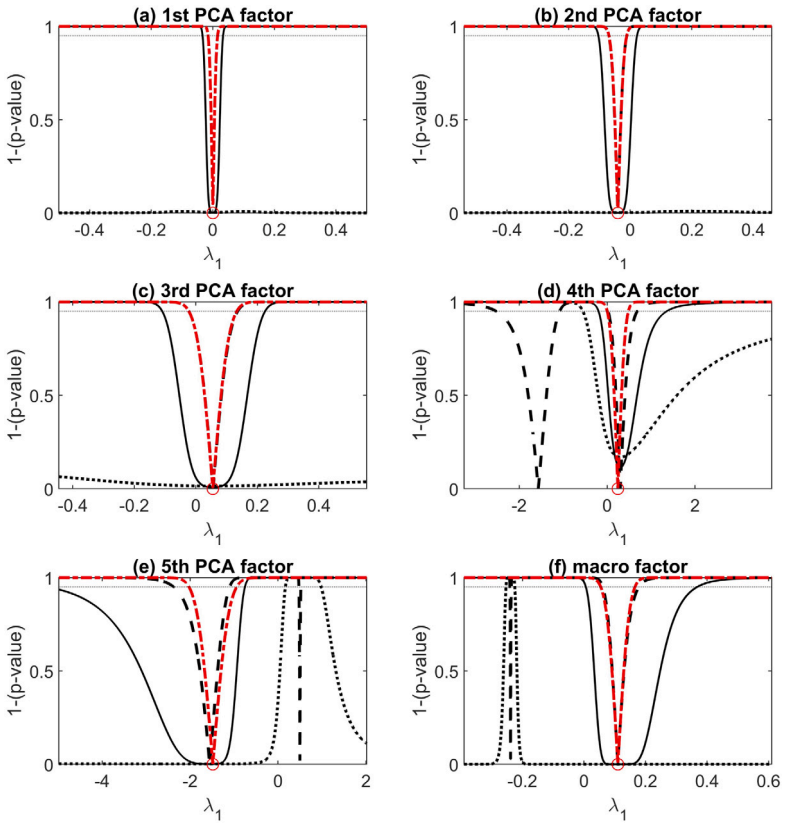

Fig. 3. one-minus-p-value curves of Wald (dash-dotted, red); FAR (solid); KLM (dashed); JKLM (dotted) for testing the Λ_1 (scalar) in a single factor model with values on the horizontal axis and dotted line at 95%. This figure uses the same data as in Table 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

driven by, e.g., the fifth principal component. Therefore, some factors may be weak but have some importance for interpreting the (time-varying) expected returns.

Fig. 3 also shows that for most cases the JKLM test leads to wide 95% confidence sets even for strong factors. While these confidence sets are wide, they remain bounded which can be shown when we use a wider grid. This results since, as shown by Fig. 2, the JKLM tests has considerably less power since it primarily tests misspecification.

Fig. 3 is in line with Tables 1 and 2 which show that the columns of the β-matrix are significantly different from zero.[2] This indicates that the risk premia resulting from a single factor model are well identified. Fig. 3 does, however, not indicate if the same applies when we consider a multi-factor setting.

5. Identification robust sub-vector testing

The identification robust tests introduced in the previous section are for testing hypotheses specified on all elements of Λ_1. We are often interested in testing hypotheses specified on just subsets of the parameters. When we analyze multi-factor models, testing whether or not a certain factor risk premium exhibits time variation would require testing a specific row of Λ_1, while testing whether a factor drives the time variation would require to test the corresponding column of Λ_1. Under our current settings, projection-based versions of the identification robust tests would allow us to test such hypotheses whilst preserving the size of the test, see Dufour and Taamouti (2005). These tests, however, lead to reduced power. We therefore extend the robust subset FAR test (sFAR) of Guggenberger et al. (2012), which concerns testing a hypothesis specified on a selection of the number of elements of a vector, i.e. a subvector, towards testing a hypothesis specified on just a row or column of the matrix Λ_1.

Without loss of generality, we consider testing the hypothesis that the risk premia associated with one specific factor, say the first, are all equal to λ_1^0:

$$\mathrm{H}_0 : \lambda_1 = \lambda_1^0, \tag{33}$$

[2] While Tables 1 and 2 report on the significance of the columns of β when using multiple factors, Table 4 further shows their significance in a single factor setting.

Table 3
KPS test (KPST) statistics for testing the null hypothesis that $H_0 : S = \Omega \otimes \Sigma$ for some $\Omega \in \mathbb{R}^{2K\times 2K}$ and $\Sigma \in \mathbb{R}^{N\times N}$ symmetric positive definite matrices. All four cases use excess returns on bonds with maturities 3, 12, 24, 60, 90, 120 months from Adrian et al. (2013). For the factors X_t, (1) uses the macro factor (real activity) and the level factor (first PCA factor), (2) uses the level factor (first PCA factor) and the slope factor (second PCA factor), (3) uses the macro factor (real activity) and the slope factor (second PCA factor) and (4) uses the macro factor (real activity) and the curvature factor (third PCA factor).

	(1)	(2)	(3)	(4)
KPST	212.0808	201.8551	197.1613	205.8054
p-value	[0.0511]	[0.1265]	[0.1809]	[0.0910]

for $\Lambda_1 = \binom{\lambda_1^{0\prime}}{\Lambda_2}$, $\lambda_1 : K\times 1$, $\Lambda_2 : (K-1)\times K$. Under $\mathrm{H}_0 : \lambda_1 = \lambda_1^0$, the $N \times 2K$ dimensional reduced rank parameter matrix in the equation for the stacked returns becomes:

$$\Phi = \begin{pmatrix}\beta_1 & \vdots & \beta_2\end{pmatrix}\begin{pmatrix}\lambda_1^{0\prime} & 1 & 0\\ \Lambda_2 & 0 & I_{K-1}\end{pmatrix}, \tag{34}$$

so post-multiplying by $\begin{pmatrix}I_K & -\lambda_1^0 & 0\\ 0 & 0 & I_{K-1}\end{pmatrix}'$ yields the $N\times(2K-1)$ matrix:

$$\Phi\begin{pmatrix}I_K & 0\\ -\lambda_1^{0\prime} & 0\\ 0 & I_{K-1}\end{pmatrix} = \begin{pmatrix}\beta_1 & \vdots & \beta_2\end{pmatrix}\begin{pmatrix}0 & 0\\ \Lambda_2 & I_{K-1}\end{pmatrix} = \beta_2\begin{pmatrix}\Lambda_2 & I_{K-1}\end{pmatrix}, \tag{35}$$

which, since the rank of $\beta_2\begin{pmatrix}\Lambda_2 & I_{K-1}\end{pmatrix}$ equals $K-1$, shows that H_0 implies that the smallest K singular values of Φ times $\begin{pmatrix}I_K & -\lambda_1^0 & 0\\ 0 & 0 & I_{K-1}\end{pmatrix}'$ equal zero. The sFAR statistic for testing $\mathrm{H}_0 : \lambda_1 = \lambda_1^0$:

$$\mathrm{sFAR}(\lambda_1) = \min_{\Lambda_2} \mathrm{FAR}(\Lambda_1(\lambda_1^0, \Lambda_2)), \tag{36}$$

therefore corresponds with a rank test of $\mathrm{H}_0 : \mathrm{rank}\left(\Phi\begin{pmatrix}I_K & 0\\ -\lambda_1^{0\prime} & 0\\ 0 & I_{K-1}\end{pmatrix}\right) = K-1$.

The bounding distribution of the limiting distribution of the sFAR statistic relies upon a Kronecker product structure (KPS) asymptotic covariance matrix of the least squares estimator of the linear model (Guggenberger et al., 2012):

$$\hat{\Phi} = \frac{1}{T}\sum_{t=1}^T R_t\begin{pmatrix}\bar{X}_t\\ \hat{v}_{t+1}\end{pmatrix}'\left[\frac{1}{T}\sum_{t=1}^T\begin{pmatrix}\bar{X}_t\\ \hat{v}_{t+1}\end{pmatrix}\begin{pmatrix}\bar{X}_t\\ \hat{v}_{t+1}\end{pmatrix}'\right]^{-1} = \begin{pmatrix}\hat{d} & \vdots & \hat{\beta}\end{pmatrix}. \tag{37}$$

The KPS thus concerns the asymptotic variance of

$$\frac{1}{\sqrt{T}}\sum_{t=1}^T\left(\begin{pmatrix}\bar{X}_t\\ \hat{v}_{t+1}\end{pmatrix}\otimes\tilde{e}_t\right), \quad \tilde{e}_t = e_t + \beta(\bar{v}_t - \hat{v}_t). \tag{38}$$

We note that $\bar{v}_t$ is not directly observed so it adds additional sampling error because of the imputed estimates, $\hat{v}_t$. To implement the sFAR test, we therefore make the following assumption.

Assumption 4. There exists $\Omega \in \mathbb{R}^{2K\times 2K}$ and $\Sigma \in \mathbb{R}^{N\times N}$ symmetric positive definite matrices such that $S = \Omega \otimes \Sigma$ and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^T\left(\begin{pmatrix}\bar{X}_t\\ \hat{v}_{t+1}\end{pmatrix}\otimes\tilde{e}_t\right) \to_d N(0, S). \tag{39}$$

The asymptotic normality stated in Assumption 4 is a direct result of Assumption 1. If $\hat{v}_t$ is directly observed, so $\tilde{e}_t = e_t$, Assumption 1 implies that $\Omega = \mathbb{E}\left((\bar{X}_t' \vdots \hat{v}_{t+1}')'(\bar{X}_t' \vdots \hat{v}_{t+1}')\right)$ and $\Sigma = \mathrm{var}(e_t)$. Because of the additional sampling error due to the generated regressor $\hat{v}_{t+1}$, Assumption 4 does, however, not provide the exact specifications of Ω and Σ.

We use the KPS test (KPST) from Guggenberger et al. (2022) to test for the proximity of a KPS matrix to the covariance matrix, S. Table 3 reports the KPST results, and shows that the KPS restriction for S is a realistic assumption since none of these tests reject the null hypothesis that the covariance matrix has a KPS at the 5% significance level. A by-product of the KPS test is the KPS factorization for $\hat{S}$, see Guggenberger et al. (2022).

