Misspecification and Weak Identification in Asset Pricing

Frank Kleibergen* Zhaoguo Zhan†

January 15, 2022

Abstract

The widespread co-existence of misspecification and weak identification in asset pricing has led to an overstated performance of risk factors. Because the conventional Fama and MacBeth (1973) methodology is jeopardized by misspecification and weak identification, we infer risk premia by using a double robust Lagrange multiplier test that remains reliable in the presence of these two empirically relevant issues. Moreover, we show how the identification, and the resulting appropriate interpretation, of the risk premia is governed by the relative magnitudes of the misspecification J-statistic and the identification IS-statistic. We revisit several prominent empirical applications and all specifications with one to six factors from the factor zoo of Feng, Giglio, and Xiu (2020) to emphasize the widespread occurrence of misspecification and weak identification.

JEL Classification: G12

Keywords: misspecification, weak identification, risk factor, asset pricing

*Corresponding author. Email: f.r.kleibergen@uva.nl. Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands.

†Email: zzhan@kennesaw.edu. Department of Economics, Finance and Quantitative Analysis, Coles College of Business, Kennesaw State University, GA 30144, USA.

1 Introduction

Over the past decades, hundreds of risk factors have been proposed to explain the cross-section of asset returns, making up the, so-called, zoo of factors; see, e.g., Cochrane (2011) and Harvey, Liu, and Zhu (2016). Much of the empirical support for these factors is based on the beta representation, where risk premia are identified by projecting expected returns on the betas, and the betas are the factor loadings in the time-series regression of asset returns on risk factors; see, e.g., Fama and MacBeth (FM, 1973). The resulting FM t-statistic on risk premia, Hansen (1982)'s J-statistic for misspecification and the cross-sectional R^2, are widely used in the asset pricing literature to show support for proposed risk factors. The reliability of these conventional statistics has, however, recently been brought into question by two empirically relevant issues: misspecification and weak identification.

Up till now, misspecification has been widely acknowledged as an inherent feature of asset pricing models; see, e.g., Gospodinov, Kan, and Robotti (2014). For the beta representation, misspecification results in expected returns that are not fully explained by the betas of the specified factors. Thus, pricing errors generally exist and should be taken into account when conducting inference on risk premia. Kan, Robotti, and Shanken (KRS, 2013) therefore propose the KRS t-test, which explicitly allows for the existence of misspecification while the FM t-test does not. As warned by Kan, Robotti, and Shanken (2013), failure to account for misspecification tends to enlarge the t-statistic on risk premia, leading to overstated pricing performance of risk factors.

Weak (or no) identification, on the other hand, is driven by poor quality risk factors or more generally, limited information contained in the data. Kan and Zhang (1999), for example, warn that risk factors proposed in the literature could be useless factors with zero betas. When betas are zero, risk premia are unidentified in the beta representation and paired with misspecification, the FM t-statistic can be spuriously large to support useless factors as shown by Kan and Zhang (1999). In addition, Kleibergen (2009) illustrates the malfunction of the FM t-statistic caused by factors that are only weakly correlated with asset

returns and proposes robust tests to address weak identification of risk premia. Moreover, Kleibergen and Zhan (2015, 2021) show that the cross-sectional R^2 and J-statistic may also spuriously favor statistically weak factors, respectively.

Despite that misspecification and weak identification are well recognized, there is currently no unified approach for identification and inference that can deal with these empirically relevant issues. We therefore advocate such an approach and apply it to show the widespread occurrence of misspecification and weak identification in applied asset pricing studies.

When misspecification is present, there is no longer a (true) value of the risk premia at which the pricing error is zero. Different estimation procedures then lead to distinct pseudo-true values of risk premia, which are the minimizers of the population objective function of the respective estimation procedure. We show that, in case of generic misspecification, the difference between the pseudo-true value of two estimation procedures, i.e. the FM two-pass approach and the continuous updating estimator (CUE) of Hansen, Heaton, and Yaron (1996), and the baseline risk premia that would apply in case of correct specification, depends on the strength of misspecification compared to the strength of identification. The strength of misspecification is reflected by the population equivalent of the J-statistic, while the strength of identification is revealed by the, so-called, identification strength (IS) statistic. The IS-statistic is a rank test statistic on the β matrix and is, by construction, always larger than or equal to the J-statistic. The difference between these two measures is indicative of whether we can interpret the pseudo-true value as a risk premium. When this difference is small, the pseudo-true value basically results from the close to reduced rank value of the β matrix, so it can be far off from the baseline risk premium under correct specification, and does not reflect a risk premium. The sample equivalents of the J and IS statistics are therefore important in gauging whether we can interpret risk premia estimates resulting from empirical studies as genuine risk premia of interest.

For eight prominent empirical studies in asset pricing: Fama and French (1993), Jagannathan and Wang (1996), Yogo (2006), Lettau and Ludvigson (2001), Savov (2011), Adrian,

Etula, and Muir (2014), Kroencke (2017) and He, Kelly, and Manela (2017) and for all (14 billion+) one to six factor specifications resulting from the factor zoo of Feng, Giglio, and Xiu (2020), we compute the J and IS statistics to illustrate the widespread co-existence of misspecification and weak identification. The resulting J-statistics are mostly large and significant, and if they are not, their smaller values are often induced by an accompanying small value of the IS-statistic. The smallish J-statistic then basically results from a close to reduced rank value of the β matrix, so the risk premia estimates do so as well and do not reflect risk premia. When we do not incorporate the zero-β return or use more time series observations, these identification issues become less problematic for some settings, such as the three-factor model from Fama and French (1993), but remain for many others.

The conventional test statistics for conducting inference on risk premia, such as the FM two-pass t-test, also with the Shanken (1992) correction of the standard errors, and the KRS t-test, become unreliable in the presence of misspecification and/or weak identification. We illustrate this by showing that the limit behavior of the FM two-pass estimator consists of four components. Two of these components are negligible in case of strong factors and correct specification, but lead to non-standard behavior in case of misspecification and/or weak identification. It is also not possible to correct for them by using the bootstrap. We therefore use the double robust Lagrange multiplier (DRLM) statistic from Kleibergen and Zhan (2021) that provenly remains reliable in case of misspecification and/or weak identification. For ease of exposition, we conduct a small simulation experiment to illustrate the pros and cons of these different statistics for testing hypotheses on risk premia.

We illustrate the practical usage of the DRLM test by using the influential Fama and French (1993) three-factor model and the conditional consumption capital asset pricing model in Lettau and Ludvigson (2001). For the Fama and French (1993) three-factor model, we show that for commonly used data sets, such as those resulting from Lettau and Ludvigson (2001) and Lettau, Ludvigson, and Ma (2019), identification of the risk premia can be problematic when incorporating the zero-β return because of the near constancy of the

β's associated with the market return. This identification issue is alleviated when using more time series observations or by removing the zero-β return. The joint 95% confidence set for the risk premia resulting from the DRLM test is then bounded (ellipsoid), while it is unbounded in the direction of the risk premia on the market return when using fewer observations and also incorporating the zero-β return. For the conditional capital asset pricing model stemming from Lettau and Ludvigson (2001), the 95% confidence sets of the risk premia resulting from the DRLM test are all unbounded, indicating limited information in the data for precisely identifying the risk premia. The 95% confidence sets from the DRLM test are, for all of these settings, fully in line with the J and IS statistics. When the J and IS statistics are relatively close, we obtain unbounded 95% confidence sets from the DRLM test, in contrast with the bounded ones resulting from the FM two-pass and KRS t-tests which are then unreliable. Also the CUE differs considerably from the FM two-pass estimator in this scenario. When the J and IS statistics are substantially apart, the results from all these procedures are, as expected, rather similar, but the 95% (projected) confidence sets from the DRLM test are notably narrower than those from the FM two-pass and KRS t-tests. It all shows the importance of jointly using the J and IS statistics to gauge the identification of risk premia while using the DRLM test to conduct inference on risk premia.

Overall, our study adds to an emerging body of research that aims to bring discipline to the zoo of factors. Harvey, Liu, and Zhu (2016), for example, propose a higher hurdle such as a t-statistic greater than 3.0 instead of the commonly used 1.96, since they are concerned that significant t-statistics documented in the existing literature could result from data mining. Unlike Harvey, Liu, and Zhu (2016) and the follow-up work, we do not provide new hurdles of t-statistics; instead, we suggest the DRLM, J and IS statistics for gauging the quality of risk factors. Feng, Giglio, and Xiu (2020) focus on whether a proposed risk factor adds explanatory power beyond the existing zoo of factors, so their methodology builds on a high-dimensional set of existing factors. In contrast, our approach is suitable for evaluating the explanatory power of each factor model individually. In order to jointly

address misspecification and no identification, Gospodinov, Kan, and Robotti (2014) propose a model selection procedure to eliminate potentially useless factors. Unlike Gospodinov, Kan, and Robotti (2014), our proposed DRLM test aims to infer risk premia regardless of whether misspecification and weak identification are present.

Finally, we note that misspecification and weak identification are not limited to the beta representation; see, e.g., Stock and Wright (2000) and Hansen and Lee (2021), who have studied weak identification and misspecification in the generalized method of moments framework, respectively. Given that asset pricing models are at best approximations of reality while lots of risk factors have little explanatory power for asset returns, misspecification and weak identification are a likely common threat to empirical asset pricing studies.

The rest of the paper is organized as follows. Section 2 starts with discussing the setup of misspecification and the consequences it has for commonly used risk premia estimators. It illustrates that the commonly used FM t-test is jeopardized by both misspecification and weak identification, while the DRLM test takes both issues into account. Section 3 uses the J and IS statistics to highlight the prevalence of misspecification and weak identification in existing studies. Section 4 contains the empirical applications for conducting inference on risk premia, while Section 5 concludes. Technical details are relegated to the Appendix.

2 Misspecification in the beta representation

Let $R_{i,t}$ be the return on the i-th asset at time t, with $i = 1, ..., N$, and $t = 1, ..., T$. The beta representation of expected returns models it as linear in the beta vector of factor loadings:

$$E(R_{i,t}) = \beta_i' \lambda_F, \tag{1}$$

where λ_F is the $K \times 1$ vector of risk premia, and β_i is the $K \times 1$ vector of factor loadings:

$$\beta_i = var(F_t)^{-1} cov(F_t, R_{i,t}), \tag{2}$$

with F_t the $K \times 1$ vector of the specified risk factors, and $K < N + 1$. We can as well represent the beta representation jointly for all assets by stacking the N equations of (1) to get

$$\mu_R = E(R_t) = \beta\lambda_F, \tag{3}$$

with $R_t = (R_{1,t}, ..., R_{N,t})'$, $\beta = (\beta_1, ..., \beta_N)'$.

In both (1) and its misspecification extension provided later on (i.e., Equation (6)), identification of λ_F relies on the quality of β_i. For instance, if some of the specified risk factors are just useless noise with zero betas, then λ_F is unidentified in the beta representation. This reasoning generalizes to the full rank condition of the $N \times K$-dimensional matrix β, i.e. for λ_F to be identified, β needs to have full rank. In the single factor case with $K = 1$, this rank condition then just requires that β should be non-zero. If the full rank condition of β is jeopardized by risk factors of poor quality, then λ_F is potentially weakly identified; see, e.g., Kan and Zhang (1999), Kleibergen (2009), and Kan, Robotti, and Shanken (2013) for a related discussion.

Remark 1: The zero-β, $\lambda_0 = 0$, restriction. A scalar λ_0 is often added to (3) so

$$E(R_t) = \iota_N\lambda_0 + \beta\lambda_F, \tag{4}$$

where ι_N is the $N \times 1$ vector of ones, and λ_0 is the, so-called, zero-beta return, or the expected return to an asset with no exposure to priced risks. Since our interest mainly lies in λ_F, we opt to focus on (3) instead of (4) for ease of exposition. This treatment is related to the $\lambda_0 = 0$ restriction, so that (4) reduces to (3). The zero restriction can be achieved by considering R_t as the excess return. Our discussion on (3), however, can be straightforwardly extended to incorporate λ_0. For example, the full rank condition of β for (3) extends to the full rank condition of $(\iota_N \vdots \beta)$ for (4) once λ_0 is allowed for.

Remark 2: Incorporating the zero-β return. If $\lambda_0 = 0$ is not assumed, we can still use (3) to focus on λ_F by considering R_t as the return in deviation of a reference asset. To

illustrate, let $\mathcal{R}_t = (\mathcal{R}_{1,t} \dots \mathcal{R}_{N+1,t})'$ be the $(N+1) \times 1$ vector of returns such that

$$E(\mathcal{R}_t) = \iota_{N+1}\lambda_0 + \mathcal{B}\lambda_F, \tag{5}$$

where $\mathcal{B}$ is the $(N+1) \times K$-dimensional matrix of factor loadings. By subtracting the $(N+1)$-th asset return, we obtain the $N \times 1$ vector $R_t : R_t = (\mathcal{R}_{1,t} \dots \mathcal{R}_{N,t})' - \iota_N \mathcal{R}_{N+1,t}$ such that (5) implies (3), with $R_t = J_N \mathcal{R}_t$, $\beta = J_N \mathcal{B}$, and $J_N = (I_N \vdots -\iota_N)$. Thus, the full rank condition of $(\iota_{N+1} \vdots \mathcal{B})$ in (5) is equivalent to the full rank condition of β in (3). Our subsequent analysis is invariant with respect to the choice of the $(N+1)$-th asset; see Kleibergen and Zhan (2020).

Remarks 1 and 2 above show that we can focus on (3) to illustrate inference on λ_F regardless of whether a zero-β return is incorporated or not.[1]

2.1 Misspecification

Under misspecification, the expected returns are not fully explained by the specified β, so $\mu_R \neq \beta\lambda_F$. We therefore consider the correctly specified setting above as the baseline, and introduce the pricing error $\tilde{e}$ in (6) to reflect misspecification. In particular, we assume the pricing error to be potentially of a different order of magnitude:

$$\begin{aligned} \mu_R &= \beta\lambda_F + \tilde{e} \\ \tilde{e} &= O(\tilde{e}) \cdot a \\ \beta &= O(\beta) \cdot b, \end{aligned} \tag{6}$$

[1] It is worth noting that whether λ_0 should be excluded already sheds light on misspecification and weak identification in asset pricing. On the one hand, if λ_0 is non-zero, then incorrectly imposing the zero restriction leads to a misspecified condition. On the other hand, if λ_0 is zero, then including the redundant intercept term in (4) potentially weakens the identification of risk premia. In particular, when there is little cross-sectional variation in β, including ι_N causes near-multicollinearity in $(\iota_N \vdots \beta)$, which will further induce the weak identification of λ_F. Misspecification and weak identification are, however, more general than whether to impose $\lambda_0 = 0$ or not.

where a is a normalized N-dimensional vector, so $O(\tilde{e})$ reflects the magnitude of misspecification; similarly, b is a normalized $N \times K$-dimensional matrix, so $O(\beta)$ reflects the magnitude of identification.

