UvA-DARE (Digital Academic Repository)

Robust Inference for Consumption-Based Asset Pricing

Kleibergen, F.; Zhan, Z.

DOI

[10.1111/jofi.12855](https://doi.org/10.1111/jofi.12855)

Publication date

2020

Document Version

Final published version

Published in

The Journal of Finance

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Kleibergen, F., & Zhan, Z. (2020). Robust Inference for Consumption-Based Asset Pricing. *The Journal of Finance*, 75(1), 507-550. <https://doi.org/10.1111/jofi.12855>

General rights

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

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Robust Inference for Consumption-Based Asset Pricing

FRANK KLEIBERGEN and ZHAOGUO ZHAN*

ABSTRACT

The reliability of traditional asset pricing tests depends on: (i) the correlations between asset returns and factors; (ii) the time series sample size T compared to the number of assets N . For macro-risk factors, like consumption growth, (i) and (ii) are often such that traditional tests cannot be trusted. We extend the Gibbons-Ross-Shanken statistic to test identification of risk premia and construct their 95% confidence sets. These sets are wide or unbounded when T and N are close, but show that average returns are not fully spanned by betas when T exceeds N considerably. Our findings indicate when meaningful empirical inference is feasible.

CLASSICAL CONSUMPTION-BASED ASSET PRICING THEORY relates asset returns to consumption risk. Yet a worrisome phenomenon is that different measures of consumption lead to different empirical findings. It has been well documented that the canonical consumption measure from the National Income and Product Accounts (NIPA) leads to small correlations between consumption growth and asset returns that could be improved by adopting alternative consumption measures, including the three-year consumption measure in Parker and Julliard (2005), the fourth-quarter to fourth-quarter consumption measure in Jagannathan and Wang (2007), the garbage measure in Savov (2011), and the unfiltered NIPA consumption measure in Kroencke (2017).

The credibility of these consumption measures for asset pricing is commonly tested using two methodologies: (i) the two-pass regression for linear asset pricing models, where expected asset returns are expressed by the beta representation (Fama and MacBeth (FM, 1973)), and (ii) the generalized method of moments (GMM) of Hansen (1982) for nonlinear asset pricing models in a stochastic discount factor (SDF) representation. Empirical findings based on

*Frank Kleiberger is at the University of Amsterdam. Zhaoguo Zhan is at Kennesaw State University. We thank the Editor, Stefan Nagel, the Associate Editor, and two anonymous referees for their comments that substantially improved the paper. Our thanks for helpful discussions also go to Lucy F. Ackert, Tomislav Ladika, and participants at the seminars as well as conferences where the paper has been presented. We thank Alexi Savov, Tim A. Kroencke, Martin Lettau, and Sydney Ludvigson for sharing their data. We have read *The Journal of Finance* disclosure policy and have no conflicts of interest to disclose.

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1111/jofi.12855

© 2019 The Authors. *The Journal of Finance* published by the American Finance Association

these two standard methodologies appear to favor some consumption measures for asset pricing over others (Kroencke (2017)).

It is well known, however, that the reliability of asset pricing tests in both the FM and the GMM settings depends on the statistical quality of consumption measures, or more generally, risk factors. For instance, Kan and Zhang (1999a) and Kleibergen (2009) warn that the t -test in the FM two-pass procedure can spuriously favor risk factors that are independent of or weakly correlated with asset returns, respectively. Similar spurious outcomes arise in the GMM setting (Stock and Wright (2000), Kleibergen (2005), Gospodinov, Kan, and Robotti (2017)). Moreover, Kan and Zhang (1999b), Lewellen, Nagel, and Shanken (2010), and Kleibergen and Zhan (2015, 2018) provide intuitive explanations for why poor quality risk factors can induce seemingly promising empirical outcomes. In light of this literature, it is important not to misinterpret potentially spurious outcomes as evidence in support of factor pricing.

In addition to the statistical quality of various consumption measures, limited sample sizes impose challenges for empirical asset pricing tests. For estimation results of asset pricing models to be representative of the span of the market, existing studies typically use a considerable number of test assets. At the same time, macro-risk factors such as consumption growth are commonly measured at annual or quarterly frequencies, in which case the number of observations is often not much larger than the number of test assets. The limited number of time series observations T compared to the number of cross-section observations N leads to large estimation error of the covariance matrix of the test assets. This “limited T versus large N ” problem is not accounted for in standard asymptotic approximations of t and Wald statistics (Bekker (1994), Newey and Windmeijer (2009)) and thus provides a second argument for why researchers cannot rely on standard t /Wald-tests to conduct dependable statistical inference.

The credibility of traditional asset pricing tests on the parameters of interest, such as risk premia, in FM and GMM thus depends on (i) the strength of identification, which is reflected by the correlations between risk factors and returns on test assets, and (ii) the number of time series observations of the risk factors relative to the number of test assets, that is, the “limited T versus large N ” problem. For asset pricing models that involve macroeconomic factors, both of these issues can threaten the reliability of standard asset pricing tests.