Theorem 2. *Under Assumption 4, and when there is a consistent estimator for S, $\hat{S}$, then in large samples $\hat{S} \approx (\hat{\Omega}\otimes\hat{\Sigma})$, where*

$$\hat{\Omega} = vecinv\left(\begin{pmatrix}\hat{L}_{11}\\ \hat{L}_{21}\end{pmatrix}/\hat{L}_{11}\right), \hat{\Sigma} = vecinv(\hat{L}_{11}\hat{\sigma}_1\hat{N}_1'), \tag{40}$$

where $\hat{L}_{11}, \hat{L}_{21}, \hat{\sigma}_1, \hat{N}_1$ are specified in the proof in the Appendix A. The KPS covariance estimator $\hat{\Omega}\otimes\hat{\Sigma}$ provides a consistent estimator for S.

Table 4
Kleibergen–Paap rank statistic testing $H_0 : \text{rank}(\beta) = K - 1$ (K denotes the number of factors) and its associated [p-value] in square brackets, for various factor combinations. The column heading by (i), for i = 1,...,5, states how many factors are used while the factor combinations are listed in separate cells. All cases use excess returns on bonds with maturities of 2, 3, 12, 60, and 120 months and different combinations of the five PCA factors from Adrian et al. (2013). We mark with one star if the lower bound of the limit sFAR (see Theorem 4) indicates a bounded 95% confidence sets in every direction, and mark with double daggers if the associated 95% confidence sets of the time-varying risk premia parameters for one or more factors are unbounded.

(1)	rank test	(2)	rank test	(3)	rank test	(4)	rank test	(5)	rank test
1*	30 075	1,2	1290	1,2,3	711.1	1,2,3,	932.9	1,2,3,	9.105
	[0.000]		[0.000]		[0.000]	4	[0.000]	4,5††	[0.003]
2*	4409	1,3	1487	1,2,4	1317	1,2,3,	5.103		
	[0.000]		[0.000]		[0.000]	5††	[0.078]		
3*	973.0	1,4	1073	1,2,5	26.03	1,2,4,	76.53		
	[0.000]		[0.000]		[0.002]	5	[0.000]		
4*	703.2	1,5	39.82	1,3,4	603.8	1,3,4,	12.36		
	[0.000]		[0.000]		[0.000]	5	[0.002]		
5*	74.62	2,3	864.7	1,3,5	13.22	2,3,4,	16.23		
	[0.000]		[0.000]		[0.004]	5	[0.000]		
		2,4	990.9	1,4,5	110.3				
			[0.000]		[0.000]				
		2,5	63.17	2,3,4	765.8				
			[0.000]		[0.000]				
		3,4	825.2	2,3,5	16.96				
			[0.000]		[0.001]				
		3,5	11.14	2,4,5	549.0				
			[0.025]		[0.000]				
		4,5	445.2	3,4,5	18.05				
			[0.000]		[0.000]				

Proof. See Appendix A. □

Because of the KPS covariance structure, $(\hat{\Omega} \otimes \hat{\Sigma})$, we can compute the sFAR statistic using the characteristic polynomial stated in Theorem 3.

Theorem 3. *Under Assumptions 3 and 4, let* $\hat{V}_{\hat{\Phi}} = (\hat{\Psi} \otimes \hat{\Sigma})$, $\hat{\Psi} = \hat{W}^{-1}\hat{\Omega}\hat{W}^{-1} = \begin{pmatrix} \hat{\Psi}_X & \hat{\Psi}_{XV} \\ \hat{\Psi}_{VX} & \hat{\Psi}_V \end{pmatrix}$, $\hat{W} = \frac{1}{T}\sum_{t=1}^{T} \begin{pmatrix} \bar{X}_t \\ \hat{v}_{t+1} \end{pmatrix}\begin{pmatrix} \bar{X}_t \\ \hat{v}_{t+1} \end{pmatrix}'$, *the subset FAR statistic, sFAR*(λ_1^0)*, for testing* $H_0 : \lambda_1 = \lambda_1^0$*, for* $\Lambda_1 = \begin{pmatrix} \lambda_1' \\ \Lambda_2 \end{pmatrix}$, $\lambda_1 : K \times 1$, $\Lambda_2 : (K-1) \times K$*, equals T times the sum of the K smallest roots of the characteristic polynomial:*

$$\left| \mu \begin{pmatrix} I_K & 0 \\ -\lambda_1^{0\prime} & 0 \\ 0 & I_{K-1} \end{pmatrix}' \hat{\Psi} \begin{pmatrix} I_K & 0 \\ -\lambda_1^{0\prime} & 0 \\ 0 & I_{K-1} \end{pmatrix} - \begin{pmatrix} I_K & 0 \\ -\lambda_1^{0\prime} & 0 \\ 0 & I_{K-1} \end{pmatrix}' \hat{\Phi}' \hat{\Sigma}^{-1} \hat{\Phi} \begin{pmatrix} I_K & 0 \\ -\lambda_1^{0\prime} & 0 \\ 0 & I_{K-1} \end{pmatrix} \right| = 0, \tag{41}$$

and $\lim_{T\to\infty} sFAR(\lambda_1) \prec \chi^2(K(N-(K-1)))$. *The bound on the limiting distribution holds regardless of Assumption 2.*

Proof. See Appendix A. □

Fig. 4 illustrates the χ^2 bound on the limiting distribution of the sFAR statistic stated in Theorem 3. The density of the limiting distribution of the sFAR statistic is simulated for a two-factor model, which uses the first two PCA factors to mimic the strongly identified case and the third and the fifth PCA factors to mimic weak identification. In Panel (a), the limiting distribution of the sFAR statistic is χ^2 when the model is strongly identified. In Panel (b), under weak identification, the limiting distribution is bounded by the χ^2 distribution. It illustrates that using χ^2 critical values for the sFAR test controls the size of the test.

For projection-based tests on λ_1, the involved test has to be computed over a grid of points concerning the partialled out parameters. This becomes computationally burdensome when the number of partialled out parameters increases because of a larger dimension of Λ_1 reflecting more factors. It makes the sFAR test more empirically appealing because it does not involve an extensive grid search. In practice, Assumption 4 can be further relaxed by using the KPST as a pre-test for conducting robust subset testing as described in Guggenberger et al. (2022).

The value of the sFAR statistic at parameter values distant from zero provides a diagnostic to indicate if the confidence sets of the hypothesized parameters are bounded. These tests are therefore indicative of weak identification.

Theorem 4. *For tests of* $H_0 : \lambda_1 = c\lambda_1^0$ *with* λ_1^0 *a fixed vector of length one and c a scalar, the realized value of the sFAR statistic at a distant value of* λ_1 *in the direction of* λ_1^0, $\lim_{c\to\infty} sFAR(c\lambda_1^0)$*, equals T times the sum of the K smallest roots of*

$$\left| \begin{pmatrix} \bar{\lambda}_{1,\perp}^0 & 0 \\ 0 & I_K \end{pmatrix}' \left[\mu\hat{\Psi} - \hat{\Phi}'\hat{\Sigma}^{-1}\hat{\Phi} \right] \begin{pmatrix} \bar{\lambda}_{1,\perp}^0 & 0 \\ 0 & I_K \end{pmatrix} \right| = 0,$$

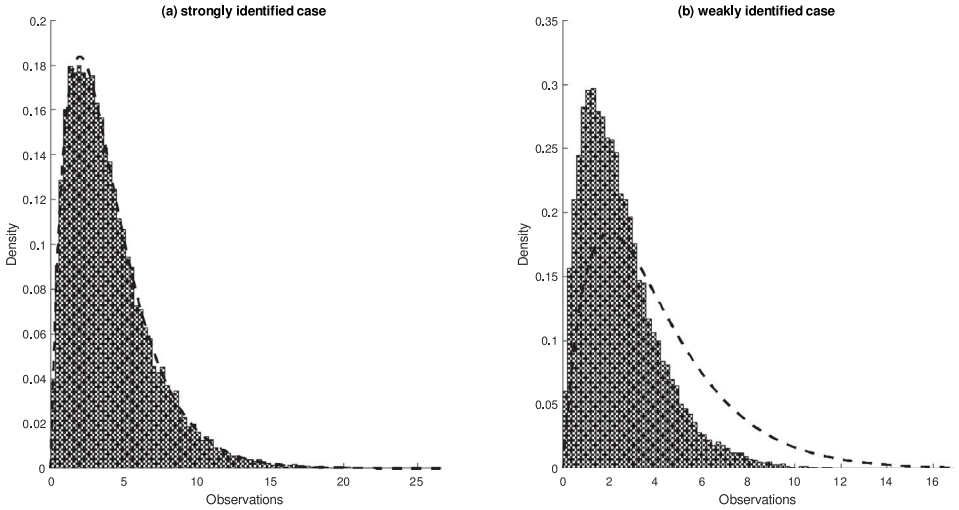

Fig. 4. Simulated density plots of the sFAR test statistic (shadowed bins) and the density function of the χ^2 -distribution (dashed black curve). Panel (a) is for strong identification while Panel (b) is for weak identification. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where $\bar{\lambda}^0_{1,\perp}$ *is a* $K \times (K-1)$ *orthonormal matrix that is orthogonal to* λ^0_1. *The limit sFAR statistic is uniformly bounded from below by the minimum eigenvalue of* $T\hat{\Psi}_V^{-1/2\prime}\hat{\beta}'\hat{\Sigma}^{-1}\hat{\beta}\hat{\Psi}_V^{-1/2}$.