The specification in (6) is without loss of generality and, for example, allows for misspecification which is much smaller than the expected return resulting from the beta representation, or proportional to it which would result in case of weak identification. To capture this, the magnitudes can be modelled to be proportional to the number of time series observations as is common in the literature on weak identification, for example, $O(\tilde{e}) = O(T^{c_{\tilde{e}}})$, $O(\beta) = O(T^{c_\beta})$, $c_{\tilde{e}}$ and c_β are finite constants; see, e.g., Staiger and Stock (1997) and Kleibergen (2009).

The misspecification is generic, so the difference from the baseline pricing error, $\tilde{e}$, lies both in the space spanned by β and outside of it. It captures the notion of the correctly specified setting as a baseline, on top of which there is an unstructured misspecification component. It resembles the estimation errors which the sample moment equations add to the population moment equations. These errors are also unstructured, and lie both in the space spanned by β and outside of it. The magnitude components, $O(\tilde{e})$ and $O(\beta)$, can then further be such that the estimation errors in the sample moment equations are of the same order of magnitude as the misspecification and identification strengths.

Next, we show that the relative magnitudes of misspecification and identification, $O(\tilde{e})$ and $O(\beta)$, drive the identification of risk premia in the FM two-pass methodology and the CUE. Since the misspecification J-statistic gauges $O(\tilde{e})$ while the identification IS-statistic reflects $O(\beta)$, the comparison of these two statistics plays a crucial role in our later discussion.

2.2 Population objective function and pseudo-true value

2.2.1 FM two-pass estimator

The population objective functions of different risk premia estimators are all quadratic forms of the pricing error but weigh it differently. The population objective function of the FM

two-pass estimator involves no weight function and is therefore just the quadratic form of the pricing error:

$$Q_{FM}(l) = (\mu_R - \beta l)'(\mu_R - \beta l), \tag{7}$$

having as its minimizer the, so-called, pseudo-true value of risk premia. Unlike the true value in case of correct specification, the pseudo-true value denoted by $\lambda^*_{F,FM}$ does not set the objective function to zero as $\mu_R = \beta\lambda_F + \tilde{e}$, with $\tilde{e} \neq 0$:

$$\lambda^*_{F,FM} = \arg\min_{l\in\mathbb{R}^K} \; Q_{FM}(l) = (\beta'\beta)^{-1}\beta'\mu_R. \tag{8}$$

Using the specification of the expected asset returns in (6), the pseudo-true value becomes:

$$\lambda^*_{F,FM} = \lambda_F + \tfrac{O(\tilde{e})}{O(\beta)} \cdot (b'b)^{-1}b'a, \tag{9}$$

where $\frac{O(\tilde{e})}{O(\beta)} \cdot (b'b)^{-1}b'a$ reflects the difference between the pseudo-true value $\lambda^*_{F,FM}$ and the generic risk premia λ_F.

The specification of the pseudo-true value (9) therefore shows that it depends on the strength of misspecification, $O(\tilde{e})$, compared to the identification strength, $O(\beta)$, unless $(b'b)^{-1}b'a = 0$ so the misspecification error is outside of the space spanned by β. For example, when we use the specification from the weak instrument/factor literature: $O(\tilde{e}) = O(T^{c_{\tilde{e}}}) = O_{c_{\tilde{e}}} \times T^{c_{\tilde{e}}}$, $O(\beta) = O(T^{c_\beta}) = O_{c_\beta} \times T^{c_\beta}$, with $c_{\tilde{e}} = c_\beta = -\frac{1}{2}$ and $O_{c_{\tilde{e}}}$, O_{c_β} non-zero finite constants:

$$\lambda^*_{F,FM} = \lambda_F + \tfrac{O_{c_{\tilde{e}}}}{O_{c_\beta}} \cdot (b'b)^{-1}b'a, \tag{10}$$

so depending on $O_{c_{\tilde{e}}}$ being larger or smaller than O_{c_β}, there can be a considerable difference between $\lambda^*_{F,FM}$ and λ_F. Hence, it is important to compare the strength of misspecification reflected by $O(\tilde{e})$ to the strength of identification reflected by $O(\beta)$, in order to gauge the difference between $\lambda^*_{F,FM}$ and λ_F.

Under correct specification, the pricing error is zero at the true value of risk premia,

and remains so when repackaging the assets. The minimizers of the population objective functions underlying different estimators all therefore have the true value of risk premia as their minimizer, and are all invariant to repackaging. Under the standard conditions, these estimators are also consistent for the true value. This is no longer the case in misspecified settings where the minimizers of the population objective functions associated with various estimators (*i.e.* the pseudo-true values) differ, since there is no longer a value of risk premia at which the population objective function is equal to zero.

Kandel and Stambaugh (1995) show that the pseudo-true value for the FM two-pass estimator is not invariant to repackaging of the assets, while the population cross-section generalized least squares (GLS) estimator is. The FM two-pass estimator provides a consistent estimator of the pseudo-true value, which makes Kandel and Stambaugh (1995) skeptical about how much of interest the cross-section ordinary least squares (OLS) estimator, or FM two-pass estimator, is under misspecification. They thus show a preference for the cross-section GLS estimator.

Theorem 1. For A an invertible $N \times N$ matrix of weights, $A\iota_N = \iota_N$ with ι_N the N-dimensional vector of ones, the FM two-pass pseudo-true value for the orginal test assets, R_t, does not equal that of the repackaged test assets, $A \times R_t$.

Proof: See the Appendix and Kandel and Stambaugh (1995).

2.2.2 Continuous updating estimator

Kandel and Stambaugh (1995)'s analysis primarily concerns the population cross-section OLS and GLS objective functions. These population objective functions treat the expected asset returns, μ_R, and β's as known while we replace them by estimators in the sample objective functions. When extending these population objective functions to sample ones, an important and empirically relevant issue occurs if the proximity of the true β's to a reduced rank value is comparable to its estimation error. The resulting risk premia estimators are

then exposed to multi-collinearity issues and test statistics, like, for example, the FM two-pass t-test and its misspecification robust extension by Kan, Robotti, and Shanken (2013), are no longer reliable for conducting tests on the pseudo-true value of the FM two-pass estimator. Test statistics which remain reliable for conducting tests on the pseudo-true value do, however, exist and are based on the continuous updating estimator (CUE) of Hansen, Heaton, and Yaron (1996). The population objective function of the CUE provides an extension of the GLS population objective function by accounting for the estimation error of all estimable components of the pricing error, *i.e.* μ_R and β :

$$\begin{aligned} Q_{CUE}(l) = & \ (\mu_R - \beta l)' \left[\text{Var}(\sqrt{T}(\hat{\mu}_R - \hat{\beta} l))\right]^{-1} (\mu_R - \beta l) \\ = & \ \tfrac{1}{1+l' Q_{FF}^{-1} l} (\mu_R - \beta l)' \Omega^{-1} (\mu_R - \beta l), \end{aligned} \tag{11}$$

where $\hat{\mu}_R = \bar{R} = \frac{1}{T}\sum_{t=1}^{T} R_t$, $\hat{\beta} = \frac{1}{T}\sum_{t=1}^{T} \bar{R}_t \bar{F}_t' \left(\frac{1}{T}\sum_{j=1}^{T} \bar{F}_j \bar{F}_j'\right)^{-1}$, $\bar{R}_t = R_t - \bar{R}$, $\bar{F}_t = F_t - \bar{F}$, $\bar{F} = \frac{1}{T}\sum_{t=1}^{T} F_t$. The specification in the last line of (11) is for a setting of i.i.d. data, so $\Omega = \text{Var}(R_t - \beta F_t)$ and $Q_{FF} = \text{Var}(F_t)$. The GLS population objective function results if we remove the first part of the expression on the bottom line of (11).

The CUE population objective function normalizes the pricing error, so it is invariant under repackaging and transformations of the factors. The (normalized) sample value of the pricing error used in (11) has unit variance, which implies that, under mild conditions (see, e.g., Shanken (1992)), the sample CUE objective function has a non-central χ^2 distribution in large samples. The pseudo-true value associated with the CUE then results as

$$\lambda^*_{F,CUE} = \arg\min_{l \in \mathbb{R}^K} Q_{CUE}(l), \tag{12}$$

which is also invariant to repackaging.

The population CUE objective function is multi-modal, and the pseudo-true value results from the smallest mode. For example, in an i.i.d. setting, it results from an eigenvalue

problem so the number of modes equals the number of eigenvalues/characteristic roots.[2] To guarantee that the characteristic root from which the pseudo-true value results, identifies risk premia, the identification strength has to exceed the misspecification strength. These identification properties result because the CUE population objective function is the optimized function in the stepwise optimization of a generalized reduced rank objective function, which imposes a reduced rank value on the $N \times (K+1)$ matrix $\left(\mu_R \vdots \beta\right)$, see Kleibergen (2007) and Kleibergen and Zhan (2021):

$$\begin{aligned} Q_{CUE}(l) = & \ \min_{B \in \mathbb{R}^{N \times K}} Q_{CUE}(l, B) \\ Q_{CUE}(l, B) = & \begin{bmatrix} \text{vec}\left(\left(\mu_R \vdots \beta\right) - B\left(l \vdots I_K\right)\right) \\ \text{vec}\left(\left(\mu_R \vdots \beta\right) - B\left(l \vdots I_K\right)\right) \end{bmatrix}' \left[\text{Var}\left(\sqrt{T}\left(\hat{\mu}' \vdots \text{vec}(\hat{\beta})'\right)'\right)\right]^{-1} \\ = & \ \text{tr}\left[Q_{FF}^{-1}\left(\left(\mu_R \vdots \beta\right) - B\left(l \vdots I_K\right)\right)' \Omega^{-1}\left(\left(\mu_R \vdots \beta\right) - B\left(l \vdots I_K\right)\right)\right] \end{aligned} \tag{13}$$

where the expression on the last line is for the setting of i.i.d. data.[3] The second expression in (13), $Q_{CUE}(l, B)$, is a normalized distance measure between the $N \times (K+1)$ matrix $\left(\mu_R \vdots \beta\right)$, which is at most of rank $K+1$, and the $N \times (K+1)$ matrix $B\left(l \vdots I_K\right)$, which is at most of rank K.[4] The sample analog of the CUE population objective function is therefore a rank test on $\left(\mu_R \vdots \beta\right)$. Its minimal value is the J-statistic for misspecification.

Since the pseudo-true value of the CUE results from a generic rank test on $\left(\mu_R \vdots \beta\right)$, it does not necessarily represent risk premia when there is misspecification. Consider, for example, a setting where the strength of misspecification, $O(\tilde{e})$, is considerably larger than

[2] For the setting of i.i.d. data, a closed-form expression of the pseudo-true value can be provided as resulting from the eigenvector associated with the smallest root of the characteristic polynomial $\left|\tau \begin{pmatrix} 1 & 0 \\ 0 & Q_{FF}^{-1} \end{pmatrix} - \begin{pmatrix} \mu_R & \beta \end{pmatrix}' \Omega^{-1} \begin{pmatrix} \mu_R & \beta \end{pmatrix}\right| = 0$ by specifying that eigenvector as $c\begin{pmatrix} 1 \\ -\lambda^*_{F,CUE} \end{pmatrix}$ with c a scalar.

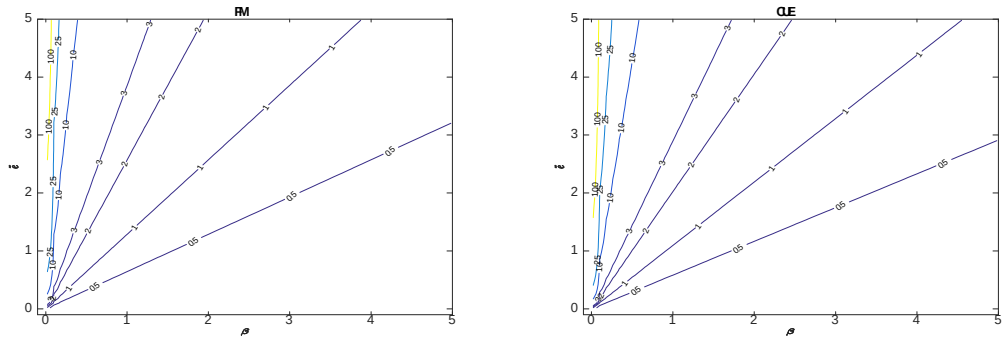
[3] $\text{vec}(A)$ is the column vectorization of a matrix A that results from stacking its columns, so $\text{vec}(A) = (a_1' \dots a_m')'$ for $A = (a_1 \dots a_m)$. $\text{tr}(A)$ is the trace, or sum of its diagonal elements, of the square matrix A.

[4] The rank of $B\left(l \vdots I_K\right)$ is at most K because B is an $N \times K$ matrix and $\left(l \vdots I_K\right)$ a $K \times (K+1)$ matrix, so the ranks of both of these are at most K.

the strength of identification, $O(\beta)$, which is tiny. The minimization over (l, B) then leads to a tiny value of B and a very large value of l that does not reflect the generic risk premia. This reasoning is thus also in line with (9) for the FM two-pass approach.

The above point is further illustrated in Figure 1. It shows, for a single factor setting, the contour lines of the pseudo-true values of the FM two-pass estimator and the CUE in deviation from the risk premium in case of correct specification, as a function of the misspecification and identification strengths. Both for FM two-pass (left panel) and CUE (right panel), Figure 1 shows that the pseudo-true values deviate considerably from the baseline risk premium when the misspecification strength exceeds the identification strength. The pseudo-true value of CUE then no longer represents a risk premium, because the closest proximity of $\left(\mu_R \vdots \beta\right)$ from a reduced rank value results mainly from the small value of β, and much less so from the combination of μ_R and β. It is therefore important to be able to diagnose if the pseudo-true value results from such a setting.

Figure 1: Contour lines that show the deviation of pseudo-true values of FM and CUE from the baseline risk premium in case of correct specification as a function of the strengths of misspecification $\tilde{e}$ and identification β.



Notes: The pseudo-true values of FM and CUE are defined in (8) and (12), respectively. The baseline risk premium is set to 2, so the contour lines show the deviation of pseudo-true values from 2. The parameters used for Figure 1 are calibrated to the data from Kroencke (2017). The magnitudes of $\tilde{e}$ and β vary from 0 to 5.