Proof. See Appendix A. □

Theorem 4 provides a way of verifying whether the confidence sets resulting from the sFAR statistic are bounded or unbounded in specific directions, see Dufour (1997), Kleibergen and Zhan (2020), Khalaf and Schaller (2016), Kleibergen et al. (2023). The minimum eigenvalue of $T\hat{\Psi}_V^{-1/2\prime}\hat{\beta}'\hat{\Sigma}^{-1}\hat{\beta}\hat{\Psi}_V^{-1/2}$ is a rank test statistic concerning the rank of the factor loading matrix β, see Kleibergen and Paap (2006), so Theorem 4 shows that the sFAR test evaluated at distant values relates to the rank of β. Theorem 4 also explains that when we encounter weak identification issues with β close to reduced rank, we have unbounded confidence sets. The lower bound is sharp when $K = 1$, as indicated in the proof of Theorem 4, for which case also Theorem 12 in Kleibergen (2021) applies. When $K = 1$ and the Kleibergen–Paap rank test is significant at the 5% significance level, Theorem 4 implies that the sFAR test leads to a bounded 95% confidence sets for λ_1.

Table 4 reports the Kleibergen–Paap rank test for different factor combinations for the data from Adrian et al. (2013). It shows that, in line with Fig. 3, all single-factor model have bounded 95% confidence sets for the time-varying risk premia, which are less likely to be bounded when we include more than three factors. The fifth factor, though identified in a single-factor setting, suffers from weak identification problems when we include other factors. Table 4 also shows that the rank test statistic is a good indicator of unboundedness as small values of the rank test statistics suggest unbounded confidence sets.

5.1. Power of the sFAR test

To illustrate the power of the sFAR test, we compute power curves for two settings calibrated to the data discussed previously. Fig. 5 therefore shows the two-dimensional power curves that result when jointly testing the two risk premia parameters associated with a single factor in a two-factor model. The left hand side of Fig. 5 shows the power curves for a strongly identified setting while the right hand side does so for a weakly identified setting. The power curves on the right hand side show that the sFAR test is not consistent for weakly identified settings since the rejection frequencies do not converge to one when we move away from the hypothesized value.

5.2. Identification robust confidence sets for risk premia

We use the sFAR test to construct confidence sets on the risk premia resulting from two and three factor models. Fig. 6 shows the 90, 95 and 99% joint confidence sets that result for the two risk premia parameters in Λ_1 resulting for one specific factor in a two-factor model using the data from Adrian et al. (2013) while Fig. 7 does so for the three risk premia parameters resulting for one specific factor in a three-factor model. Size correct confidence sets for the individual risk premia result by projecting the joint confidence sets on the axes. When using four or more factors, the number of risk premia concerning one factor is at least four so we have to use projection-based tests based on the sFAR statistic to be able to visualize these confidence sets. For expository purposes

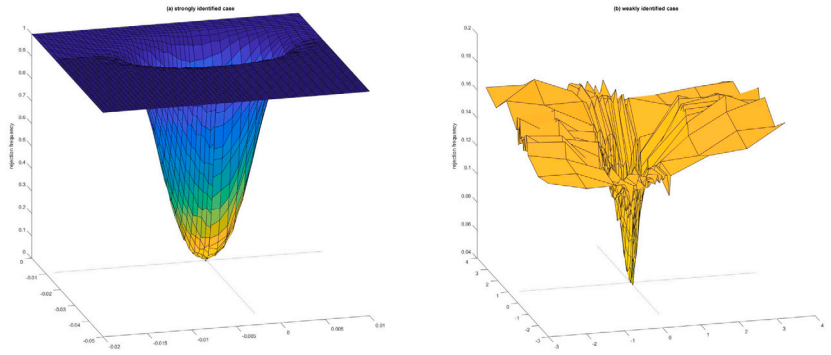

Fig. 5. Simulated power surfaces (rejection frequencies) of sFAR tests on Λ_1's w.r.t. the first factor of a two-factor model: the left panel (a) is a strong identified setting calibrated to the two-factor model with excess returns of bonds with maturities 3, 60, 120 months using the first and second PCA factors, while the right panel (b) is a weakly identified setting calibrated to the two-factor model with excess returns of bonds with maturities 3, 60, 120 months using the third and fifth PCA factors. Dotted lines $(\gamma_i, y, 0.05)$ are drawn to mark the positions of the calibrated risk premia values, γ_i's, at 5% level.

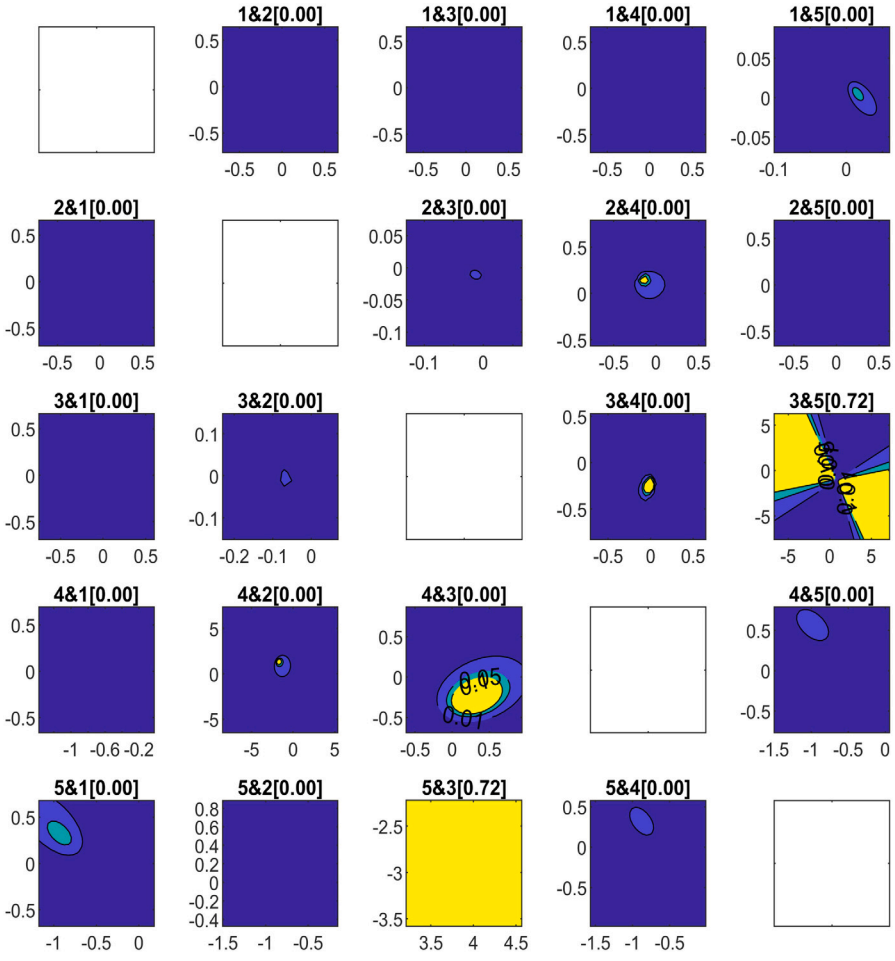

Fig. 6. Joint confidence sets from the sFAR test for the two risk premia parameters of the first of the two listed factors in a two factor model. (yellow 90%, light green 95%, light blue 99%, dark blue area contains the remaining values). Excess returns on bonds with maturities 3, 60, 120 months are used. [p-value] of Kleibergen–Paap rank statistic testing H_0 : rank$(\beta) = 1$ in square brackets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

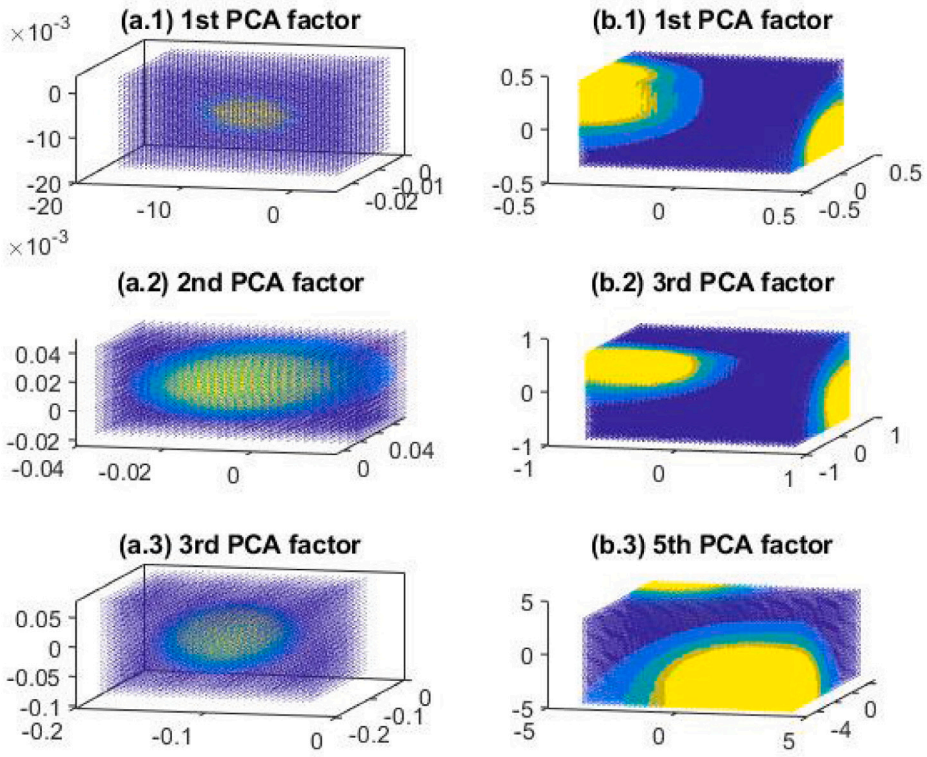

Fig. 7. Joint confidence sets from the sFAR test for three-factor models (yellow 90%, light green 95%, light blue 99%, dark blue area contains the remaining values) with excess returns on bonds with maturities 3, 60, 120 months, (a.i) testing on Λ_1's w.r.t. the ith factor when using first, second and third PCA factors Kleibergen–Paap rank statistic testing H_0 :rank(β) = 2 equals 310.9294 [p-value: 0.0000]); (b.i) testing on Λ_1's w.r.t. the $(i+2)$th factor when using first, third and fifth PCA factors. Kleibergen–Paap rank statistic testing H_0 :rank(β) = 2 equals 0.0654 [p-value 0.7981]). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and since Table 4 shows that some of these confidence sets are unbounded, for example, the one that results when using all five factors, we therefore refrain from using more than three factors.[3]

Fig. 6 shows all two dimensional confidence sets for the two risk premia resulting for one factor for all different specifications using the five PCA factors discussed previously in a two factor model. The two dimensional confidence sets in Fig. 6 vary a lot. Quite a few are empty so all values of the parameters are rejected at significance levels which exceed 99%. This occurs, for example, when using the first and either the second, third and fourth PCA factor so the model is misspecified. There are also settings where the confidence set is bounded and well behaved which occurs, for example, when using the third and fourth PCA factor. Other confidence sets are unbounded and/or cover the whole two-dimensional space, which occurs, for instance, when using the third and fifth PCA. For this combination the 90% confidence set for the two risk premia on the third factor is unbounded but excludes an area in the parameter space while the 90% confidence set of the two risk premia on the fifth factor covers the whole two-dimensional space. Table 4 also shows that the combination of the third and fifth PCA factors leads to a smaller rank test statistic than other factor combinations when using two-factor models which is in line with the unbounded confidence set in Fig. 6 which relates to the third and fifth PCA factors. This is all indicative of weak identification when using both the third and fifth factors.

Fig. 7 shows the joint confidence sets for the three risk premia associated with a single factor in a three-factor model. The first column of Fig. 7 does so for a factor model containing the first three principal components as factors while the second column does so using the first, third and fifth principal components as factors. Unlike when using two factors, the first column shows that the confidence sets are no longer empty but bounded which shows that the parameters for the first three PCA factors are well identified and that the model is no longer misspecified. This is confirmed by the p-value of the rank test on the β's. This is in contrast when using the first, third and fifth principal components as factors. The confidence sets in the second column of Fig. 7 are namely all unbounded indicating weak identification of the risk premia which is further reflected by the p-value of the rank test on the β's. Table 4 also shows that the model including the first, third, and fifth PCA factors has a much smaller rank test statistic than when using the first three PCA factors in a three-factor model.

[3] The rank tests in Table 4 and Figs. 6–7 are not identical. For expository purposes, we choose a smaller number of test assets in Figs. 6–7.

The left hand side column of Fig. 7 shows the confidence sets for the first three PCA factors. These factors are proxies for the "level", "slope" and "curvature" factors which are constructed from observed yields, i.e. the level factor equals the difference between long and short yields, and the slope and curvature factors are a combination of short, intermediate, and long yields. Instead of the first three PCA factors, the level, slope and curvature factors could as well be used. The confidence sets for the risk premia in Fig. 7 for the first three PCA factors are all bounded which shows that they are well identified which is similarly confirmed by the large value of the rank test. The standard risk premia on the level, slope and curvature factors can also be computed using their average return on a trading strategy mimicking the respective factor. The confidence sets in Fig. 7, are, however, for the time-varying component of the risk premia which could as well be computed using a multi-pass estimation procedure using the return from a trading strategy mimicking the respective factor.[4]

Fig. 7 shows the three dimensional confidence sets that result from the sFAR test. It results from partialling out the six risk premia parameters associated with the other two factors. Hence, when we compute these confidence sets using projection with the identification robust tests, we have to specify a nine-dimensional grid for the parameters and compute the identification robust tests for all values on this nine-dimensional grid. This is, or is close to be, computationally infeasible. Hence, the sFAR test provides a computationally tractable manner to conduct identification robust tests on larger number of parameters.

5.3. Practical takeaway for when weak identification occurs

The identification robust confidence sets discussed previously can have different shapes and not always conform with the bounded Wald type confidence sets most researchers are used to. When the risk premia parameters are well identified, the resulting identification robust confidence sets are bounded and convex and comparable to the Wald type confidence sets researchers are used to. In case of weak identification, the identification robust confidence sets can, however, be unbounded which indicates that the data is not informative about the risk premia parameters. The Wald type confidence sets remain bounded when identification is weak but are unreliable because they mistakenly give the impression that the data is informative about the risk premia parameter while it is not. Weak identification is very unsatisfactory because it leaves the researcher kind of empty-handed, and no conclusions can be drawn about the parameters of interest. In such settings, it is important to try to improve the identification strength which can be achieved in different ways:

1. Limit the number of factors or remove weak factors from the model. Our previous analyzes show that the higher order PCA and macro factors are much weaker than the first three PCA factors. When estimating risk premia parameters using all these factors, they become weakly identified so little can be learned about them. When we instead just use the first three principal components as factors, Fig. 7 shows that bounded confidence sets are obtained which lead to meaningful inference.
2. Impose restrictions on the risk premia. When certain risk premia are thought to be identical, identification can be improved by equating them. It would especially be of interest to use economic restrictions for this purpose, like, for example, no-arbitrage conditions.

6. Conclusion

We propose identification robust test procedures for testing hypotheses on risk premia in DATSMs. The robust subset factor Anderson–Rubin test extends the sFAR test from the linear asset pricing model to allow for tests on multiple risk premia parameters and, unlike projection based testing, provides a computationally tractable manner to conduct identification robust tests on larger number of parameters. Our empirical results show that especially in case of multiple factors, weak identification is pervasive and traditional tests are likely misleading. We use the empirical settings from the literature on affine term structure models, see e.g. Adrian et al. (2013) and Ang and Piazzesi (2003)), to illustrate our results and the importance of using weak identification robust test procedures.

Appendix: Proofs

Appendix A. Proof of Theorem 2

Proof of Theorem 2. Suppose that there exists a consistent estimator for S, $\hat{S}$, a by-product of the Kronecker Product Structure (KPS) test is the KPS factorization for $\hat{S}$ (Guggenberger et al., 2022). We briefly discuss how it operates. Let $R \in \mathbb{R}^{kp\times kp}$ be a matrix with a block structure

$$R := \begin{pmatrix} R_{11} & \cdots & R_{1p} \\ \vdots & \ddots & \vdots \\ R_{p1} & \cdots & R_{pp} \end{pmatrix},$$

where $R_{jl} \in \mathbb{R}^{k\times k}, j,l = 1,\ldots,p$. Define $\mathcal{R}(R) := \begin{pmatrix} R_1' \cdots R_p' \end{pmatrix}' \in \mathbb{R}^{p^2\times k^2}$ with $R_j := \begin{pmatrix} \operatorname{vec}(R_{1j}) \cdots \operatorname{vec}(R_{pj}) \end{pmatrix}' \in \mathbb{R}^{p\times k^2}$, $j = 1,\ldots,p$.

[4] We thank an anonymous referee for pointing us at the relationship between the first three PCA factors and the level, slope and curvature factors.

Consider a singular value decomposition (SVD) of $\mathcal{R}(\hat{S})$: $\mathcal{R}(\hat{S}) = \hat{L}\hat{\Sigma}\hat{N}'$ where $p = 2K$, $k = N$, $\hat{\Sigma} := \text{diag}\left(\hat{\sigma}_1 \dots \hat{\sigma}_{\min(p^2,k^2)}\right)$ denotes a $p^2 \times k^2$ dimensional diagonal matrix with the singular values $\hat{\sigma}_j$ $\left(j = 1, \dots, \min\left(p^2, k^2\right)\right)$ non-increasingly ordered on the main diagonal, and $\hat{L} \in \mathbb{R}^{p^2 \times p^2}$, $\hat{N} \in \mathbb{R}^{k^2 \times k^2}$ orthonormal matrices. Decompose

$$\hat{L} := \begin{pmatrix} \hat{L}_{11} & \hat{L}_{12} \\ \hat{L}_{21} & \hat{L}_{22} \end{pmatrix} = \left(\hat{L}_1 \ \vdots \ \hat{L}_2\right), \hat{\Sigma} := \begin{pmatrix} \hat{\sigma}_1 & 0 \\ 0 & \hat{\Sigma}_2 \end{pmatrix},$$

$$\hat{N} := \begin{pmatrix} \hat{N}_{11} & \hat{N}_{12} \\ \hat{N}_{21} & \hat{N}_{22} \end{pmatrix} = \left(\hat{N}_1 \ \vdots \ \hat{N}_2\right),$$

with $\hat{L}_{11} : 1 \times 1$, $\hat{L}_{12} : 1 \times \left(p^2 - 1\right)$, $\hat{L}_{21} : \left(p^2 - 1\right) \times 1$, $\hat{L}_{22} : \left(p^2 - 1\right) \times \left(p^2 - 1\right)$, $\hat{\sigma}_1 : 1 \times 1$, $\hat{\Sigma}_2 : \left(p^2 - 1\right) \times \left(k^2 - 1\right)$, $\hat{N}_{11} : 1 \times 1$, $\hat{N}_{12} : 1 \times \left(k^2 - 1\right)$, $\hat{N}_{21} : \left(k^2 - 1\right) \times 1$, $\hat{N}_{22} : \left(k^2 - 1\right) \times \left(k^2 - 1\right)$ dimensional matrices. The consistency arises from the consistency of $\hat{S}$ and Theorem 1 from Guggenberger et al. (2022). □

Appendix B. Proof of Theorem 3

We provide a proof of Theorem 3. Before we do so, we first state two lemmas that enable us to prove it in a succinct manner. The proofs of these Lemmas are given in Appendix B.1 which succeeds the concise proof of Theorem 3.

Lemma 1. *Under Assumptions of Theorem 3, the sFAR(λ_1^0) statistic equals the sum of the K smallest roots of the following polynomial:*

$$\left|\mu I_{2K-1} - \hat{\Xi}'\hat{\Xi}\right| = 0,$$

where $\hat{\Xi} = [\xi_u \ \vdots \ \xi_{\beta_2}] \sim_a N(\mathcal{M}, I_{N(2K-1)})$, "$\sim_a$" denotes an approximate distribution in large samples. The specifications of ξ_u, ξ_{β_2} and the parameter matrix $\mathcal{M}$ are stated in the proof.

Comments on Lemma 1.