2.3 IS-statistic versus J-statistic

To obtain a diagnostic for interpreting the pseudo-true value as a risk premium, we use that the CUE population objective function evaluated at the pseudo-true value equals a rank test on $(\mu_R \vdots \beta)$ as shown by (13), which is also the population equivalent of a J-statistic:

$$J = Q_{CUE}(\lambda^*_{F,CUE}). \tag{14}$$

It is then always smaller than or equal to an appropriately specified rank test statistic conducted on any sub-matrix of $\left(\mu_R \vdots \beta\right)$. An example of a rank test on a sub-matrix of $(\mu_R \vdots \beta)$ is a rank test on just β, whose sample analog is typically used to test for the identification of the risk premia under correct specification; see, e.g., Cragg and Donald (1997), Kleibergen and Paap (2006) and Robin and Smith (2000). This rank test statistic is thus always larger than or equal to the CUE objective function evaluated at the pseudo-true value (i.e. the J-statistic in (14)), and provides a measure of the identification strength of λ_F, see Kleibergen and Zhan (2021):

$$\begin{aligned}
IS = &\ \min_{d\in\mathbb{R}^{(K-1)}} Q_\beta(d) \\
Q_\beta(d) = &\ \tbinom{1}{-d}'\beta' \left[\left(\tbinom{1}{-d}\otimes I_N\right)' \mathrm{Var}\left(\sqrt{T}\mathrm{vec}(\hat{\beta})\right)\left(\tbinom{1}{-d}\otimes I_N\right)\right]^{-1}\beta\tbinom{1}{-d} \\
= &\ \min_{G\in\mathbb{R}^{N\times(K-1)}} Q_{p,r}(d,G) \\
Q_{p,r}(d,G) = &\ \left[\mathrm{vec}\left(\beta - G\left(d \vdots I_{K-1}\right)\right)\right]' \left[\mathrm{Var}\left(\sqrt{T}\mathrm{vec}(\hat{\beta})\right)\right]^{-1}\left[\mathrm{vec}\left(\beta - G\left(d \vdots I_{K-1}\right)\right)\right].
\end{aligned} \tag{15}$$

When the J misspecification measure is close to the IS identification measure, the pseudo-true value basically results from a close to reduced rank value of β. It implies that the pseudo-true value $\lambda^*_{F,CUE}$ is very large and does not represent risk premia. Hence, only when the J misspecification measure is considerably less than the IS identification strength measure does the pseudo-true value represent risk premia. It shows that when misspecification is present, the identification of risk premia is no longer just reflected by the rank strength value of β, as

is the case under correct specification, but by the difference between a measure of the degree of misspecification and a measure of the rank strength of β. In case of correct specification, the misspecification measure equals zero, so the identification then solely results from the rank value of β, but not so when misspecification is present. The cut-off for identification is when the measures of misspecification and rank strength of β are equal in which case λ_F is not identified, while it can be identified as risk premia when the latter exceeds the former.

To summarize, the sample analog of the IS identification measure is a rank test statistic on β, while the sample analog of the minimal value of the population CUE objective function is the J-statistic for misspecification. It is thus important to compare the IS-statistic on β with the J-statistic for misspecification. When using aligning specifications for the J-statistic and the IS-statistic on β, the former is always less than or equal to the latter. Close values, however, indicate an issue with identifying risk premia. The estimated values of the risk premia are then also typically very large, which sheds further doubt on whether they can be interpreted as risk premia.[5]

2.4 Tests of risk premia

To conduct inference on the pseudo-true values of risk premia, it is important to have tests that remain reliable for a wide range of values of the misspecification and identification strengths. We show that this does not hold for the FM t-test and its misspecification robust extension by Kan, Robotti, and Shanken (2013), which are size distorted when the identification strength of risk premia is minor. We conduct a small simulation exercise to highlight this sensitivity. Thereafter we state the double robust Lagrange multiplier (DRLM) test from Kleibergen and Zhan (2021), which is size correct for all settings of the misspecification and identification strengths.

[5]It is important to note that we cannot test for the equality of the J-test and rank test for β, since they involve the same estimators. For example, in case β is of reduced rank while the expected returns, μ_R, are different from zero, the population values of these two test statistics would be the same, and the joint sampling distribution of their estimators would be degenerate, so we cannot establish a test for their equality.

2.4.1 Large sample behavior of t-tests based on the FM estimator under misspecification and weak identification

We use a single factor setting, so $K = 1$, to show that the FM two-pass t-test and its misspecification robust extension by Kan, Robotti, and Shanken (2013), for conducting inference on the FM two-pass pseudo-true value become unreliable, when the value of the β-matrix is close to rank deficient. For the single factor setting, a close to rank deficient value of β implies that β is close to zero. Because of the commonality of misspecification paired with weak identification, we construct the large sample behavior of the FM two-pass estimator for a setting where both the β's and the misspecification are small. We model this using the weak factor/small β and misspecification assumption (6) and (10), and further postulate that both β and the misspecification $\mu_R - \beta\lambda_F$ are drifting to zero at rate $1/\sqrt{T}$:

$$\beta = \beta_T = \tfrac{b}{\sqrt{T}},\ \mu_R - \beta\lambda_F = \tfrac{a}{\sqrt{T}},\ \lambda^*_{F,FM} = \lambda_F + (b'b)^{-1}b'a, \tag{16}$$

with b and a N-dimensional vectors of constants. The small β assumption (16) is common in the weak identification literature; see, e.g., Staiger and Stock (1997), Kleibergen (2005, 2009) and Kleibergen and Zhan (2020, 2021). It leads to the small values of the F-statistic testing the joint significance of the β's that we often observe; see Kleibergen and Zhan (2015). The small misspecification assumption further accommodates the small but significant values of the J-statistic that are regularly seen.

Theorem 2: For i.i.d. data and under the weak β and misspecification assumption (16), the large sample distribution of the FM two-pass estimator, $\hat{\lambda}_F = (\hat{\beta}'\hat{\beta})^{-1}\hat{\beta}'\hat{\mu}_R$, consists of four components:

$$\hat{\lambda}_F \underset{d}{\to} \underbrace{\lambda^*_{F,FM}}_{1} + \underbrace{\frac{\psi'_\mu(b+\psi_\beta)}{(b+\psi_\beta)'(b+\psi_\beta)}}_{2} - \underbrace{\lambda^*_{F,FM}\frac{\psi'_\beta(b+\psi_\beta)}{(b+\psi_\beta)'(b+\psi_\beta)}}_{3} + \underbrace{\frac{e'\psi_\beta}{(b+\psi_\beta)'(b+\psi_\beta)}}_{4}, \tag{17}$$

with $e = a - b(b'b)^{-1}b'a$; ψ_μ and ψ_β are independent normally distributed N-dimensional random vectors with mean zero and covariance matrices $\text{var}(R_t) = \Omega + \beta Q \beta'$ and ΩQ^{-1}, $\Omega = \text{Var}(R_t - \beta F_t)$ and $Q_{FF} = \text{Var}(F_t)$.

Proof: See the Appendix.

The four different components of the large sample behavior of the FM two-pass estimator in Theorem 2 are characterized by:

1. The object of interest: $\lambda^*_{F,FM}$, the pseudo-true value of the FM two-pass estimator.

2. $\frac{\psi'_\mu(b+\psi_\beta)}{(b+\psi_\beta)'(b+\psi_\beta)}$: Under i.i.d. data and assumption (16), $\sqrt{T}\hat{\beta} \underset{d}{\to} b + \psi_\beta$ and $\sqrt{T}\hat{\mu}_R \underset{d}{\to} e + b\lambda^*_{F,FM} + \psi_\mu$, so it shows the large sample behavior of $\frac{(\hat{\mu}_R - \mu_R)'\hat{\beta}}{\hat{\beta}'\hat{\beta}}$, which is such that $\frac{\psi'_\mu(b+\psi_\beta)}{\sqrt{(b+\psi_\beta)'(b+\psi_\beta)}} \sim N(0, \Omega + \beta Q \beta')$, since ψ_μ is independent of ψ_β.

3. $-\lambda^*_{F,FM}\frac{\psi'_\beta(b+\psi_\beta)}{(b+\psi_\beta)'(b+\psi_\beta)}$: Since the ψ_β elements in the numerator are positively correlated, it creates a negative bias in the FM two-pass estimator for small values of b. It also implies that the large sample distribution of the FM two-pass estimator is not a normal one for such values of b. When b is much larger than ψ_β, so we are in a setting of sizeable β's, ψ_β becomes negligible compared to b in the $(b + \psi_\beta)$ elements. This is the setting covered by Shanken (1992), who provides the correction for the standard errors of the FM two-pass estimator to incorporate the contribution of this component for the variance of the FM two-pass estimator.

4. $\frac{e'\psi_\beta}{(b+\psi_\beta)'(b+\psi_\beta)}$: It appears in the large sample distribution of the FM two-pass estimator because of misspecification. When $e = 0$ or the identification strength, $b'b$, is much larger than the length of e, or the amount of misspecification, it has little effect on the large sample distribution of the FM two-pass estimator. Because of the dependence between the ψ_β elements in the numerator and denominator, this component only has a normal distribution for large values of b, so ψ_β becomes negligible in the $(b + \psi_\beta)$ elements. This is the setting covered by Kan, Robotti, and Shanken (2013), who provide a correction of the standard errors to incorporate this component.

The third and fourth components lead to a large sample distribution for the FM two-pass estimator which is not a normal one for the empirically relevant setting of small values of b. Test statistics based on the FM two-pass estimator, like the FM t-test and the KRS t-test, therefore become size distorted for such small values. To illustrate this, Figure 2 shows the simulated rejection frequencies of 5% significance FM and KRS t-tests on $\lambda^*_{F,FM}$ for a range of values of the misspecification and identification strengths.[6] Figure 2.1 shows the rejection frequencies for $H_0 : \lambda^*_{F,FM} = 0$, and Figure 2.2 is for $\lambda^*_{F,FM}$ corresponding with the values resulting from Figure 1.

Figure 2: Rejection frequencies of 5% significance FM and KRS t-tests for varying identification strengths β and misspecification $\tilde{e}$.

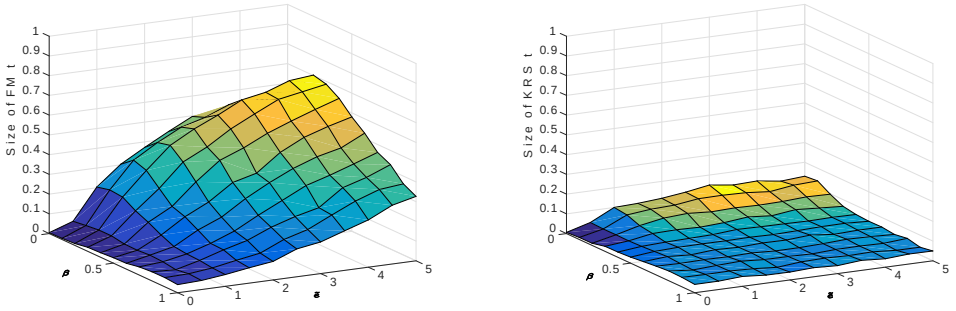


Figure 2.1: $H_0 : \lambda^*_{F,FM} = 0$

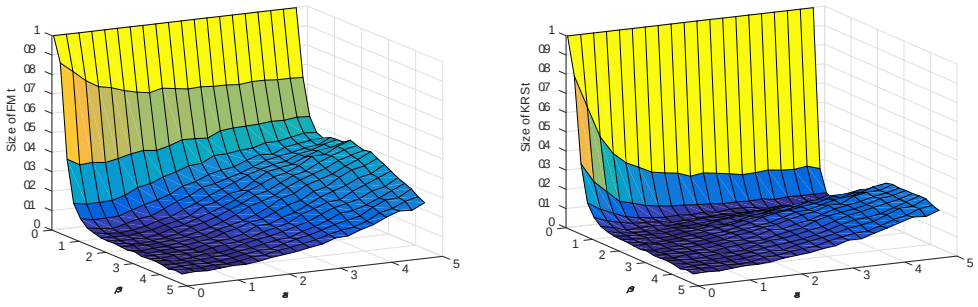


Figure 2.2: $H_0 : \lambda^*_{F,FM}$ results from Figure 1

[6] Throughout the paper, simulated data are generated from the linear factor model using parameters calibrated to the data from Kroencke (2017).

Figure 2.1 shows that the FM t-test and the KRS t-test are conservative or correctly sized when testing $H_0 : \lambda^*_{F,FM} = 0$ and there is no misspecification so $\tilde{e} = 0$. For increasing values of $\tilde{e}$, however, the FM t-test and the KRS t-test over-reject for small values of β, i.e. the rejection frequencies are larger than the nominal 5%. For the FM t-test, this over-rejection also extends to larger values of β.

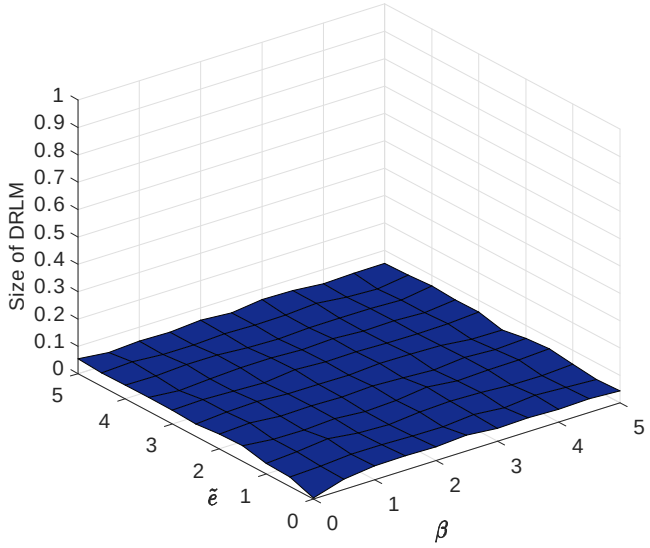
For testing $H_0 : \lambda^*_{F,FM}$ corresponding with the pseudo-true values resulting from Figure 1, Figure 2.2 shows that the FM t-test and the KRS t-test severely over-reject for small values of β which are less than $\tilde{e}$. When β approaches zero, the rejection frequencies are even equal to one, which is in line with Kan and Zhang (1999).

2.4.2 The DRLM test which remains size correct under misspecification and weak identification

The decomposition in Theorem 2 can also be conducted for the GLS risk premia estimator, so test statistics based on it, such as the GLS t-test, become similarly size distorted. We therefore use a statistic based on the CUE: the DRLM statistic proposed in Kleibergen and Zhan (2021), whose asymptotic distribution is bounded by the χ^2_K distribution. When using appropriate critical values from the χ^2_K distribution, the DRLM test of hypotheses specified on the pseudo-true value of the CUE remains size correct for general levels of misspecification and identification. In the Appendix, we state the DRLM statistic, and provide a brief discussion of its implementation.