1. The $N \times (2K - 1)$ dimensional matrix $\mathcal{M}$ is non-random (and defined in (B.47)). It depends on the parameters of the data generating process. Under the null hypothesis $H_0 : \lambda_1 = \lambda_1^0$, $\mathcal{M}$ can be expressed as $(0_{N,K}, M)$ for some nonrandom $N \times (K-1)$ dimensional matrix M. When, for example, all β's are zero, we have $M = 0$. The eigenvalues of $\mathcal{M}'\mathcal{M}$ indicate the strength of identification.
2. In large samples, $\hat{\Xi}'\hat{\Xi}$ is approximately (non-central) Wishart distributed with N degrees of freedom, covariance matrix I_{2K-1}, and non-centrality parameter matrix $\mathcal{M}'\mathcal{M}$, denoted by $\mathcal{W}_{2K-1}\left(N, I_{2K-1}, \mathcal{M}'\mathcal{M}\right)$.
3. Instead of being equal to the smallest root of a characteristic polynomial which is the setting in, e.g., Guggenberger et al. (2012), Guggenberger et al. (2019)), Lemma 1 shows that the subset FAR statistic equals the sum of the K smallest roots. Theorem 3 therefore provides an important extension to the bounding result in Guggenberger et al. (2012).

Lemma 2. *Let:*

(a) $\mathcal{M}_T$ be a sequence of the parameter matrix $\mathcal{M}$ from Lemma 1;
(b) $\mathbb{M}$ denote the collection of all such sequences under the null hypothesis $H_0 : \lambda_1 = \lambda_1^0$;
(c) $\kappa_i(A)$ denote the ith eigenvalue of $A'A$ ordered non-increasingly;
(d) $\mathbb{M}_\infty$ be a subset of $\mathbb{M}$ that contains all sequences $\mathbb{M}_T$ satisfying $Q_T \to Q$ with Q_T being an orthonormal matrix whose columns are eigenvectors of $\mathcal{M}_T'\mathcal{M}_T$ and $\kappa_1(\mathcal{M}_T) > \dots > \kappa_K(\mathcal{M}_T) = \kappa_{K+1}(\mathcal{M}_T) = \dots = 0, \kappa_{K-1}(\mathcal{M}_T) \to \infty$;
(e) $\tilde{\Xi}_{\mathcal{M}} = Z + \mathcal{M}$ with Z being an $N \times (2K - 1)$ random matrix and $vec(Z) \sim N(0_{N(2K-1),1}, I_N \otimes I_{2K-1})$;

then:

i for any M_T in M, $\kappa_j(\tilde{\Xi}_{\mathcal{M}_T}) = O_p(1)$, $j \geq K$;
ii for any M_T in M, we can find a parameter sequence $\tilde{\mathcal{M}}_h$ in M_∞ such that $\limsup_{T\to\infty} \kappa_j(\tilde{\Xi}_{\mathcal{M}_T}) \leq \liminf_{h\to\infty} \kappa_j(\tilde{\Xi}_{\tilde{\mathcal{M}}_h}), j \geq K$;
iii under the assumptions of Lemma 1 and any sequence of parameter matrices in M_∞, the ordered smallest K eigenvalues, $\hat{\kappa}_j = \kappa_j(\hat{\Xi}), j \geq K$, converge in distribution to the (ordered) smallest eigenvalues of W, with W (central) Wishart distributed with N degrees of freedom and covariance matrix I_K, i.e., $W \sim W_K\left(N - (K-1), I_K, 0_K\right)$.

Proof of Theorem 3. The equality of the sFAR statistic with the sum of the K smallest eigenvalues of a non-central Wishart distributed random matrix results from Lemma 1.

Lemma 2.(ii) implies that for deriving an upper bound on the limiting distribution of the sFAR statistic, we only have to analyze the limiting distributions of the K smallest eigenvalues that result under parameter sequences in $\mathbb{M}_\infty$.

Lemma 2.(iii) and because the trace of a central Wishart distributed random matrix has a (central) χ^2 distribution, see, e.g., Muirhead (2009), then lead to the bounding result $\lim_{T\to\infty} \text{sFAR}(\lambda_1^0) \prec \chi^2(K(N-(K-1)))$. □

B.1. Proofs of Lemmas 1 and 2

This section contains proofs of the lemmas needed for the proof of Theorem 3.

Proof of Lemma 1. The FAR statistic reads,

$$\text{FAR}(\Lambda_1^0) = \quad T \times f_T\,(\Lambda_1^0, X)'\hat{V}_{ff}(\Lambda_1^0)^{-1} f_T(\Lambda_1^0, X). \tag{B.42}$$

The sample moments $f_T(\Lambda_1, X)$ of the model under consideration for given Λ_1 are (22):

$$f_T(\Lambda_1, X) = \frac{1}{T}\sum_{t=1}^{T}(\bar{X}_{t-1} \otimes (\bar{R}_t - \hat{\beta}\hat{V}_t)) - \left(\hat{Q}_{XX} \otimes \hat{\beta}\right) \text{vec}(\Lambda_1).$$

Using that $\hat{d}$ is the least squares estimator resulting from step 2 of the three-step procedure:

$$\text{vec}(\hat{d}) = \left(\hat{Q}_{XX} \otimes I_N\right)^{-1} \frac{1}{T}\sum_{t=1}^{T}(\bar{X}_{t-1} \otimes (\bar{R}_t - \hat{\beta}\hat{V}_t)),$$

we have that

$$f_T(\Lambda_1, X) = \left(\hat{Q}_{XX} \otimes I_N\right) \text{vec}\left(\hat{\Phi} A(\lambda_1)\pi_{\Lambda_2}\right), \tag{B.43}$$

with

$$\hat{\Phi} = \tfrac{1}{T}\textstyle\sum_{t=1}^{T} R_t \begin{pmatrix} \bar{X}_t \\ \hat{v}_{t+1} \end{pmatrix}' \left[\tfrac{1}{T}\textstyle\sum_{t=1}^{T} \begin{pmatrix} \bar{X}_t \\ \hat{v}_{t+1} \end{pmatrix}\begin{pmatrix} \bar{X}_t \\ \hat{v}_{t+1} \end{pmatrix}'\right]^{-1} = \left(\hat{d} \;\vdots\; \hat{\beta}\right), \; A(\lambda_1) = \begin{pmatrix} I_K & -\lambda_1 & 0 \\ 0 & 0 & I_{K-1} \end{pmatrix}', \; \pi_{\Lambda_2} = \begin{pmatrix} I_K \\ -\Lambda_2 \end{pmatrix}.$$

Provided that Assumption 4 holds, the above result implies that

$$\begin{aligned} \hat{V}_{ff}(\Lambda_1^0) &= \left(\hat{Q}_{XX}\pi'_{\Lambda_2^0} A(\lambda_1^0)' \otimes I_N\right) \hat{V}_{\hat{\Phi}} \left(A(\lambda_1^0)\pi_{\Lambda_2^0}\hat{Q}_{XX} \otimes I_N\right) \\ &= \left(\hat{Q}_{XX} \otimes I_N\right)\left(\pi'_{\Lambda_2^0} A(\lambda_1^0)'\hat{\Psi} A(\lambda_1^0)\pi_{\Lambda_2^0} \otimes \hat{\Sigma}\right)\left(\hat{Q}_{XX} \otimes I_N\right). \end{aligned} \tag{B.44}$$

Substituting (B.43) and (B.44) into (B.42) then gives

$$\begin{aligned} \text{FAR}(\lambda_1^0, \Lambda_2^0) &= T\text{vec}\left(\hat{\Phi} A(\lambda_1^0)\pi_{\Lambda_2^0}\right)'\left(\pi'_{\Lambda_2^0} A(\lambda_1^0)'\hat{\Psi} A(\lambda_1^0)\pi_{\Lambda_2^0} \otimes \hat{\Sigma}\right)^{-1} \text{vec}\left(\hat{\Phi} A(\lambda_1^0)\pi_{\Lambda_2^0}\right) \\ &= T\text{vec}\left(\hat{\Phi} A(\lambda_1^0)\pi_{\Lambda_2^0}\right)' \text{vec}\left(\hat{\Sigma}^{-1}\hat{\Phi} A(\lambda_1^0)\pi_{\Lambda_2^0}\left(\pi'_{\Lambda_2^0} A(\lambda_1^0)'\hat{\Psi} A(\lambda_1^0)\pi_{\Lambda_2^0}\right)^{-1}\right), \end{aligned}$$

which using the trace operator can be rewritten as

$$\text{FAR}(\lambda_1^0, \Lambda_2^0) = T\text{tr}\left(\pi'_{\Lambda_2^0} A(\lambda_1^0)'\hat{\Phi}'\hat{\Sigma}^{-1}\hat{\Phi} A(\lambda_1^0)\pi_{\Lambda_2^0}\left[\pi'_{\Lambda_2^0} A(\lambda_1^0)'\hat{\Psi} A(\lambda_1^0)\pi_{\Lambda_2^0}\right]^{-1}\right).$$

Let $\widetilde{\text{FAR}}(\lambda_1^0, q) = T\text{tr}\left(q' A(\lambda_1^0)'\hat{\Phi}'\hat{\Sigma}^{-1}\hat{\Phi} A(\lambda_1^0) q\left[q' A(\lambda_1^0)'\hat{\Psi} A(\lambda_1^0) q\right]^{-1}\right)$, where q is a $(2K-1)\times K$ matrix and $q' A(\lambda_1^0)'\hat{\Psi} A(\lambda_1^0) q = I_K$ and let q^{**} be a collection of such q's. Denote

$$\hat{q} = \arg\min_{q \in q^{**}} \widetilde{\text{FAR}}(\lambda_1^0, q).$$

Theorem 1.2 from Sameh and Wisniewski (1982) implies that $\widetilde{\text{FAR}}(\lambda_1^0, \hat{q})$ equals T times the sum of the K smallest eigenvalues of the eigenvalue problem:

$$A(\lambda_1^0)'\hat{\Phi}'\hat{\Sigma}^{-1}\hat{\Phi} A(\lambda_1^0)x = \kappa\hat{\Psi}(\lambda_1^0)x,$$

where $\hat{\Psi}(\lambda_1^0) = A(\lambda_1^0)'\hat{\Psi} A(\lambda_1^0)$, x is a $(2K-1)$-eigenvector, κ is a scalar eigenvalue, and $\hat{q}$ corresponds to the associated eigenvectors of $A(\lambda_1^0)'\hat{\Phi}'\hat{\Sigma}^{-1}\hat{\Phi} A(\lambda_1^0)$. Let $\hat{w}_u$ denote the upper $K\times K$ block matrix of $\hat{q}$, and let $\hat{w}_d$ denote the lower block matrix. It follows that $\text{FAR}(\lambda_1^0, \hat{\Lambda}_2) = \widetilde{\text{FAR}}(\lambda_1^0, \hat{q})$ where $\hat{\Lambda}_2 = -\hat{w}_d\hat{w}_u^{-1}$, as constructed.