Figure 3 presents the rejection frequencies of a 5% significance DRLM test of $H_0 : \lambda^*_{F,CUE} = 0$. We note that, unlike the large sample distributions of the FM and KRS t-tests, the limiting distributions of the statistics underlying the DRLM test do not depend on the tested parameter or the covariance matrices Ω and Q_{FF}. In contrast with Figure 2 for the simulated sizes of the FM and KRS t-tests, Figure 3 shows that the DRLM test is size correct (i.e. the rejection frequencies do not exceed the nominal 5%), and is conservative for combined small identification and misspecification strengths.

Figure 3: Rejection frequencies of the 5% significance DRLM test of $H_0 : \lambda^*_{F,CUE} = 0$ for varying identification strengths β and misspecification $\tilde{e}$.

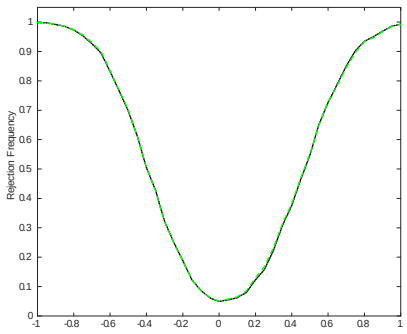


2.5 Power comparison

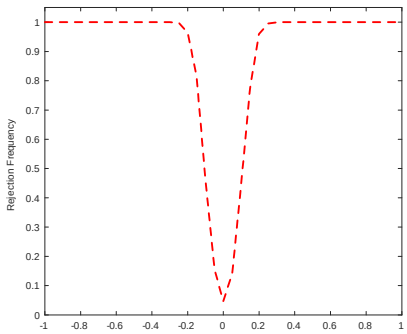
We briefly illustrate the power of the FM and KRS t-tests compared to the DRLM test. A more extensive power study of the DRLM test is conducted in Kleibergen and Zhan (2021). We first show the power for a correct specification, so all the aforementioned tests target the same value of the risk premium, while there is also strong identification. Thus, we have the ideal setting that β is sizeable while $\tilde{e} = 0$ in the data generation process. All the examined tests are therefore size correct in Figure 4a-b at the hypothesized value, i.e. all the rejection frequencies are near the nominal level of 5% when the distance to the tested risk premium is zero. Since there is no misspecification, the power curves of FM t (dotted black) and KRS t (dashed green) largely overlap in Figure 4a. The comparison of Figure 4a and 4b, however, also shows that the DRLM test can have more power than FM and KRS t-tests. This results since the DRLM test is based on the GLS framework, while the FM and KRS t-tests for Figure 4 are based on OLS.

Figure 4: Power comparison of FM, KRS t-tests and DRLM at the 5% level.

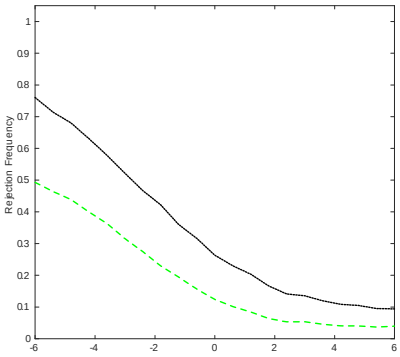
a-b: Correct specification, strong identification; c-d: Misspecification, weak identification.



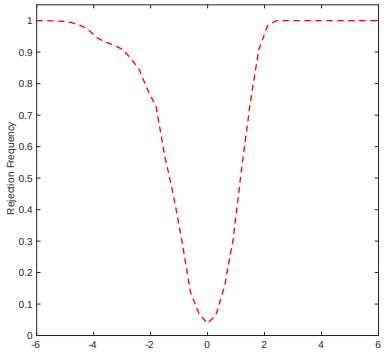
(a) FM t (black) and KRS t (green)



(b) DRLM



(c) FM t (black) and KRS t (green)



(d) DRLM

In contrast with Figure 4a-b, we allow for misspecification as well as weak identification in Figure 4c-d. This is achieved by considering smallish β and nonzero $\tilde{e}$ in the data generation process. For this scenario, Figure 4c shows that the FM and KRS t-tests are no longer size correct, while their power curves substantially differ from those in Figure 4a. In contrast, the DRLM test, which involves the power improvement rule mentioned in the Appendix, remains size correct in Figure 4d.

Overall, Figure 4 illustrates that it is appealing to use the DRLM test for conducting inference on risk premia. In contrast, the conventional FM t-test, together with the KRS

t-test, is jeopardized by joint misspecification and weak identification, both of which are prevalent in empirical studies as we show next.

3 Diagnostic statistics

Before we turn to empirical applications of the tests on the risk premia, we first apply the diagnostic statistics which help to gauge whether we can interpret the risk premia accordingly. The IS-statistic, which tests for a reduced rank value of β, is commonly used to test for identification of the risk premia in correctly specified settings. In misspecified settings, it is, however, no longer just the IS-statistic which governs the identification and interpretation of the pseudo-true value of the risk premia, but its difference with the J-statistic, which equals the minimal value of the sample CUE objective function. This is, for example, shown by the deviation of the FM two-pass pseudo-true value in (9) from its counterpart under correct specification, and further illustrated in Figure 1. The J and IS statistics have well established limiting distributions under their hypotheses of interest, which are χ^2_{N-K} for the J-statistic under correct specification, and χ^2_{N-K+1} for the IS-statistic under a reduced rank β matrix; see the Appendix for their explicit expressions. We are, however, mainly interested in these statistics because of the inequality between them, J-statistic $\leq$ IS-statistic, and because a close proximity between these two statistics shows that the pseudo-true value does not identify risk premia as discussed in Section 2. We therefore compute these two statistics first for eight well known specifications of the linear asset pricing model, and second for the specifications resulting from the factor zoo of Feng, Giglio, and Xiu (2020).

3.1 Empirical identification of risk premia

Figure 5 shows a scatter plot of the J and IS statistics for eight well known specifications of the linear asset pricing model: Fama and French (1993), Jagannathan and Wang (1996), Yogo (2006), Lettau and Ludvigson (2001), Savov (2011), Adrian, Etula, and Muir (2014),

Kroencke (2017) and He, Kelly, and Manela (2017).[7] In line with common practice, we incorporate the zero-β return, λ_0, while the factors used in the eight different specifications are:

1. Fama and French (1993), the prominent three, so-called Fama-French, factors: the market return R_m, SMB (small minus big), and HML (high minus low). We use the quarterly data from Lettau, Ludvigson, and Ma (2019) over 1963Q3 to 2013Q4, so $T = 202$, for the three factors, and the twenty-five size and book-to-market sorted portfolios as test assets.

2. Jagannathan and Wang (1996), three factors: R_m, corporate bond yield spread, and per capita labor income growth. We use their monthly data from July 1963 to December 1990 so $T = 330$, while one hundred size and beta sorted portfolios are used as test assets.

3. Yogo (2006), three factors: R_m, durable and nondurable consumption growth. The sample period is from 1951Q1 to 2001Q4 so $T = 204$, with twenty-five size and book-to-market sorted portfolios as test assets.

4. Lettau and Ludvigson (2001), three factors: consumption-wealth ratio, consumption growth, and their interaction. We use the quarterly data from 1963Q3 to 1998Q3 so $T = 141$, while the test assets are the twenty-five Fama-French portfolios.

5. Savov (2011), one factor: garbage growth. We use the same annual data, 1960 - 2006, while the test assets are the twenty-five Fama-French portfolios augmented by the ten industry portfolios, as suggested by Lewellen, Nagel, and Shanken (2010).

6. Adrian, Etula, and Muir (2014), one factor: leverage. Following Lettau, Ludvigson,

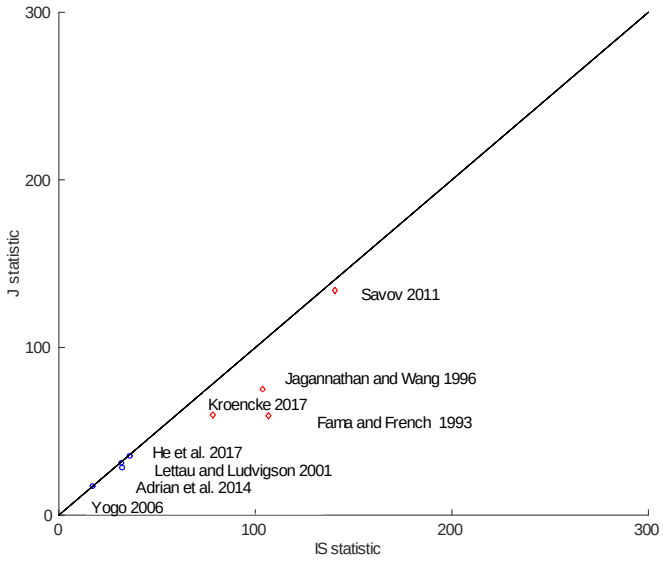
[7] We thank the authors of Jagannathan and Wang (1996), Yogo (2006), Lettau and Ludvigson (2001), Savov (2011), and Kroencke (2017) for sharing their data. For the models of Fama and French (1993), Adrian, Etula, and Muir (2014), and He, Kelly, and Manela (2017), we use the extended data of risk factors and test assets as in Lettau, Ludvigson, and Ma (2019).

and Ma (2019), we extend the time period to 1963Q3 - 2013Q4, and use twenty-five size and book-to-market sorted portfolios as test assets.

7. Kroencke (2017), one factor: unfiltered annual consumption growth. We use the post-war 1960 - 2014 sample from Kroencke (2017), while thirty portfolios, sorted by size, value and investment alongside the market portfolio, are used as test assets.

8. He, Kelly, and Manela (2017), two factors: banking equity-capital ratio and R_m. The data are also taken from Lettau, Ludvigson, and Ma (2019) for the period 1963Q3 - 2013Q4, and twenty-five size and book-to-market sorted portfolios are the test assets.

Figure 5: Scatter plot of J and IS statistics for different specifications



Notes: The zero-β return is incorporated. We revisit eight models and their associated factors. Fama and French (1993): R_m, SMB, and HML; Jagannathan and Wang (1996): R_m, corporate bond yield spread, and per capita labor income growth; Yogo (2006): R_m, durable and nondurable consumption growth; Lettau and Ludvigson (2001): consumption growth, consumption wealth ratio and their interaction; Savov (2011): garbage growth; Adrian, Etula, and Muir(2014): leverage; Kroencke (2017): unfiltered consumption growth; He, Kelly, and Manela (2017): R_m and the banking equity-capital ratio. For detailed descriptions of risk factors and test assets, we refer to the published articles.

The plotted points in Figure 5 roughly exhibit two patterns, which we illustrate by using different colors (red and blue). For Fama and French (1993), Jagannathan and Wang (1996), Savov (2011) and Kroencke (2017), we observe both large J and IS statistics, so the corresponding models appear to be misspecified. In contrast, for Lettau and Ludvigson (2001), Yogo (2006), Adrian, Etula, and Muir (2014), and He, Kelly, and Manela (2017), we encounter small J and IS statistics, so these models are likely to be weakly identified.

The IS-statistics for the first set of four specifications are large and mostly significant but they are also close to their respective J-statistics, which is revealed by the proximity of these points to the 45-degree line. We can therefore not conclude from these IS-statistics that we identify risk premia. The most likely setting for identification is for Fama and French (1993), whose (IS, J) point is most distant from the 45-degree line.

For the second set of four points, the J-statistics are small and mostly insignificant, indicating that the models might not be misspecified. However, their corresponding IS-statistics are also similarly small, which implies that the low values of the J-statistics are induced by the low values of the IS-statistics, since J-statistics are necessarily exceeded by IS-statistics. Hence, we cannot identify risk premia for these specifications.

Table 1: J **and** IS **statistics**

Panel A contains the J and IS statistics plotted in Figure 5, for which the zero-β return is incorporated. In Panel B, the zero-β return is removed so $\lambda_0 = 0$. Significance at 1%, ***; 5%, **; 10%, *.

	(A) Impose $\lambda_0 = 0$: No		(B) Impose $\lambda_0 = 0$: Yes	
	J-statistic	IS-statistic	J-statistic	IS-statistic
Fama and French (1993)	59.34***	106.81***	87.47***	974.39***
Jagannathan and Wang (1996)	75.07	103.54	86.46	103.56
Lettau and Ludvigson (2001)	31.11*	31.75*	37.15**	40.90**
Yogo (2006)	17.14	17.34	19.42	19.60
Savov (2011)	134.27***	140.68***	268.60***	296.78***
Adrian, Etula, and Muir (2014)	28.42	31.97	30.41	42.03**
Kroencke (2017)	59.84***	78.47***	60.03***	102.77***
He, Kelly, and Manela (2017)	35.32**	35.88**	44.44***	59.74***

3.2 Removing the zero-β return

Panel A in Table 1 states the values of the J and IS statistics plotted in Figure 5. Panel B states these statistics when the zero-β return is removed, so $\lambda_0 = 0$. All statistics in Panel B therefore exceed their corresponding counterparts in Panel A. Removing the zero-β return thus increases both the misspecification and identification. The misspecification has increased since we removed a parameter from the pricing equation, while identification has improved since removing the zero-β return allows to identify risk premia when the β's are constant over the assets, which was not so when the zero-β return was included. For some of the specifications, the increase of the J and IS statistics is disproportional. This is most notably so for the Fama and French (1993) specification, where the IS-statistic increases ninefold while the J-statistic does so only minorly. For the other specifications, the increase of the IS-statistic typically exceeds that of the J-statistic, but not by an amount which makes it clear that the risk premia become well identified as is the case for the Fama and French (1993) specification. For more than half of the specifications, the removal of the zero-β return has therefore little effect on the identification of risk premia.

We note that the J and IS statistics presented in Figure 5 are just for illustrative purposes. We do not aim to use these statistics to strictly reject or favor any model in Figure 5, since there are many other issues involved. These issues include, for example, that the models used for Figure 5 contain various numbers of risk factors; in addition, their corresponding empirical studies have used non-identical test assets; furthermore, the considered time periods and frequencies also vary to a large extent. To address these concerns, we next use the factor zoo data from Feng, Giglio, and Xiu (2020) to investigate misspecification and weak identification in a more systematic manner.

3.3 Misspecification and weak identification in the factor zoo

How prevalent are misspecification and weak identification in empirical asset pricing studies? With this question in mind, we extend our analysis to the factor zoo collected by Feng, Giglio,

and Xiu (2020), which covers one hundred and fifty risk factors from July 1976 to December 2017, so $T = 498$. We use the twenty-five Fama and French size and book-to-market portfolios as test assets, which have been widely used as the default choice.

3.3.1 Prevalence of misspecification and weak identification

In line with existing studies, we evaluate all possible specifications of linear factor models resulting from the factor zoo with $K = 1$, 2, 3, 4, 5 and 6 factors. For $K = 1$, we therefore consider $C_{150}^1 = 150$ single factor models. Similarly, we examine $C_{150}^2 = 11,175$ $(= 150 \times 149/2)$ two-factor models; $C_{150}^3 = 551,300$ three-factor models; $C_{150}^4 = 20,260,275$ four-factor models; $C_{150}^5 = 591,600,030$ five-factor models; and $C_{150}^6 = 14,297,000,725$ six-factor models.[8] For each model, we compute its J and IS statistics while incorporating the zero-β return.

Table 2 states the frequencies of, at the 5% level, significant values of the J-statistics, and insignificant values of the IS-statistics, signaling misspecification and weak identification, respectively. Table 2 shows that, when we increase the number of factors, the resulting factor models tend to be less misspecified, while becoming weaker identified. This results naturally since on the one hand, adding more factors helps to better explain asset returns, so the resulting models are less likely to be rendered misspecified; while on the other hand, since some factors could be closely related to others, including more factors decreases the distance of the β matrix from a reduced rank value, which leads to weaker identified models. We note that the reported percentages in Table 2 likely understate the severeness of misspecification, because the misspecification J-statistic tends to be insignificant under weak identification, as we have discussed for Figure 5.

Overall, Table 2 shows that the majority of the examined models seem to suffer from misspecification and/or weak identification. Apparently, there is also a trade-off between the reported percentages of misspecification and weak identification in Table 2. We previously

[8] We used high-performance computers to conduct this study.

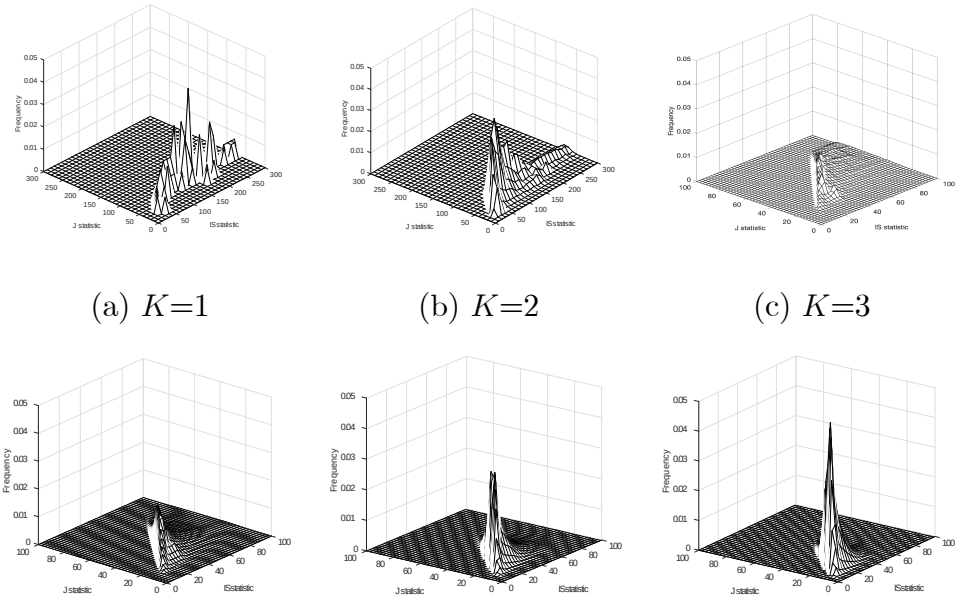
have, however, shown that we cannot analyze the J and IS statistics in isolation, as in Table 2, to determine whether risk premia are identified, so we next analyze them jointly.

Table 2: **Prevalence of misspecification and weak identification**

The data of one hundred and fifty risk factors are taken from Feng, Giglio, and Xiu (2020). The test assets are the twenty-five Fama and French size and book-to-market portfolios from July 1976 to December 2017. Models are deemed misspecified at the 5% level, if the p-value of the J-statistic does not exceed 5%. Models are deemed weakly identified at the 5% level, if the p-value of the IS-statistic exceeds 5%. The $\lambda_0 = 0$ restriction is not imposed.

	Number of Factors, K					
	1	2	3	4	5	6
Number of Models (C_{150}^K)	150	11175	551300	20260275	591600030	14297000725
Misspecified (%)	98.67%	95.84%	87.11%	68.65%	43.58%	18.37%
Weakly Identified (%)	0.67%	2.08%	5.49%	15.40%	33.35%	50.28%

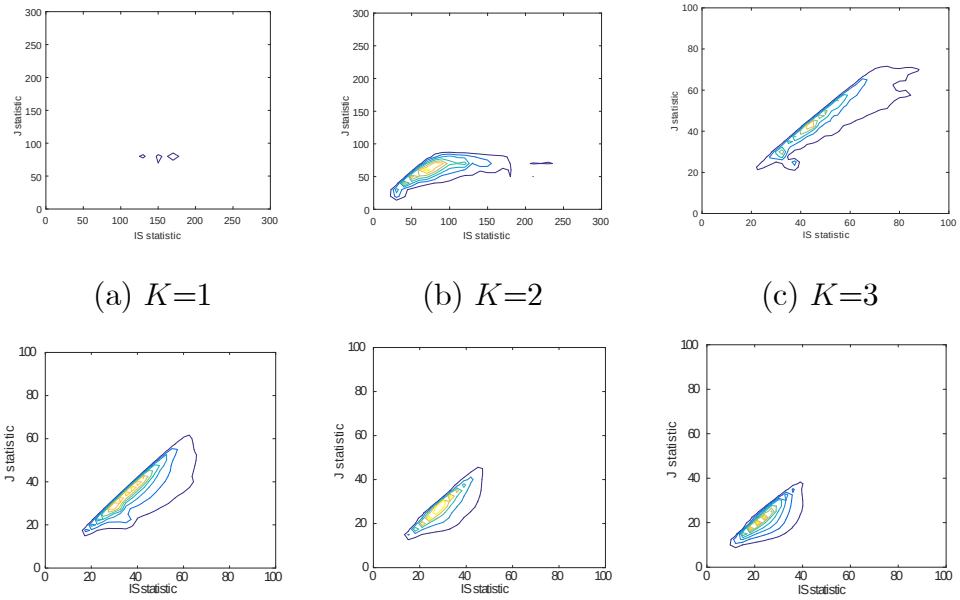
Figure 6: Joint empirical density of J and IS statistics over the specifications with K factors from the factor zoo.



(a) K=1 (b) K=2 (c) K=3

(d) K=4 (e) K=5 (f) K=6

Figure 7: Contour lines of the joint empirical density of J and IS statistics over the specifications with K factors from the factor zoo.



(a) K=1 (b) K=2 (c) K=3

(d) K=4 (e) K=5 (f) K=6

3.3.2 Joint empirical density of J and IS statistics resulting from the different specifications

Because the J and IS statistics jointly indicate if the risk premia are identified, the six pictures in Figure 6 show the bivariate empirical density functions (histograms) of the J and IS statistics resulting from all the specifications having one to six factors presented in Table 2. Figure 7 shows the contour lines of these six empirical density functions.

When $K = 1$, Figures 6a and 7a show that almost all single factor models are associated with large significant (at the 5% level) J-statistics. This is as expected, since it is unlikely that a single factor explains all the variation in the cross-section of expected asset returns. Specifically, 148 out of 150 single factor models are deemed misspecified by their significant J-statistics at the 5% level, leading to $148/150 \approx 98.67\%$ in Table 2. The remaining two single factor models have small J-statistics (see the bottom left of Figure 6a), but their IS-statistics

are also small. One of these two IS-statistics is even insignificant at the 5% level, leading to the $1/150 \approx 0.67\%$ for weak identification in Table 2, while the other is insignificant at the 1% level. Thus, these two factors are of poor quality, and their resulting single factor models should also be deemed misspecified. The J-statistic fails to signal misspecification for these two models, because the small IS-statistic forces it to be very small, given that the J-statistic is less than or equal to the IS-statistic. Joint misspecification and weak identification therefore occur in these two single factor models.

When we increase the number of factors so $K = 2$, 3, 4, 5 and 6, Figures 6b-f and 7b-f show two clear patterns:

1. The J and IS statistics are overall decreasing.

2. The empirical bivariate density of the J and IS statistics moves closer to the 45-degree line.

While Table 2 shows that for $K = 6$, more than 50% of the examined specifications are weakly identified, the second pattern listed above implies that many more specifications are weakly identified, and similarly so for specifications involving fewer factors. This is analogous to what we observed for Figure 5. Taken all together, it all shows that joint misspecification and weak identification is a common problem that needs to be addressed.

Encouragingly, Figures 6 and 7 also show that there are specifications for which the (IS , J) combination is distant from the 45-degree line. For these specifications, the IS-statistics exceed their J-statistics to a larger extent when compared to those in Figure 5. From a purely statistical point of view, the specifications with J much smaller than IS are worth investigation, since their pricing errors tend to be small while the risk premia are likely to be well identified. It would be interesting if researchers could further relate these models to economic theories. Yet given the large number of such models as shown by Figure 6, we leave them for future research.

4 Application

We show the importance of J, IS and DRLM statistics for applied asset pricing by revisiting two prominent examples: the Fama and French (1993) model and the conditional consumption capital asset pricing model from Lettau and Ludvigson (2001). Table 3 therefore reports the estimation results for the three-factor specification used in Lettau and Ludvigson (2001), and for Fama and French (1993) using three different data sets. Accordingly, Figure 8 shows the joint 95% confidence sets resulting from DRLM for all four specifications in Table 3.

4.1 Fama and French (1993)

The Fama and French (1993) three-factor model has been widely used as a benchmark in the asset pricing literature; see, e.g., Lettau and Ludvigson (2001) and Lettau, Ludvigson, and Ma (2019). In line with the large, significant (at the 1% level) J-statistics in Table 3, the Fama and French (1993) three-factor model is well acknowledged to be misspecified. Yet this model appears able to explain the cross-section of asset returns, as reflected by the reported large cross-sectional R^2.

Using quarterly data from Lettau and Ludvigson (2001) and Lettau, Ludvigson, and Ma (2019), the two-pass procedure of Fama and MacBeth (1973) suggests that the risk premia on SMB and HML are both positive; in contrast, the point estimate of the market premium is negative in Lettau, Ludvigson, and Ma (2019), but positive in Lettau and Ludvigson (2001); see Panel A versus Panel B in Table 3. Because Lettau, Ludvigson, and Ma (2019) use longer time series than Lettau and Ludvigson (2001), its IS-statistic is considerably larger (106.81 vs. 47.01), reflecting that more information is available for identification in Lettau, Ludvigson, and Ma (2019). On the other hand, the larger J-statistic for Lettau, Ludvigson, and Ma (2019) (59.34 vs. 45.38) implies more severe misspecification. A question thus arises: how do we reconcile the seemingly conflicting findings in Lettau, Ludvigson, and Ma (2019) and Lettau and Ludvigson (2001) for R_m in the Fama and French (1993) three-factor model?

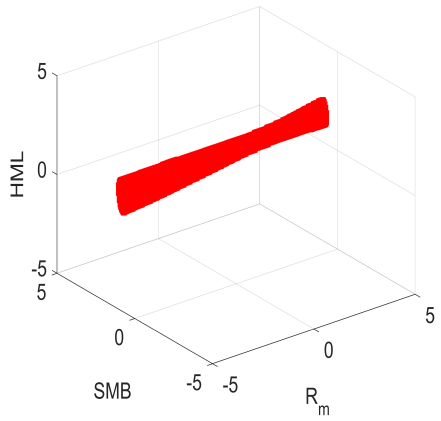
Table 3: **Risk Premia λ_F for the three-factor models of Fama and French (1993) and Lettau and Ludvigson (2001)**

The test assets are the 25 Fama-French portfolios from 1963Q3 - 2013Q4 used in Lettau, Ludvigson, and Ma (2019) for Panel A, from 1963Q3 - 1998Q3 used in Lettau and Ludvigson (2001) for Panel B and Panel C, and from July 1963 - June 2021 downloaded from French's website for Panel D, respectively. The estimates of λ_F result from the Fama and MacBeth (1973) two-pass procedure, and are identical to those reported in Lettau, Ludvigson, and Ma (2019) and Lettau and Ludvigson (2001) when using the same data. The presented FM t-statistic and its resulting 95% confidence interval (C.I.) of risk premia do not use the Shanken (1992) correction, while Shanken t does. The KRS t is computed by using the programs provided by Kan, Robotti, and Shanken (2013). The zero-β return is incorporated.

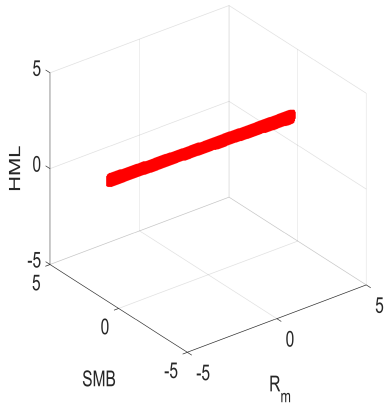
	A: Lettau, Ludvigson, and Ma (2019)			B: Lettau and Ludvigson (2001)		
	R_m	SMB	HML	R_m	SMB	HML
$\lambda_{F,FM}$	-1.96	0.70	1.35	1.33	0.47	1.46
$\lambda_{F,CUE}$	-4.15	0.82	0.86	-11.26	0.69	1.52
FM t	-1.72	1.67	2.64	0.83	0.94	3.24
95% C.I.	(-4.18, 0.27)	(-0.12, 1.52)	(0.35, 2.35)	(-1.81, 4.46)	(-0.51, 1.45)	(0.58, 2.34)
Shanken t	-1.64	1.66	2.60	0.78	0.94	3.22
95% C.I.	(-4.29, 0.38)	(-0.13, 1.52)	(0.33, 2.37)	(-2.02, 4.68)	(-0.51, 1.45)	(0.57, 2.35)
KRS t	-1.33	1.65	2.52	0.63	0.96	3.26
95% C.I.	(-4.83, 0.92)	(-0.13, 1.53)	(0.30, 2.40)	(-2.78, 5.44)	(-0.49, 1.43)	(0.58, 2.34)
R^2		0.73			0.80	
J-statistic		59.34			45.38	
IS-statistic		106.81			47.01	

	C: Lettau and Ludvigson (2001)			D: Monthly data from July 1963 - June 2021		
	$\triangle c$	cay	$\triangle c \times cay$	R_m	SMB	HML
$\lambda_{F,FM}$	0.02	-0.13	0.06	-0.53	0.13	0.34
$\lambda_{F,CUE}$	-1.45	-3.80	0.01	-0.52	0.19	0.32
FM t	0.20	-0.43	3.12	-1.75	1.09	2.93
95% C.I.	(-0.20, 0.25)	(-0.70, 0.45)	(0.02, 0.09)	(-1.11, 0.06)	(-0.11, 0.37)	(0.11, 0.57)
Shanken t	0.15	-0.31	2.25	-1.74	1.09	2.92
95% C.I.	(-0.29, 0.34)	(-0.93, 0.68)	(0.01, 0.11)	(-1.12, 0.07)	(-0.11, 0.38)	(0.11, 0.57)
KRS t	0.08	-0.27	1.95	-1.67	1.09	2.94
95% C.I.	(-0.53, 0.58)	(-1.03, 0.78)	(-0.00, 0.11)	(-1.14, 0.09)	(-0.11, 0.37)	(0.11, 0.57)
R^2		0.70			0.75	
J-statistic		31.11			52.25	
IS-statistic		31.75			425.55	

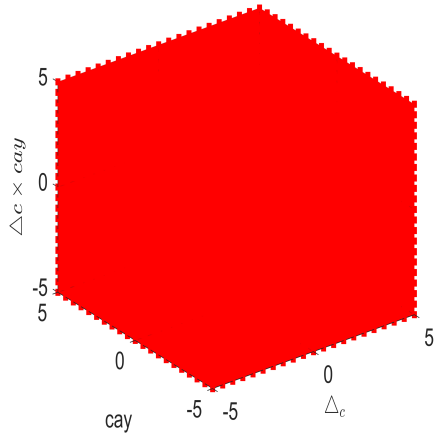
Figure 8: Joint 95% confidence sets of risk premia from the DRLM test.