T times the sum of the K smallest eigenvalues, κ's, of the above eigenvalue problem equals T times the sum of the K smallest roots of the characteristic polynomial:

$$\left|\mu\hat{\Psi}(\lambda_1^0) - T A(\lambda_1^0)'\hat{\Phi}'\hat{\Sigma}^{-1}\hat{\Phi} A(\lambda_1^0)\right| = 0. \tag{B.45}$$

Pre/post-multiplying $\left|\mu\hat{\Psi}(\lambda_1^0) - T A(\lambda_1^0)'\hat{\Phi}'\hat{\Sigma}^{-1}\hat{\Phi} A(\lambda_1^0)\right| = 0$ by $\begin{pmatrix} I_K & -\Lambda_2' \\ 0 & I_{K-1} \end{pmatrix}$ gives

$$\left|\mu\hat{\Psi}(\lambda_1^0, \Lambda_2) - T\left(\hat{u}, \hat{\beta}_2\right)'\hat{\Sigma}^{-1}\left(\hat{u}, \hat{\beta}_2\right)\right| \quad = 0 \;, \tag{B.46}$$

where

$$\hat{\Psi}(\lambda_1^0, \Lambda_2) = \begin{pmatrix} I_K & -\Lambda_2' \\ 0 & I_{K-1} \end{pmatrix}' \hat{\Psi}(\lambda_1^0) \begin{pmatrix} I_K & -\Lambda_2' \\ 0 & I_{K-1} \end{pmatrix} = \begin{pmatrix} \hat{\Psi}_u & \hat{\Psi}_{u\beta_2} \\ \hat{\Psi}_{u\beta_2}' & \hat{\Psi}_{\beta_2} \end{pmatrix},$$

and $\hat{u} = \hat{d} - \hat{\beta}_1 \lambda_1^{0\prime} - \hat{\beta}_2 \Lambda_2 = \epsilon + \hat{\beta}_1 \left(\lambda_1 - \lambda_1^0\right)'$. Note that $C\hat{\Psi}(\lambda_1^0, \Lambda_2)C' = I_{2K-1}$ holds with

$$C = \begin{pmatrix} \hat{\Psi}_{uu}^{-\frac{1}{2}} & 0 \\ -\hat{\Psi}_{\beta_2\beta_2\cdot u}^{-\frac{1}{2}} \hat{\Psi}_{u\beta_2}' \hat{\Psi}_{uu}^{-1} & \hat{\Psi}_{\beta_2\beta_2\cdot u}^{-\frac{1}{2}} \end{pmatrix}, \hat{\Psi}_{\beta_2\beta_2\cdot u} = \hat{\Psi}_{\beta_2\beta_2} - \hat{\Psi}_{u\beta_2}' \hat{\Psi}_{uu}^{-1} \hat{\Psi}_{u\beta_2}.$$

Pre/post-multiplying Eq. (B.46) by $|C|$ yields $\left|\mu I_{2K-1} - \hat{\Xi}'\hat{\Xi}\right| = 0$, where

$$\begin{aligned} &\xi_u = \sqrt{T}\hat{\Sigma}^{-\frac{1}{2}} \hat{u} \hat{\Psi}_{uu}^{-\frac{1}{2}}, \xi_{\beta_2} = \sqrt{T}\hat{\Sigma}^{-\frac{1}{2}} \left(\hat{\beta}_2 - \hat{u}\hat{\Psi}_{uu}^{-1}\hat{\Psi}_{u\beta_2}'\right) \hat{\Psi}_{\beta_2\beta_2\cdot u}^{-\frac{1}{2}}, \\ &\mathcal{M} = \left(\sqrt{T}\Sigma^{-\frac{1}{2}} \beta_1 (\lambda_1 - \lambda_1^0)' \Psi_{uu}^{-\frac{1}{2}} \ \vdots \ \sqrt{T}\Sigma^{-\frac{1}{2}} \left(\beta_2 - \beta_1(\lambda_1 - \lambda_1^0)'\Psi_{uu}^{-1} \Psi_{u\beta_2}'\right) \Psi_{\beta_2\beta_2\cdot u}^{-\frac{1}{2}}\right), \end{aligned} \tag{B.47}$$

and $\hat{\Psi}_{uu}, \hat{\Psi}_{u\beta_2}, \hat{\Psi}_{\beta_2\beta_2\cdot u}$ converge to $\Psi_{uu}, \Psi_{u\beta_2}, \Psi_{\beta_2\beta_2\cdot u}$ respectively in probability. Therefore, the eigenvalue problem (B.45) is equivalent to the eigenvalue problem $\left|\mu I_{2K-1} - \hat{\Xi}'\hat{\Xi}\right| = 0$, and the test statistic equals the sum of the K smallest eigenvalues of $\hat{\Xi}'\hat{\Xi}$. □

Proof of Lemma 2. (i) Under the null hypothesis $H_0 : \lambda_1 = \lambda_1^0$, $\mathcal{M} = \left(0_{N,K}, \sqrt{T}\Sigma^{-\frac{1}{2}}\beta_2\Psi_{\beta_2\beta_2\cdot u}^{-\frac{1}{2}}\right)$, so the rank of $\mathcal{M}$ is smaller than or equal to $K-1$. Therefore, the null space of $\mathcal{M}$, denoted by $N(\mathcal{M})$, is at least K dimensional. In the following of this proof, we use $\Xi = \tilde{\Xi}_{\mathcal{M}} = Z + \mathcal{M}$.

For an arbitrary n-by-n real symmetric matrix A, the kth largest eigenvalue, via min–max characterization (also known as the Courant–Fisher expression), can be expressed as $\min_{\substack{U \ s.t. \\ \dim(U)=n-k+1}} \max_{x\in U}\{x'Ax : \|x\| = 1\}$, where the first minimum is over all $(n-k+1)$-dimensional subspaces U of $\mathbb{R}^n$. Therefore, employing this characterization, the jth eigenvalue of $\Xi'\Xi$ is

$$\hat{\kappa}_j = \min_{\substack{U \ s.t. \\ \dim(U)=2K-1-j+1}} \max_{x\in U}\{x'(Z+\mathcal{M})'(Z+\mathcal{M})x : \|x\| = 1\}.$$

For $j \geq K$, $2K-1-j+1 \leq K$, we can choose $U \subseteq N(\mathcal{M})$, and note that

$$\max_{x\in N(\mathcal{M})}\{x'(Z+\mathcal{M})'(Z+\mathcal{M})x : \|x\| = 1\} = \max_{x\in N(\mathcal{M})}\{x'Z'Zx : \|x\| = 1\} = O_p(1),$$

which implies that $\hat{\kappa}_j = O_p(1)$. □

Proof of Lemma 2. (ii) We show that for any given sequence $\mathcal{M}_n$ in $\mathbb{M}$, there exists a sequence $\tilde{\mathcal{M}}_h$ that belongs to $\mathbb{M}_\infty$, satisfying the condition $\limsup_{n\to\infty} \kappa_j(\tilde{\Xi}_{\mathcal{M}_n}) \leq \liminf_{h\to\infty} \kappa_j(\tilde{\Xi}_{\tilde{\mathcal{M}}_h}), j \geq K$.

To construct the sequence, we start by selecting a sub-sequence of $\mathcal{M}_n$, denoted by $\mathcal{M}_{n_h}$, such that $\lim_{h\to\infty} \kappa_i(\tilde{\Xi}_{\mathcal{M}_{n_h}}) = \limsup_{n\to\infty} \kappa_i(\tilde{\Xi}_{\mathcal{M}_n})$, and $Q_{n_h} \to Q$, where $\mathcal{M}_{n_h}'\mathcal{M}_{n_h} = Q_{n_h}P_{n_h}Q_{n_h}'$, Q_{n_h} is an orthogonal matrix whose columns are eigenvectors of $\mathcal{M}_{n_h}'\mathcal{M}_{n_h}$ and P_{n_h} is a diagonal matrix whose entries are the eigenvalues $\kappa_i(\mathcal{M}_{n_h})$. Denote by $Q_{n_h,i}$ the eigenvector associated with $\kappa_i(\mathcal{M}_{n_h})$, and by construction $Q_{n_h,i} \to Q_i$. Since we allow weak identification, $\kappa_i(\mathcal{M}_{n_h})$ with $i \leq K-1$ can be close to zero when all factors are weak, e.g., $\beta_2 = O(1/\sqrt{T})$. Parameter sequences in $\mathbb{M}_\infty$ thus correspond to "non-weakly identified" data generating processes.