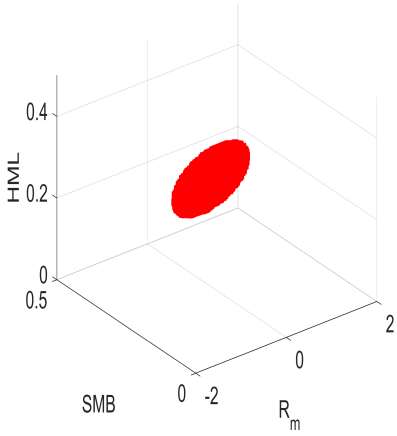
(a) Data: Lettau, Ludvigson, and Ma (2019)



(b) Data: Lettau and Ludvigson (2001)



(c) Data: Lettau and Ludvigson (2001)



(d) Data: Monthly data, July 1963 - June 2021

Notes: The red region consists of risk premia values that are not rejected by the DRLM test at the 5% significance level. The test assets are the twenty-five Fama-French portfolios taken from Lettau, Ludvigson, and Ma (2019) over 1963Q3 - 2013Q4 for (a), Lettau and Ludvigson (2001) over 1963Q3 - 1998Q3 for (b) and (c), and monthly data from July 1963 - June 2021 downloaded from French's website for (d), respectively. (a)(b)(c)(d) correspond to Panels A, B, C, D of Table 3. The zero-β return is incorporated.

To help answer this question, we employ the DRLM test, since this test is robust to misspecification as well as the strength of identification. Specifically, we test every risk premium value between -5 and 5, since Table 3 Panel A and Panel B indicate that risk premia on all three factors likely lie within this range. The resulting joint confidence sets at the 5% level are presented in Figure 8a-b, using the data from Lettau, Ludvigson, and Ma (2019) and Lettau and Ludvigson (2001), respectively.

Interestingly, Figure 8a-b show that the risk premia on SMB and HML are strongly identified, while the risk premium on R_m is not. These findings remain similar no matter whether we adopt the data from Lettau, Ludvigson, and Ma (2019) or Lettau and Ludvigson (2001). In both Figure 8a-b, we can not reject any risk premium value on R_m between -5 and 5, which thus helps reconcile the difference in the confidence intervals for the risk premium on R_m reported in Table 3 Panel A and Panel B.

For SMB and HML, however, the DRLM test yields tight 95% confidence sets for their risk premia in Figure 8. For example, Figure 8a implies that the range of the risk premium on SMB is (0.59, 1.06), while the range of the risk premium on HML is (0.04, 2.00). Similar ranges can be derived from Figure 8b. All these sets thus largely overlap with those resulting from the FM t-test as reported in Table 3 Panel A and Panel B.

One might wonder what causes the large difference in the risk premia between R_m and SMB, HML. To illustrate, we present their β estimates and associated t-statistics in Table 4. It is clear in Table 4 that the three Fama and French (1993) factors are all closely related to the test asset returns as reflected by the significant t-statistics. There is, however, little cross-sectional variation in the estimated betas of R_m, i.e. the β estimates on R_m are all close to 1. Thus, if the zero-β return is incorporated, we have near-multicollinearity in the $(\iota_N \vdots \beta)$ matrix for the cross-sectional regression, causing the market risk premium to be weakly identified. Consequently, we observe in Table 3 Panels A and B the seemingly conflicting risk premium values on R_m, but not for SMB and HML. For the same reason, we observe in Figure 8a-b the wide range for the risk premium on R_m, but narrower ones for SMB, HML.

Table 4: β **with 25 portfolio returns**

The test assets are the 25 Fama-French portfolios. For R_m, SMB, and HML, we use the data from 1963Q3 - 2013Q4 used by Lettau, Ludvigson, and Ma (2019). For $\triangle c$, cay, $\triangle c \times cay$, we use the data from 1963Q3 - 1998Q3 used by Lettau and Ludvigson (2001). The reported beta estimates and their associated t-statistics result from the first pass time-series regression of the Fama and MacBeth (1973) methodology, see also Kleibergen, Kong, and Zhan (2020) for the same results reported for Lettau and Ludvigson (2001).

	R_m		SMB		HML		$\triangle c$		cay		$\triangle c \times cay$	
	$\hat{\beta}$	t-stat	$\hat{\beta}$	t-stat	$\hat{\beta}$	t-stat	$\hat{\beta}$	t-stat	$\hat{\beta}$	t-stat	$\hat{\beta}$	t-stat
(1)	1.08	26.50	1.50	25.11	-0.31	-6.36	6.35	2.32	4.25	2.96	-4.56	-0.22
(2)	0.94	32.73	1.33	31.52	0.02	0.69	6.30	2.61	3.66	2.90	2.25	0.12
(3)	0.86	29.98	1.14	27.08	0.19	5.56	5.11	2.28	3.22	2.75	3.22	0.19
(4)	0.81	26.16	1.10	24.32	0.29	8.02	5.50	2.58	2.96	2.65	5.12	0.32
(5)	0.94	28.29	1.17	24.04	0.57	14.37	5.79	2.53	2.65	2.21	11.36	0.65
(6)	1.11	36.22	1.05	23.49	-0.33	-8.97	4.38	1.76	4.55	3.49	-15.82	-0.84
(7)	0.94	36.04	0.97	25.32	0.05	1.48	3.63	1.65	3.25	2.83	0.02	0.00
(8)	0.92	33.18	0.76	18.80	0.22	6.81	3.92	1.97	3.07	2.95	-1.08	-0.07
(9)	0.91	29.94	0.68	15.32	0.42	11.67	3.52	1.91	2.66	2.76	6.93	0.50
(10)	0.98	30.95	0.81	17.55	0.66	17.47	4.83	2.41	2.16	2.06	8.48	0.56
(11)	1.08	39.19	0.76	18.65	-0.39	-11.96	2.70	1.21	4.53	3.88	-22.20	-1.31
(12)	0.99	35.98	0.60	14.70	0.07	2.16	2.76	1.47	3.64	3.69	-3.35	-0.23
(13)	0.91	29.02	0.49	10.71	0.28	7.51	2.92	1.69	2.63	2.91	4.40	0.34
(14)	0.95	28.29	0.40	8.19	0.47	11.66	2.58	1.57	2.73	3.18	0.02	0.00
(15)	0.91	23.16	0.62	10.77	0.57	12.20	3.71	1.98	2.15	2.19	4.78	0.34
(16)	1.07	38.14	0.43	10.37	-0.42	-12.67	1.97	1.02	4.22	4.20	-20.28	-1.39
(17)	1.01	31.06	0.32	6.63	0.13	3.31	2.62	1.49	3.28	3.58	-10.24	-0.77
(18)	1.00	32.59	0.22	4.94	0.32	8.74	1.94	1.21	2.57	3.06	-4.98	-0.41
(19)	0.98	30.21	0.22	4.56	0.40	10.44	2.50	1.56	2.32	2.76	0.60	0.05
(20)	1.05	27.69	0.36	6.41	0.61	13.46	3.78	2.05	2.26	2.34	3.28	0.23
(21)	1.02	46.70	-0.20	-6.30	-0.27	-10.39	1.61	1.01	2.77	3.33	-21.22	-1.76
(22)	0.99	39.05	-0.19	-5.16	0.07	2.25	1.16	0.80	2.53	3.33	-4.20	-0.38
(23)	0.93	32.10	-0.23	-5.34	0.27	7.93	2.32	1.88	2.46	3.80	-8.27	-0.88
(24)	0.94	36.21	-0.16	-4.29	0.45	14.52	1.24	0.94	2.24	3.25	-10.49	-1.05
(25)	0.99	26.11	-0.13	-2.40	0.56	12.31	3.07	2.13	1.59	2.11	-1.25	-0.11

4.2 Lettau and Ludvigson (2001)

To compare with Fama and French (1993), we consider the conditional consumption capital asset pricing model from Lettau and Ludvigson (2001) where the three factors are consumption growth, $\triangle c$, consumption-wealth ratio, cay, and their interaction, $\triangle c \times cay$. The significant FM t and Shanken t-statistics on the interaction $\triangle c \times cay$, the relatively small, insignificant (at the 5% level) J-statistic, together with the large cross-sectional R^2 reported in Panel C of Table 3 provide considerable motivation for this model for asset pricing.

The small IS-statistic in Panel C of Table 3, however, shows that the three-factor model of Lettau and Ludvigson (2001) is just weakly identified. Furthermore, given its proximity to the J-statistic, the small value of the J-statistic also results from it, since the J-statistic is at most equal to the IS-statistic. The risk premia can therefore not be identified. Consequently, the significant FM t-statistic on $\triangle c \times cay$ should be interpreted with caution, since the FM t-test is now unreliable and this reasoning similarly applies to the KRS t-test. Also, as warned by Kleibergen and Zhan (2015), weakly identified models can yield spuriously large cross-sectional R^2's, which should then not be taken as the evidence in support of asset pricing.

Next, we apply the DRLM test to the three-factor model of Lettau and Ludvigson (2001). As shown by Figure 8c, we can not reject any hypothesized value for risk premia in the range of [-5, 5], reflecting that the risk premia can not be identified for the conditional consumption capital asset pricing model in Lettau and Ludvigson (2001), which is in line with the proximity of the J and IS statistics.

The challenge to identify the risk premia in Lettau and Ludvigson (2001) is further explained by Table 4, which contains the β estimates and their associated t-statistics. Table 4 shows that $\triangle c \times cay$ is poorly correlated with asset returns, which makes a column of β statistically close to zero (i.e. tiny t-statistics in the last column). Thus, the full rank condition of β is problematic, which leads to the weak identification problem in Lettau and Ludvigson (2001). Consequently, we observe a small IS-statistic, which tests for a full rank

value of β, in Panel C of Table 3, and uninformative confidence sets in Figure 8c.

In light of the findings in Figure 8a-c, one might wonder when the DRLM test could yield an informative confidence set for the risk premia. We show that a bounded confidence set of the risk premia is feasible if we just have more time series observations, or if we remove the zero-β return, so $\lambda_0 = 0$, as presented in the next two subsections, respectively.

Table 5: β **with 25 portfolio returns using alternative data for Fama and French (1993)**

The test assets are the 25 Fama-French portfolios. For R_m, SMB, and HML in the left panel, we use the data from 1963Q3 - 1998Q3 as in Lettau and Ludvigson (2001); and from July 1963 - June 2021, downloaded from French's website, for the right panel, respectively. The reported beta estimates and their associated t-statistics result from the first pass time-series regression of the Fama and MacBeth (1973) methodology.

	R_m		SMB		HML		R_m		SMB		HML	
	$\hat{\beta}$	t-stat	$\hat{\beta}$	t-stat	$\hat{\beta}$	t-stat	$\hat{\beta}$	t-stat	$\hat{\beta}$	t-stat	$\hat{\beta}$	t-stat
(1)	1.00	21.40	1.53	25.43	-0.29	-4.44	1.09	49.32	1.38	43.79	-0.48	-14.87
(2)	0.98	30.05	1.37	32.47	0.11	2.49	0.95	53.68	1.31	51.69	-0.17	-6.58
(3)	0.97	30.66	1.21	29.54	0.28	6.39	0.92	75.00	1.09	61.84	0.15	8.16
(4)	0.96	34.84	1.15	32.46	0.43	11.38	0.87	72.34	1.07	61.78	0.32	17.81
(5)	1.03	32.38	1.24	30.15	0.76	17.24	0.93	52.58	1.08	42.39	0.53	20.15
(6)	1.04	27.17	1.08	21.77	-0.49	-9.31	1.12	76.44	1.03	48.82	-0.51	-23.70
(7)	1.01	31.82	1.00	24.45	0.00	0.10	1.00	82.03	0.92	52.68	-0.01	-0.73
(8)	1.02	35.96	0.83	22.65	0.23	5.96	0.96	75.66	0.76	41.93	0.28	14.97
(9)	1.02	36.41	0.69	19.10	0.47	12.23	0.94	80.20	0.72	42.72	0.47	27.34
(10)	1.07	36.32	0.83	21.88	0.76	18.66	1.07	84.69	0.88	48.20	0.65	35.21
(11)	1.06	31.54	0.75	17.17	-0.46	-10.02	1.10	78.20	0.74	36.81	-0.51	-24.97
(12)	1.01	31.29	0.65	15.68	0.00	0.05	1.01	73.17	0.60	26.93	0.06	2.91
(13)	1.00	34.68	0.54	14.47	0.31	7.83	0.97	68.10	0.44	21.75	0.35	16.85
(14)	1.02	35.48	0.43	11.56	0.50	12.58	0.98	72.25	0.44	22.51	0.55	27.59
(15)	1.07	30.24	0.61	13.28	0.78	15.87	1.07	63.75	0.57	23.75	0.74	29.92
(16)	1.02	30.70	0.39	9.04	-0.48	-10.42	1.07	74.82	0.39	19.12	-0.45	-21.72
(17)	1.06	31.66	0.32	7.48	0.02	0.49	1.06	69.04	0.23	10.45	0.17	7.52
(18)	1.05	34.63	0.22	5.50	0.32	7.54	1.03	65.57	0.18	8.07	0.40	17.55
(19)	1.07	31.94	0.22	5.03	0.50	10.92	1.02	66.30	0.23	10.28	0.56	24.57
(20)	1.12	25.63	0.44	7.86	0.70	11.67	1.15	61.12	0.30	11.00	0.78	28.16
(21)	1.00	34.39	-0.23	-6.20	-0.39	-9.61	0.98	95.17	-0.23	-15.89	-0.31	-20.80
(22)	1.02	33.17	-0.21	-5.19	-0.05	1.16	0.97	75.34	-0.18	-9.65	0.11	5.68
(23)	0.91	24.75	-0.23	-4.87	0.16	3.09	1.05	87.14	-0.13	-7.54	0.35	20.06
(24)	1.00	35.20	-0.16	-4.28	0.44	11.26	1.05	82.15	-0.20	-10.70	0.37	19.72
(25)	0.99	20.47	-0.09	-1.39	0.69	10.30	1.06	77.70	-0.26	-13.49	0.39	19.42

4.3 More time series observations

With more time series observations, we expect more information in the data and consequently, stronger identification of the risk premia. Figure 8d therefore uses monthly data from July 1963 to June 2021, so $T = 696$, for the Fama and French (1993) three factors and the twenty-five size and book-to-market sorted portfolios, which are downloaded from French's online data library. Table 5 (right panel) shows the resulting β estimates from which it is clear that the standard errors of the estimates have almost halved, which doubles the t-statistics testing their significance, while there is also more variation in the β estimates for R_m when compared to the left panel of Table 5 with fewer observations. The estimate of β is consequently distant from a reduced rank value as shown by the large IS-statistic, 425.55, in Panel D of Table 3. It considerably exceeds the resulting J-statistic, 52.25, indicating strong identification of the risk premia.