We choose $\tilde{\mathcal{M}}_h = \tilde{P}_h^{\frac{1}{2}} Q_{n_h}'$ with $\tilde{P}_h$ being a diagonal matrix with diagonal entries $\tilde{P}_{h,ii} = \kappa_{h,i}$ such that $\kappa_{h,1} > \cdots > \kappa_{h,K} = \kappa_{h,K+1} = \cdots = 0$, $\kappa_{h,K-1} \to \infty$ and $\kappa_{h,1} = o\left(\left(\sum_{1\leq i\leq 2K-1} \|Q_{n_h,i} - Q_i\|_{\max} + h^{-1}\right)^{-1}\right)$. Note that the following two properties hold by construction,

1. For $j \geq K$,

$$\liminf_h \kappa_j(\tilde{\Xi}_{\tilde{\mathcal{M}}_h}) = \liminf_h \min_{\substack{U \ s.t. \\ \dim(U)= \\ 2K-1-j+1; \\ U\perp \\ \text{span}(Q_i, i<K)}} \max_{\substack{x\in U, \\ \|x\|=1}} \{x'\tilde{\Xi}_{\tilde{\mathcal{M}}_h}'\tilde{\Xi}_{\tilde{\mathcal{M}}_h}x\}, \tag{B.48}$$

   where the minimum is over all $(2K-1-j+1)$-dimensional subspaces U of $\mathbb{R}^{2K-1}$ that are orthogonal to the linear spaces spanned by vectors $Q_i, i < K$.
2. For $j \geq K$,

$$\liminf_{h\to\infty} \kappa_j(\tilde{\Xi}_{\tilde{\mathcal{M}}_h}) = \lim_{h\to\infty} \kappa_j(\tilde{\Xi}_{\tilde{\mathcal{M}}_h}). \tag{B.49}$$

These two properties lead directly to the final conclusion that for $j \geq K$,

$$\begin{aligned}
&\limsup_{n\to\infty} \kappa_j(\tilde{\Xi}_{\mathcal{M}_n}) = \lim_{h\to\infty} \kappa_j(\tilde{\Xi}_{\mathcal{M}_{n_h}}) \\
= &\lim_{h\to\infty} \min_{\substack{U \text{ s.t.} \\ \dim(U)=2K-1-j+1}} \max_{x\in U} \{x' \tilde{\Xi}'_{\mathcal{M}_{n_h}} \tilde{\Xi}_{\mathcal{M}_{n_h}} x : \|x\| = 1\} \\
\leq &\lim_{h\to\infty} \min_{\substack{U \text{ s.t.} \\ \dim(U)=2K-1-j+1, \\ U\perp \text{span}(Q_{n_h,i}, i<K)}} \max_{x\in U} \{x' \tilde{\Xi}'_{\mathcal{M}_{n_h}} \tilde{\Xi}_{\mathcal{M}_{n_h}} x : \|x\| = 1\} \\
= &\lim_{h\to\infty} \min_{\substack{U \text{ s.t.} \\ \dim(U)=2K-1-j+1, \\ U\perp \text{span}(Q_{n_h,i}, i<K)}} \max_{x\in U} \{x' Z' Z x : \|x\| = 1\} \\
= &\lim_{h\to\infty} \min_{\substack{U \text{ s.t.} \\ \dim(U)=2K-1-j+1, \\ U\perp \text{span}(Q_{i}, i<K)}} \max_{x\in U} \{x' Z' Z x + o_p(1) : \|x\| = 1\} \\
= &\lim_{h\to\infty} \min_{\substack{U \text{ s.t.} \\ \dim(U)=2K-1-j+1, \\ U\perp \text{span}(Q_{i}, i<K)}} \max_{x\in U} \{x' \tilde{\Xi}'_{\tilde{\mathcal{M}}_h} \tilde{\Xi}_{\tilde{\mathcal{M}}_h} x : \|x\| = 1\} \\
= &\lim_{h\to\infty} \kappa_j(\tilde{\Xi}_{\tilde{\mathcal{M}}_h}) = \liminf_{h\to\infty} \kappa_j(\tilde{\Xi}_{\tilde{\mathcal{M}}_h}),
\end{aligned}$$

where the second last equality results from using an increasing rate of $\kappa_{h,1}$ such that $\kappa_{h,1} = o\left(\frac{1}{\sum_{1\leq i\leq 2K-1} \|Q_{n_h,i} - Q_i\|_{\max} + h^{-1}}\right)$.

To complete the proof, we now prove the above two properties.

1. Note that

$$\|\tilde{\mathcal{M}}_h x - \tilde{P}_h^{\frac{1}{2}} Q' x\| \leq \kappa_{h,1}^{\frac{1}{2}} \left(\sum_{1\leq i\leq 2K-1} \|Q_{n_h,i} - Q_i\|_{\max}\right) \|x\| = o(1)$$

because $\kappa_{h,1} = o\left(\frac{1}{\sum_{1\leq i\leq 2K-1} \|Q_{n_h,i} - Q_i\|_{\max} + h^{-1}}\right)$. Hence, for the sequence $\{\tilde{\mathcal{M}}_h, h \geq 1\}$, if, for example, for an arbitrary $i < K$ such that $Q_i' x \neq 0$, $\|x\| = 1$, then with probability approaching one as h increases,

$$\begin{aligned}
x' \tilde{\Xi}'_{\tilde{\mathcal{M}}_h} \tilde{\Xi}_{\tilde{\mathcal{M}}_h} x &= x' Z' Z x + x' \tilde{\mathcal{M}}_h' Z x + x' Z' \tilde{\mathcal{M}}_h x + x' \tilde{\mathcal{M}}_h' \tilde{\mathcal{M}}_h x \\
&\geq \tfrac{1}{2} \kappa_{h,K-1} x' Q_i' Q_i x \to \infty,
\end{aligned}$$

Combining the above equation with Lemma 2.(i), we know $\kappa_j(\tilde{\Xi}_{\mathcal{M}_h})$, the minimum in Eq. (B.48), should be achieved over linear spaces that are orthogonal to $Q_i, i < K$ as h increases. This completes the proof of the first property (B.48).
2. For $j \geq K$, by construction, $\kappa_j(\tilde{\Xi}_{\tilde{\mathcal{M}}_h}) = Q'_{n_h,i} \tilde{\Xi}'_{\tilde{\mathcal{M}}_h} \tilde{\Xi}_{\tilde{\mathcal{M}}_h} Q_{n_h,i} = Q'_{n_h,i} Z' Z Q_{n_h,i} \to Q_i' Z' Z Q_i$. Hence, $\liminf_{h\to\infty} \kappa_j(\tilde{\Xi}_{\tilde{\mathcal{M}}_h}) = \lim_{h\to\infty} \kappa_j(\tilde{\Xi}_{\tilde{\mathcal{M}}_h})$. This completes the proof of the second property (B.49). □

Proof of Lemma 2. (iii) The proof follows the proof of Theorem 8.4.(d) in the supplementary material of Andrews and Guggenberger (2017). Note that $n\hat{U}_n' \hat{D}_n' \widehat{W}_n' \widehat{W}_n \hat{D}_n \hat{U}_n$ in Andrews and Guggenberger (2017) corresponds with our $\hat{\Xi}' \hat{\Xi}$. Equations (15.10), (16.37) and Theorem 8.4.(d) in the supplementary material of Adrian et al. (2013) then lead to $W \sim \mathcal{W}_K\left(N - (K-1), I_K, 0_K\right)$ under any sequence of parameter matrices in $\mathbb{M}_\infty$. □

Appendix C. Proof of Theorem 4

Proof of Theorem 4. The smallest K roots are calculated from the polynomial

$$\left| \begin{pmatrix} I_K & 0 \\ -\lambda_1^{0\prime} & 0 \\ 0 & I_{K-1} \end{pmatrix}' \left[\mu \hat{\Psi} - \hat{\Phi}' \hat{\Sigma}^{-1} \hat{\Phi}\right] \begin{pmatrix} I_K & 0 \\ -\lambda_1^{0\prime} & 0 \\ 0 & I_{K-1} \end{pmatrix} \right| = 0,$$

where $\hat{\Psi} = \hat{\Psi}' = \begin{pmatrix} \hat{\Psi}_X & \hat{\Psi}_{XV} \\ \hat{\Psi}_{VX} & \hat{\Psi}_V \end{pmatrix}$ and $\hat{\Psi}_X$ and $\hat{\Psi}_V$ are $K \times K$ submatrices of $\hat{\Psi}$, We specify $\lambda_1^0 = c\bar{\lambda}_1^0$, with $c = (\lambda_1^{0\prime} \lambda_1^0)^{\frac{1}{2}}$, so $\bar{\lambda}_1^{0\prime} \bar{\lambda}_1^0 = 1$ and $\bar{\lambda}_{1,\perp}^0 : K \times (K-1)$, $\bar{\lambda}_{1,\perp}^{0\prime} \bar{\lambda}_1^0 \equiv 0$, $\bar{\lambda}_{1,\perp}^{0\prime} \bar{\lambda}_{1,\perp}^0 \equiv I_{K-1}$. Hence, $\text{diag}((\bar{\lambda}_{1,\perp}^0 \ \vdots \ \bar{\lambda}_1^0), I_{K-1})$ is an invertible orthonormal matrix. Pre- and post multiplying the matrices in the determinant with it does therefore not affect the characteristic roots of he following polynomial:

$$\left| \begin{pmatrix} \bar{\lambda}_{1,\perp}^0 & \bar{\lambda}_1^0 & 0 \\ 0 & 0 & I_{K-1} \end{pmatrix}' \begin{pmatrix} I_K & 0 \\ -\lambda_1^{0\prime} & 0 \\ 0 & I_{K-1} \end{pmatrix}' \left[\mu \hat{\Psi} - \hat{\Phi}' \hat{\Sigma}^{-1} \hat{\Phi}\right] \begin{pmatrix} I_K & 0 \\ -\lambda_1^{0\prime} & 0 \\ 0 & I_{K-1} \end{pmatrix} \begin{pmatrix} \bar{\lambda}_{1,\perp}^0 & \bar{\lambda}_1^0 & 0 \\ 0 & 0 & I_{K-1} \end{pmatrix} \right| = 0,$$

which can be rewritten as

$$\left|\begin{pmatrix} 1 & 0 & 0 \\ 0 & -c & 0 \\ 0 & 0 & I_{2(K-1)-1} \end{pmatrix}' \begin{pmatrix} \bar{\lambda}^0_{1,\perp} & -\bar{\lambda}^0_1/c & 0 \\ 0 & 1 & 0 \\ 0 & 0 & I_{K-1} \end{pmatrix}' \left[\mu\hat{\Psi} - \hat{\Phi}'\hat{\Sigma}^{-1}\hat{\Phi}\right] \begin{pmatrix} \bar{\lambda}^0_{1,\perp} & -\bar{\lambda}^0_1/c & 0 \\ 0 & 1 & 0 \\ 0 & 0 & I_{K-1} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -c & 0 \\ 0 & 0 & I_{2(K-1)-1} \end{pmatrix}\right| = 0.$$

Hence, when c goes to infinity, the characteristic polynomial becomes:

$$\left|\begin{pmatrix} \bar{\lambda}^0_{1,\perp} & 0 \\ 0 & I_K \end{pmatrix}' \left[\mu\hat{\Psi} - \hat{\Phi}'\hat{\Sigma}^{-1}\hat{\Phi}\right] \begin{pmatrix} \bar{\lambda}^0_{1,\perp} & 0 \\ 0 & I_K \end{pmatrix}\right| = 0.$$

Since $\hat{\Phi} = (\hat{d} \vdots \hat{\beta})$, we have $\hat{\Phi} \times \text{diag}(\bar{\lambda}^0_{1,\perp}, I_{K-1}) = (\hat{d}\bar{\lambda}^0_{1,\perp} \vdots \hat{\beta})$. The subset AR statistic now equals the sum of the K smallest root of the above characteristic polynomial which depend on $\bar{\lambda}^0_{1,\perp}$. For example, when $\bar{\lambda}_1 = e_{1,k}$, $\bar{\lambda}^0_{1,\perp} = \binom{0}{I_{k-1}}$ so $\hat{d}\bar{\lambda}^0_{1,\perp} = \hat{d}_2$ etc.