In line with such strong identification, Figure 4d shows that the joint 95% confidence set of the risk premia on the Fama and French (1993) factors resulting from the DRLM test lies in a tight region of the 3-dimensional space. Specifically, this confidence set contains all values of the risk premia that are not rejected by the DRLM test at the 5% level. Combining all these values results in the red joint confidence set plotted in Figure 4d. Moreover, projecting the joint confidence set to each axis yields the 95% confidence intervals for each risk premium: (-1.22, 0.18) for R_m, (0.12, 0.26) for SMB, and (0.24, 0.41) for HML, respectively. These 95% confidence intervals are also comparable to, but often smaller than those reported in Panel D of Table 3 by inverting the FM t, Shanken t and KRS t tests.

4.4 Remove the zero-β return, impose $\lambda_0 = 0$

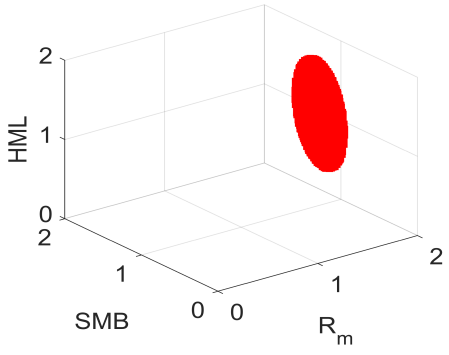
Instead of using richer data, we can also remove the zero-β return, so impose $\lambda_0 = 0$, to improve the identification of the risk premia. When we do so by using the data from Lettau, Ludvigson, and Ma (2019), the J-statistic equals 87.47 and the IS-statistic becomes 974.39. Hence, the IS-statistic has increased ninefold while the J-statistic remains almost the same

compared with the specification including the zero-β return in Panel A of Table 3. It indicates that the risk premia are now well identified.

To illustrate further, we use DRLM to construct the joint 95% confidence set in Figure 9, where $\lambda_0 = 0$ is imposed for the Fama and French (1993) model. Figure 9 is to be compared with Figure 8a. They use the same data, but $\lambda_0 = 0$ is imposed for Figure 9 but not for Figure 8a. Clearly, the $\lambda_0 = 0$ restriction strongly improves the identification of the risk premia, so we observe a bounded 95% confidence set in Figure 9. Projecting the joint confidence set to each axis leads to 95% confidence intervals for the risk premia on R_m, SMB, and HML, respectively: $(1.52, 1.84)$ for R_m, $(0.66, 1.06)$ for SMB, and $(0.59, 1.93)$ for HML.

Since the three factors are traded, we can also infer the 95% confidence sets on their risk premia by using the average value of the respective factor $\pm$ 1.96 $\times$ standard deviation, which leads to $(0.41, 2.80)$ for R_m, $(0.18, 1.78)$ for SMB and $(0.05, 1.81)$ for HML. These confidence intervals are comparable to those derived above by the projection method, but are considerably wider. All these empirical findings therefore lend credibility to the proposed DRLM test.

Figure 9: Joint 95% confidence set of risk premia by the DRLM test for the Fama and French (1993) model with the $\lambda_0 = 0$ restriction.



Notes: The red region consists of risk premia values that are not rejected by the DRLM test at the 5% significance level. The test assets are the twenty-five Fama-French portfolios taken from Lettau, Ludvigson, and Ma (2019) over 1963Q3 - 2013Q4.

4.5 Further discussion

Though misspecified, the Fama and French (1993) three-factor model has been well documented to exhibit relatively strong correlations with asset returns. Indeed, the model leads to large IS-statistics as we report in Table 3. Astonishingly, even when the IS-statistic is significant and exceeds 100 (see Panel A of Table 3), we still find that the risk premium on R_m is only weakly identified in the beta representation. Given that a large number of empirical studies, as indicated by Table 1, have weaker strength of identification than Fama and French (1993), it is therefore unlikely that their risk premia could be precisely identified. To achieve stronger identification, we therefore need to adopt richer data or impose extra restrictions, as indicated by Figure 8d and Figure 9, respectively.

The conditional consumption capital asset pricing model in Lettau and Ludvigson (2001) serves as an example where the FM t-statistic, the misspecification J-statistic, and the cross-sectional R^2 line up nicely to show support for the model. Yet the model appears to lack identification, as reflected by the tiny difference between the J and IS statistics, and the wide confidence set resulting from the DRLM test, since the rank condition of β is likely violated. These empirical findings thus clearly show the importance of adopting the J, IS statistics and the DRLM test for assessing asset pricing models.

In the Appendix, we further extend our empirical analysis to the five-factor model of Fama and French (2015). Similar to Figure 9, we find that the DRLM test can lead to tight 95% confidence sets for the risk premia on R_m, SMB, HML. For the other two factors (profitability, RMW; investment, CMA), we find that the 95% confidence sets resulting from DRLM appear relatively wide, especially for the investment factor. These findings are consistent with those in Kleibergen and Zhan (2015), who show that the commonly used 25 Fama-French portfolios have a strong factor structure, i.e., the majority of variations in these portfolios are likely captured by three factors, so it is hard to precisely identify risk premia on all five factors by using such test assets.

5 Conclusion

An alarmingly large number of factors in the asset pricing literature are able to yield significant t-statistics on their risk premia, together with seemingly promising misspecification J-statistics and cross-sectional R^2's. The credibility of these conventional statistics, however, is threatened by misspecification and weak identification, both of which are prevalent as we document in this paper. We show that failure to account for both misspecification and weak identification could easily lead to erroneous conclusions. To remedy these problems, we suggest that the J-statistic for misspecification, the IS-statistic for identification strength, and the DRLM test for risk premia become part of a toolkit that helps to provide trustworthy diagnosis and inference in future studies.

References

[1] Adrian, T., E. Etula and T. Muir. Financial Intermediaries and the Cross-Section of Asset Returns. *Journal of Finance*, **69**:2557–2596, 2014.

[2] Cochrane, J. H. Presidential address: Discount rates. *Journal of Finance*, 66(4):1047–1108, 2011.

[3] Cragg, J.C. and S.G. Donald. Inferring the rank of a matrix. *Journal of Econometrics*, **76**:223–250, 1997.

[4] Dufour, J.-M. Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models. *Econometrica*, **65**:1365–388, 1997.

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

[6] Fama, E.F. and K.R. French. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, **33**:3–56, 1993.

[7] Fama, E.F. and K.R. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, 2015.

[8] Feng, G. and Giglio, S. and Xiu, D. Taming the factor zoo: A test of new factors. *Journal of Finance*, 75(3):1327–1370, 2020.

[9] Gospodinov, N. R. Kan and C. Robotti. Misspecification-Robust Inference in Linear Asset-Pricing Models with Irrelevant Factors. *Review of Financial Studies*, **27**:2139–2170, 2014.

[10] Hansen, B. E. and Lee, S. Inference for iterated gmm under misspecification. *Econometrica*, 89(3):1419–1447, 2021.

[11] Hansen, L.P. Large Sample Properties of Generalized Method Moments Estimators. *Econometrica*, **50**:1029–1054, 1982.

[12] Hansen, L.P., J. Heaton and A. Yaron. Finite Sample Properties of Some Alternative GMM Estimators. *Journal of Business and Economic Statistics*, **14**:262–280, 1996.

[13] Harvey, C. R. and Liu, Y. and Zhu, H. ... and the cross-section of expected returns. *Review of Financial Studies*, 29(1):5–68, 2016.

[14] He, Z., B. Kelly and A. Manela. Intermediary Asset Pricing: New Evidence from Many Asset Classes. *Journal of Financial Economics*, **126**:1–35, 2017.

[15] Jagannathan, R. and Z. Wang. The Conditional CAPM and the Cross-Section of Expected Returns. *Journal of Finance*, **51**:3–53, 1996.

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

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

[18] Kandel, S. and R.F. Stambaugh. Portfolio Inefficiency and the Cross-section of Expected Returns. *Journal of Finance*, **50**:157–184, 1995.

[19] Kleibergen, F. Testing Parameters in GMM without assuming that they are identified. *Econometrica*, **73**:1103–1124, 2005.

[20] Kleibergen, F. Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics*, **139**:181–216, 2007.

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

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

[23] Kleibergen F. and Z. Zhan. Unexplained factors and their effects on second pass r-squared's. *Journal of Econometrics*, **189**:101–116, 2015.

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

[25] Kleibergen F. and Z. Zhan. Double Robust Inference for Continuous Updating GMM. Working paper, University of Amsterdam, 2021.

[26] Kleibergen, F., L. Kong and Z. Zhan. Identification Robust Testing of Risk Premia in Finite Samples. Working paper, University of Amsterdam, 2020.

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

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

[29] Lettau, M., S. Ludvigson and S. Ma. Capital Share Risk in U.S. Asset Pricing. *Journal of Finance*, **74**:1753–1792, 2019.

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

[31] Robin, J.-M. and R.J. Smith. Tests of Rank. *Econometric Theory*, **16**:151–175, 2000.

[32] Savov, A. Asset Pricing with Garbage. *Journal of Finance*, **72**:47–98, 2011.

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

[34] Staiger, D. and J.H. Stock. Instrumental Variables Regression with Weak Instruments. *Econometrica*, **65**:557–586, 1997.

[35] Stock, J.H. and J.H. Wright. GMM with Weak Identification. *Econometrica*, **68**:1055–1096, 2000.

[36] Yogo, M. A consumption-based explanation of expected stock returns. *Journal of Finance*, **61**:539–580, 2006.

Appendix

Proof of Theorem 1: Consider repackaging the assets to a new set of N^* assets by an invertible $N \times N$ weight matrix A :

$$R_t^* = AR_t.$$

The pseudo-true value resulting from the FM two-pass population objective function is:

$$\lambda_{F,FM}^* = \left(\beta^{*\prime}\beta^*\right)^{-1}\beta^{*\prime}\mu_{R^*} = \left(\beta' A' A\beta\right)^{-1}\beta' A' A\mu_R,$$

with $\beta^* = A\beta$, $\mu_{R^*} = A\mu_R$, which is not equal to the pseudo-true value resulting from the orginal set of assets unless A is orthogonal. Since an orthogonal matrix A does not lead to a set of portfolios, the FM pseudo-true value is not invariant under repackaging. Under correct specification, $\mu_R = \beta\lambda_F$, so the risk premium is invariant to repackaging.

The pseudo-true value resulting from the CUE objective function:

$$\begin{aligned}
\lambda_{F,CUE}^* = &\ \arg\min_{\lambda_F}(\mu_R - \beta\lambda_F)' \left[\mathrm{var}(\hat{\mu}_R - \hat{\beta}\lambda_F)\right]^{-1}(\mu_R - \beta\lambda_F) \\
= &\ \arg\min_{\lambda_F}(A\mu_R - A\beta\lambda_F)' \left[\mathrm{var}(A\hat{\mu}_R - A\hat{\beta}\lambda_F)\right]^{-1}(A\mu_R - A\beta\lambda_F) \\
= &\ \arg\min_{\lambda_F}(\mu_{R^*} - \beta^*\lambda_F)' \left[\mathrm{var}(\hat{\mu}_{R^*} - \hat{\beta}^*\lambda_F)\right]^{-1}(\mu_{R^*} - \beta^*\lambda_F),
\end{aligned}$$

is clearly invariant to repackaging by an invertible matrix A, but not to repackaging by an $N^* \times N$ matrix with N^* smaller than N.