Let $\hat{\Psi}(\bar{\lambda}^0_{1,\perp}) = \begin{pmatrix} \bar{\lambda}^0_{1,\perp} & 0 \\ 0 & I_K \end{pmatrix}' \hat{\Psi} \begin{pmatrix} \bar{\lambda}^0_{1,\perp} & 0 \\ 0 & I_K \end{pmatrix}$. Then the above characteristic polynomial can be rewritten, by pre- and post multiplying the matrices in the determinant with $\hat{\Psi}(\bar{\lambda}^0_{1,\perp})^{-1/2\prime}$ and $\hat{\Psi}(\bar{\lambda}^0_{1,\perp})^{-1/2}$ respectively, as

$$\left|\mu I_{2K-1} - \hat{\Psi}(\bar{\lambda}^0_{1,\perp})^{-1/2\prime}(\hat{d}\bar{\lambda}^0_{1,\perp} \vdots \hat{\beta})'\hat{\Sigma}^{-1}(\hat{d}\bar{\lambda}^0_{1,\perp} \vdots \hat{\beta})\hat{\Psi}(\bar{\lambda}^0_{1,\perp})^{-1/2}\right| = 0.$$

The lower $K \times K$ principal submatrix of

$$\hat{\Psi}(\bar{\lambda}^0_{1,\perp})^{-1/2\prime}(\hat{d}\bar{\lambda}^0_{1,\perp} \vdots \hat{\beta})'\hat{\Sigma}^{-1}(\hat{d}\bar{\lambda}^0_{1,\perp} \vdots \hat{\beta})\hat{\Psi}(\bar{\lambda}^0_{1,\perp})^{-1/2}$$

by construction is $\hat{\Psi}_V^{-1/2\prime}\hat{\beta}'\hat{\Sigma}^{-1}\hat{\beta}\hat{\Psi}_V^{-1/2}$, and thus Cauchy's interlacing inequality implies the sum of the K smallest roots of the above polynomial is bounded from below by the minimum eigenvalue of $\hat{\Psi}_V^{-1/2\prime}\hat{\beta}'\hat{\Sigma}^{-1}\hat{\beta}\hat{\Psi}_V^{-1/2}$. Therefore, the limit sFAR statistic is bounded from below uniformly by the minimum eigenvalue of $\hat{\Psi}_V^{-1/2\prime}\hat{\beta}'\hat{\Sigma}^{-1}\hat{\beta}\hat{\Psi}_V^{-1/2}$. □

Appendix D. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2024.105728.

References

Adrian, T., Crump, R.K., Moench, E., 2013. Pricing the term structure with linear regressions. J. Financ. Econ. 110 (1), 110–138.

Anatolyev, S., Mikusheva, A., 2022. Factor models with many assets: strong factors, weak factors, and the two-pass procedure. J. Econometrics 229 (1), 103–126.

Andrews, D.W., Cheng, X., 2012. Estimation and inference with weak, semi-strong, and strong identification. Econometrica 80 (5), 2153–2211.

Andrews, D.W., Guggenberger, P., 2017. Asymptotic size of Kleibergen's LM and conditional LR tests for moment condition models. Econometric Theory 33 (5), 1046–1080.

Ang, A., Piazzesi, M., 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. J. Monetary Econ. 50 (4), 745–787.

Antoine, B., Lavergne, P., 2022. Identification-robust nonparametric inference in a linear IV model. J. Econometrics.

Antoine, B., Renault, E., 2009. Efficient GMM with nearly-weak instruments. Econom. J. 12, S135–S171.

Antoine, B., Renault, E., 2012. Efficient minimum distance estimation with multiple rates of convergence. J. Econometrics 170 (2), 350–367.

Antoine, B., Renault, E., 2020. Testing identification strength. J. Econometrics 218 (2), 271–293.

Barnichon, R., 2010. Building a composite help-wanted index. Econom. Lett. 109 (3), 175–178.

Beaulieu, M.-C., Dufour, J.-M., Khalaf, L., 2013. Identification-robust estimation and testing of the zero-beta CAPM. Rev. Econom. Stud. 80 (3), 892–924.

Beaulieu, M.-C., Gagnon, M.-H., Khalaf, L., 2016. Less is more: Testing financial integration using identification-robust asset pricing models. J. Int. Financial Mark., Inst. Money 45, 171–190.

Buraschi, A., Jiltsov, A., 2005. Inflation risk premia and the expectations hypothesis. J. Financ. Econ. 75 (2), 429–490.

Cattaneo, M.D., Crump, R.K., Wang, W., 2022. Beta-sorted portfolios. arXiv preprint arXiv:2208.10974.

Cochrane, J.H., Piazzesi, M., 2005. Bond risk premia. Amer. Econ. Rev. 95 (1), 138–160.

Cox, J.C., Ingersoll Jr., J.E., Ross, S.A., 1985. A theory of the term structure of interest rates. Econometrica 385–407.

Crump, R.K., Gospodinov, N., 2022. On the factor structure of bond returns. Econometrica 90 (1), 295–314.

Dovonon, P., Renault, E., 2013. Testing for common conditionally heteroskedastic factors. Econometrica 81 (6), 2561–2586.

Dufour, J.-M., 1997. Some impossibility theorems in econometrics with applications to structural and dynamic models. Econometrica 1365–1387.

Dufour, J.-M., Taamouti, M., 2005. Projection-based statistical inference in linear structural models with possibly weak instruments. Econometrica 73 (4), 1351–1365.

Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. J. Polit. Econ. 607–636.

Goliński, A., Zaffaroni, P., 2016. Long memory affine term structure models. J. Econometrics 191 (1), 33–56.

Gospodinov, N., Kan, R., Robotti, C., 2017. Spurious inference in reduced-rank asset-pricing models. Econometrica 85 (5), 1613–1628.

Guggenberger, P., Kleibergen, F., Mavroeidis, S., 2019. A more powerful subvector Anderson Rubin test in linear instrumental variables regression. Quant. Econ. 10 (2), 487–526.

Guggenberger, P., Kleibergen, F., Mavroeidis, S., 2022. A test for kronecker product structure covariance matrix. J. Econometrics.

Guggenberger, P., Kleibergen, F., Mavroeidis, S., Chen, L., 2012. On the asymptotic sizes of subset Anderson–Rubin and Lagrange multiplier tests in linear instrumental variables regression. Econometrica 80 (6), 2649–2666.

Gürkaynak, R.S., Sack, B., Wright, J.H., 2007. The US Treasury yield curve: 1961 to the present. J. Monetary Econ. 54 (8), 2291–2304.

Hamilton, J.D., Wu, J.C., 2012. Identification and estimation of Gaussian affine term structure models. J. Econometrics 168 (2), 315–331.

Joslin, S., Priebsch, M., Singleton, K.J., 2014. Risk premiums in dynamic term structure models with unspanned macro risks. J. Finance 69 (3), 1197–1233.

Kan, R., Zhang, C., 1999. Two-pass tests of asset pricing models with useless factors. J. Finance 54 (1), 203–235.

Khalaf, L., Schaller, H., 2016. Identification and inference in two-pass asset pricing models. J. Econom. Dynam. Control 70, 165–177.

Kleibergen, F., 2005. Testing parameters in GMM without assuming that they are identified. Econometrica 73 (4), 1103–1123.
Kleibergen, F., 2009. Tests of risk premia in linear factor models. J. Econometrics 149 (2), 149–173.
Kleibergen, F., 2021. Efficient size correct subset inference in homoskedastic linear instrumental variables regression. J. Econometrics 221 (1), 78–96.
Kleibergen, F., Kong, L., Zhan, Z., 2023. Identification robust testing of risk premia in finite samples. J. Financ. Econom. 21 (2), 263–297.
Kleibergen, F., Paap, R., 2006. Generalized reduced rank tests using the singular value decomposition. J. Econometrics 133 (1), 97–126.
Kleibergen, F., Zhan, Z., 2015. Unexplained factors and their effects on second pass R-squared's. J. Econometrics 189 (1), 101–116.
Kleibergen, F., Zhan, Z., 2020. Robust inference for consumption-based asset pricing. J. Finance 75 (1), 507–550.
Ludvigson, S.C., Ng, S., 2009. Macro factors in bond risk premia. Rev. Financ. Stud. 22 (12), 5027–5067.
Moreira, M.J., 2003. A conditional likelihood ratio test for structural models. Econometrica 71 (4), 1027–1048.
Muirhead, R.J., 2009. Aspects of Multivariate Statistical Theory. John Wiley & Sons.
Sameh, A.H., Wisniewski, J.A., 1982. A trace minimization algorithm for the generalized eigenvalue problem. SIAM J. Numer. Anal. 19 (6), 1243–1259.
Stock, J.H., Wright, J.H., 2000. GMM with weak identification. Econometrica 68 (5), 1055–1096.
Vasicek, O., 1977. An equilibrium characterization of the term structure. J. Financ. Econ. 5 (2), 177–188.
Wu, J.C., Xia, F.D., 2016. Measuring the macroeconomic impact of monetary policy at the zero lower bound. J. Money Credit Bank. 48 (2–3), 253–291.