Proof of Theorem 2: The FM two-pass estimator is given by:

$$\hat{\lambda}_F = (\hat{\beta}'\hat{\beta})^{-1}\hat{\beta}'\hat{\mu}_R.$$

We characterize the limit behavior of the FM two-pass estimator for the one factor setting, so $\hat{\lambda}_F = \frac{\hat{\beta}'\hat{\mu}_R}{\hat{\beta}'\hat{\beta}}$. It results from the joint limit behavior of its two different elements:

$$\sqrt{T}\begin{pmatrix} \hat{\mu}_R - \mu_R \\ \hat{\beta} - \beta \end{pmatrix} \underset{d}{\to} \begin{pmatrix} \psi_\mu \\ \psi_\beta \end{pmatrix},$$

with $\psi_\mu \sim N(0, \Omega + \beta Q \beta')$ and $\psi_\beta \sim N(0, \Omega Q^{-1})$ independently distributed which corresponds with Shanken (1992, Lemma 1). To focus on a setting where both the misspecification and betas are small and perhaps just borderline significant, we use the weak factor/small β and misspecification assumption (16):

$$\beta = \beta_T = \tfrac{b}{\sqrt{T}},\ \mu_R - \beta\lambda_F = \tfrac{a}{\sqrt{T}},\ \lambda_F^* = \lambda_F + (b'b)^{-1}b'a,$$

with b and a N-dimensional vectors of constants. Under the small misspecification and β assumption, the limit behavior of the least squares estimator $\hat{\beta}$ and $\hat{\mu}_R$ are characterized by:

$$\sqrt{T}\hat{\beta} \underset{d}{\to} b + \psi_\beta,\ \sqrt{T}\hat{\mu}_R \underset{d}{\to} b\lambda_F + a + \psi_\mu,$$

which we use to characterize the behavior of the FM risk premia estimator

$$\begin{aligned}\hat{\lambda}_F = &\ \tfrac{\hat{\mu}_R'\hat{\beta}}{\hat{\beta}'\hat{\beta}} = \tfrac{(\hat{\mu}_R - \hat{\beta}\lambda_F^* + \hat{\beta}\lambda_F^*)'\hat{\beta}}{\hat{\beta}'\hat{\beta}} = \lambda_F^* + \tfrac{(\hat{\mu}_R - \hat{\beta}\lambda_F^*)'\hat{\beta}}{\hat{\beta}'\hat{\beta}} \\ = &\ \lambda_F^* + \tfrac{[\hat{\mu}_R - \mu_R + (\mu_R - \beta\lambda_F^*) - (\hat{\beta} - \beta)\lambda_F^*]'\hat{\beta}}{\hat{\beta}'\hat{\beta}}\end{aligned}$$

with $\lambda_F^* = \frac{\beta'\mu_R}{\beta'\beta} = \lambda_F + (b'b)^{-1}b'a$, the pseudo-true value, so for small values of the betas and misspecification:

$$\begin{aligned}\hat{\lambda}_F = &\ \lambda_F^* + \tfrac{\left[\sqrt{T}(\hat{\mu} - \mu_R) + \sqrt{T}(\mu_R - \beta\lambda_F^*) - \sqrt{T}(\hat{\beta} - \beta)\lambda_F^*\right]'(\sqrt{T}\hat{\beta})}{(\sqrt{T}\hat{\beta})'(\sqrt{T}\hat{\beta})} \\ \underset{d}{\to} &\ \lambda_F^* + \tfrac{[\psi_\mu + b\lambda_F + a - b(\lambda_F + (b'b)^{-1}b'a) - \psi_\beta\lambda_F^*]'(b + \psi_\beta)}{(b+\psi_\beta)'(b+\psi_\beta)} \\ = &\ \lambda_F^* + \tfrac{(\psi_\mu + e - \psi_\beta\lambda_F^*)'(b+\psi_\beta)}{(b+\psi_\beta)'(b+\psi_\beta)} \\ = &\ \lambda_F^*\left(1 - \tfrac{\psi_\beta'(b+\psi_\beta)}{(b+\psi_\beta)'(b+\psi_\beta)}\right) + \tfrac{\psi_\mu'(b+\psi_\beta)}{(b+\psi_\beta)'(b+\psi_\beta)} + \tfrac{e'(b+\psi_\beta)}{(b+\psi_\beta)'(b+\psi_\beta)},\end{aligned}$$

with $e = a - b(b'b)^{-1}b'a$, which shows that the limit behavior of the FM two-pass estimator consists of four components.

The DRLM test: Kleibergen and Zhan (2021) propose the double robust Lagrange multiplier (DRLM) statistic for testing hypotheses on the pseudo-true value of the CUE. They show that under the hypothesis of interest, $\mathrm{H}_0 : \lambda_{F,CUE}^* = \lambda_{F,CUE,0}^*$, the limiting distribution of the DRLM statistic is bounded by a χ_K^2 distribution for general values of the identifica-

tion and misspecification strengths under weak conditions. The DRLM statistic involves a recentered estimator of the β's that depends on the hypothesized value of the pseudo-true value, which we indicate by the K-dimensional vector l :

$$\begin{aligned}\hat{D}(l) = & -\hat{\beta} - \Big[\hat{V}_{\beta_1(\mu_R-\beta l)}(l)\hat{V}_{(\mu_R-\beta l)}(l)^{-1}(\hat{\mu}_R - \hat{\beta}l)\ldots \\ & \qquad \hat{V}_{\beta_K(\mu_R-\beta l)}(l)\hat{V}_{(\mu_R-\beta l)}(l)^{-1}(\hat{\mu}_R - \hat{\beta}l)\Big] \\ = & -\hat{\beta} - (\bar{R} - \hat{\beta}l)(1 + l'\hat{Q}_{\bar{F}\bar{F}}^{-1}l)^{-1}l'\hat{Q}_{\bar{F}\bar{F}}^{-1} \\ = & -\tfrac{1}{T}\textstyle\sum_{t=1}^{T} R_t(\bar{F}_t + l)'\left[\tfrac{1}{T}\sum_{t=1}^{T}(\bar{F}_t + l)(\bar{F}_t + l)'\right]^{-1},\end{aligned}$$

where $\hat{V}_{\beta_i(\mu_R-\beta l)}$ is the estimator of the covariance between $\hat{\beta}_i$ and $\hat{\mu}_R - \hat{\beta}l$, for $i = 1, \ldots, K$, $\hat{\beta} = (\hat{\beta}_1 \ldots \hat{\beta}_K)$ and $\hat{V}_{(\mu_R-\beta\lambda_f)}(l)$ is the covariance matrix estimator of the sample pricing error $\hat{\mu}_R - \hat{\beta}l$. The identical expressions on the last two lines are for a setting of i.i.d. errors with $\hat{Q}_{\bar{F}\bar{F}}$ the estimator of the covariance matrix of the factors, $\hat{Q}_{\bar{F}\bar{F}} = \frac{1}{T}\sum_{t=1}^{T}\bar{F}_t\bar{F}_t'$, $\bar{F}_t = F_t - \bar{F}$, $\bar{F} = \frac{1}{T}\sum_{t=1}^{T} F_t$. The DRLM statistic for testing $\mathrm{H}_0 : \lambda^*_{F,CUE} = l$ then reads:

$$\begin{aligned}DRLM(l) = & \; T \times (\hat{\mu}_R - \hat{\beta}l)'\hat{V}_{(\mu_R-\beta l)}(l)^{-1}\hat{D}(l)\Big[\hat{D}(l)'\hat{V}_{(\mu_R-\beta l)}(l)^{-1}\hat{D}(l) + \\ & \Big(I_N \otimes \hat{V}_{(\mu_R-\beta l)}(l)^{-1}(\hat{\mu}_R - \hat{\beta}l)\Big)'\hat{V}_{\hat{D}(l)}(l)\Big(I_N \otimes \hat{V}_{(\mu_R-\beta l)}(l)^{-1}(\hat{\mu}_R - \hat{\beta}l)\Big)\Big]^{-1} \\ & \hat{D}(l)'\hat{V}_{(\mu_R-\beta l)}(l)^{-1}(\hat{\mu}_R - \hat{\beta}l),\end{aligned}$$

where $\hat{V}_{\hat{D}(l)}$ is the estimator of the covariance matrix of $\mathrm{vec}(\hat{D}(l))$. For the i.i.d. setting, it simplifies to:

$$DRLM(l) = \; \hat{\mu}(l)^{*\prime}\hat{D}(l)^*\left[\hat{\mu}(l)^{*\prime}\hat{\mu}(l)^* I_N + \hat{D}(l)^{*\prime}\hat{D}(l)^*\right]^{-1}\hat{D}(l)^{*\prime}\hat{\mu}(l)^*,$$

with $\hat{\mu}(l)^* = \sqrt{T}\hat{\Omega}^{-\frac{1}{2}}(\bar{R} - \hat{\beta}l)(1 + l'\hat{Q}_{\bar{F}\bar{F}}^{-1}l)^{-\frac{1}{2}}$, and $\hat{D}(l)^* = \sqrt{T}\hat{\Omega}^{-\frac{1}{2}}\hat{D}(l)(\hat{Q}_{\bar{F}\bar{F}} + ll')^{\frac{1}{2}}$.

The $100\times(1-\alpha)\%$ confidence set for $\lambda^*_{F,CUE}$ (denoted by $\mathrm{CS}_{\lambda^*_{F,CUE}}(\alpha)$ below) that results from the DRLM test consists of all values of l for which the DRLM test does not reject using the $100 \times (1-\alpha)\%$ critical value that results from the χ^2_K-distribution:

$$\mathrm{CS}_{\lambda^*_{F,CUE}}(\alpha) = \left\{l : \mathrm{DRLM}(l) \leq \chi^2_K(\alpha)\right\},$$

where $\chi^2_K(\alpha)$ is the upper α-th quantile of the χ^2_K distribution. DRLM(l) is not a quadratic function of l, so it cannot directly be inverted to obtain the confidence set. The confidence set does therefore not have the usual expression of an estimator plus or minus a multiple of the standard error. Instead, we have to specify a K-dimensional grid of values for l, and compute the DRLM statistic for each value of l on the K-dimensional grid to determine if it does not exceed the appropriate critical value so l is part of the confidence set.

The DRLM statistic is a quadratic form of the derivative of the sample CUE objective function with respect to l. It is therefore equal to zero at all stationary points of the sample CUE objective function. Since the DRLM statistic is not a quadratic function of l, there can be multiple stationary points where it equals zero. This affects the discriminatory power of the DRLM test. Kleibergen and Zhan (2021) therefore propose a power improvement rule which rejects values of l at the α% significance level, alongside values of l where the DRLM statistic is significant at the α% level also, when there are significant values of the DRLM statistic on every line going from the hypothesized value to the CUE. Kleibergen and Zhan (2021) show that the power improvement rule does not affect the size of the DRLM test and improves power considerably.

Since the DRLM test is size correct, the coverage of a $100 \times (1-\alpha)$% confidence set is at least $100 \times (1-\alpha)$%. By projecting these confidence sets on the K different axes, we obtain $100 \times (1-\alpha)$% univariate confidence sets for the individual risk premium whose coverage is also at least $100 \times (1-\alpha)$%. When we plug in estimators for some of the risk premia, the coverage of the resulting confidence sets is not guaranteed nor is the size of the resulting subset DRLM test.

When using the power improvement rule, the confidence sets resulting from the DRLM test can have two distinct shapes.

1. Bounded and convex: there is a closed compact set of values of l for which the DRLM statistic does not exceed the critical value.

2. Unbounded: this occurs either when there are no values of l for which the DRLM statistic exceeds the critical value (unbounded and convex), or when there is a bounded set of values of l for which the DRLM statistic exceeds the critical value (unbounded

and disjoint).

Bounded and convex confidence sets occur when the risk premia are well identified. Unbounded confidence sets are indicative of identification failure. Dufour (1997, Theorems 3.3 and 3.6) formally proves that a size correct test on a parameter which is potentially not identified must have a positive probability of producing an unbounded 95% confidence set. Conversely, any test procedure, such as the FM t-test, that cannot generate an unbounded 95% confidence set, cannot be a size correct test procedure when the tested parameter can be non-identified.

J and IS statistics:

1. If $\lambda_0 = 0$ is not imposed: consider $\mathcal{R}_t = (\mathcal{R}_{1,t} \dots \mathcal{R}_{N+1,t})'$: $(N+1)\times 1$ vector of returns; F_t: $K \times 1$ vector of risk factors, $t = 1, ..., T$. By subtracting the $(N+1)$-th asset return, we obtain the $N \times 1$ column vector R_t:

$$R_t = (\mathcal{R}_{1,t} \dots \mathcal{R}_{N,t})' - \iota_N \mathcal{R}_{N+1,t}.$$

2. If $\lambda_0 = 0$ is imposed: consider R_t as the observed $N \times 1$ vector of returns, and F_t as the $K \times 1$ vector of risk factors.

Estimation of the auxiliary linear factor model $R_t = \alpha + \beta F_t + u_t$ yields

$$\hat{\beta} = \sum_{t=1}^{T} \bar{R}_t \bar{F}_t' \left(\sum_{t=1}^{T} \bar{F}_t \bar{F}_t' \right)^{-1}, \ \hat{\Omega} = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t \hat{u}_t', \quad \text{and} \quad \hat{Q}_{FF} = \frac{1}{T} \sum_{t=1}^{T} \bar{F}_t \bar{F}_t',$$

where $\bar{F}_t = F_t - \bar{F}$, $\bar{F} = \frac{1}{T}\sum_{t=1}^{T} F_t$, $\bar{R}_t = R_t - \bar{R}$, $\bar{R} = \frac{1}{T}\sum_{t=1}^{T} R_t$, and $\hat{u}_t = (R_t - \bar{R}) - \hat{\beta}(F_t - \bar{F})$ is the residual at time t.

Let rk be the smallest root of

$$\left| \mu \hat{Q}_{FF}^{-1} - \hat{\beta}' \hat{\Omega}^{-1} \hat{\beta} \right| = 0,$$

which is identical to the smallest eigenvalue of the matrix $\hat{Q}_{FF} \hat{\beta}' \hat{\Omega}^{-1} \hat{\beta}$. The IS-statistic for

$\mathrm{H}_0 : \mathrm{rank}(\beta) = K - 1$, reads:

$$IS = T \times \mathrm{rk} \preceq \chi^2_{N-K+1}.$$

Let "Eigen$_{min}$" be the smallest root of

$$\left| \mu \begin{pmatrix} 1 & 0 \\ 0 & \hat{Q}_{FF}^{-1} \end{pmatrix} - \begin{pmatrix} \bar{R} & \hat{\beta} \end{pmatrix}' \hat{\Omega}^{-1} \begin{pmatrix} \bar{R} & \hat{\beta} \end{pmatrix} \right| = 0,$$

which is equal to the smallest eigenvalue of $\begin{pmatrix} 1 & 0 \\ 0 & \hat{Q}_{FF} \end{pmatrix} \begin{pmatrix} \bar{R} & \hat{\beta} \end{pmatrix}' \hat{\Omega}^{-1} \begin{pmatrix} \bar{R} & \hat{\beta} \end{pmatrix}$. The misspecification J-statistic for testing $\mathrm{H}_0 : E(R_t) = \beta\lambda_F$ reads:

$$J = T \times \mathrm{Eigen}_{min} \xrightarrow{d} \chi^2_{N-K}.$$

DRLM for the Fama and French (2015) five-factor model: See Table 6.

Table 6: **Application to the Fama and French (2015) five-factor model**

The five factors (R_m, SMB, HML, RMW, CMA) of Fama and French (2015) are downloaded from French's website, together with the 25 Fama-French portfolios from July 1963 - Nov 2021 as the test assets. The zero-β return, $\lambda_0 = 0$ restriction is imposed for the DRLM test. The resulting J-statistic is 68.16, and IS-statistic is 113.05.

	R_m	SMB	HML	RMW	CMA
mean	0.59	0.23	0.27	0.27	0.26
s.e.	0.17	0.11	0.11	0.08	0.07
Comparison of 95% C.I. of λ_F					
(I) C.I. by mean $\pm$ 1.96 s.e.	(0.26, 0.92)	(0.00, 0.45)	(0.05, 0.48)	(0.11, 0.43)	(0.12, 0.41)
(II) C.I. by DRLM	(0.55, 0.69)	(0.15, 0.31)	(0.15, 0.32)	(0.01, 0.31)	(0.15, 1.29)
Ratio of the C.I. widths, $\frac{\text{(II)}}{\text{(I)}}$	21%	36%	40%	94%	393%