In this paper, we propose two straightforward asset pricing tests that, unlike traditional tests, are valid for all possible strengths of identification of the risk premia and for scenarios in which the time series sample size exceeds the number of test assets. Our tests are simple extensions of the Gibbons, Ross, and Shanken (GRS, 1989) test, which tests the joint significance of the so-called alphas, or constant terms, in a set of regressions of test asset returns on a constant and risk factors. The first of our proposed statistics tests the identification of the risk premia, while the second tests for specific values of the risk premia to allow for the construction of confidence sets.

In the linear beta representation of expected asset returns, risk premia are identified through the variation of the betas over the cross-section of test assets.

Our first proposed identification statistic therefore examines whether there is sufficient variation in the betas to identify risk premia. In the case of a single risk factor, the risk premium is not identified in the beta representation that includes a zero-beta return if the betas do not vary over the cross-section, so our identification test statistic simply tests whether all betas are identical. The GRS statistic that we use for this purpose allows the number of cross-section and time series observations to be close and has an F -distribution in finite samples, as in GRS.

We apply our identification test to data from Kroencke (2017) to investigate the identification of the risk premia on the five consumption measures mentioned above (i.e., NIPA, Parker and Julliard (2005), Jagannathan and Wang (2007), Savov (2011), and Kroencke (2017)). At the 5% significance level, our identification test cannot reject the possibility that the betas of these five consumption measures are constant. Consequently, despite the significance of the betas for some consumption measures, we cannot reject that the risk premia are unidentified under these consumption measures. Simulation studies calibrated to these data further show that the FM t -test on risk premia is indeed unreliable for each of the five consumption growth series. The reason is that when the betas are almost constant, they basically become proportional to the constant term in the second pass of the FM two-pass procedure. The resulting near multicollinearity violates the assumptions needed for the asymptotic validity of the FM t -test and explains why its rejection frequencies do not accord with conventional asymptotics in the simulation experiment.

Unlike the FM t -test, whose validity is subject to the questionable asymptotic normal approximation of the risk premia estimator, our second proposed test statistic does not involve a risk premia estimator but rather considers all restrictions imposed by factor pricing at the hypothesized value of the risk premia. It does so by noting that factor pricing implies zero alphas in an appropriately specified set of regressions that use the prespecified value of the risk premia. We test the joint significance of all of these alphas using the GRS test, which for this purpose, we refer to as the GRS-Factor Anderson-Rubin (GRS-FAR) statistic to indicate that it also provides an extension of the well-known Anderson-Rubin statistic in instrumental variables regression (Anderson and Rubin (1949), Kleibergen (2009)). Since it is just a GRS test, the GRS-FAR statistic has an exact F -distribution so its reliability does not depend on either the identification of the risk premia or the number of time series observations compared to the number of test assets.

We report confidence sets that consist of the hypothesized values of the risk premia for which the GRS-FAR statistic is insignificant at the appropriate significance level. By varying the value of the risk premia used in the GRS-FAR statistic, we can map out the confidence set, but we can also compute it with a closed-form expression (Dufour and Taamouti (2005)). Using the data from Kroencke (2017), we find that the 95% confidence sets for the risk premia do not exclude any pre-set value for all five consumption measures. Our 95% confidence sets are thus unbounded. This might seem odd but it should come

out naturally because we cannot reject at the 5% significance level that the risk premia are unidentified for all five consumption growth measures. When a risk premium is not identified, it can take any real value. For comparison, we also use the market return as a risk factor. The hypothesis of no identification of the market return's risk premium is just barely rejected at the 5% level. This minor rejection leads to a very wide 95% confidence set for the risk premium on the market return.

Next, we analyze the extent to which identification of the risk premium improves when we remove the zero-beta return, that is, when we exclude the intercept in the second pass of the FM two-pass procedure. The risk premium is now identified when the betas differ from zero, so the betas can be constant, which is not allowed for when the zero-beta return is included. Except for the garbage consumption measure, the 95% confidence sets of the risk premia on all consumption growth measures remain unbounded. The length of the 95% confidence set for the risk premium on the garbage consumption measure is considerably lower, but it remains rather wide and includes zero. The 95% confidence set for the risk premium on the market return, however, shrinks sufficiently that it becomes comparable to the 95% confidence set of the mean of the market return. Taken together, the results show that precise statements on risk premia are not likely to be made for data sets whose time series and cross-section dimensions are close unless we introduce stronger identification conditions or restrictions, such as no zero-beta return. In the Kroencke data that we study, the cross-section sample size of 31 is just too close to the time series sample size of 55. The number of elements to be estimated in the covariance matrix of returns is thus too large given the limited number of observations in the time series.

Our proposed identification and risk premia tests also apply to multifactor settings. To illustrate, we use our GRS-based tests on the data from Lettau and Ludvigson (2001) for which the time series dimension, 141, is considerably larger than the cross-section sample size, 25. Using the three Fama and French (FF, 1993) factors, we now obtain an empty 95% confidence set for their risk premia. This implies that the alphas are significant for each value of the risk premia, that is, the average asset returns are not fully spanned by the betas of the three Fama-French factors, and thus there is misspecification. However, when we use consumption-based risk factors, for example, when we jointly use consumption growth, the (lagged) consumption-wealth ratio, and their interaction, the 95% confidence sets for their risk premia do not exclude any value and hence are all unbounded.

To align these seemingly contradictory findings, we conduct a simulation experiment calibrated to real data. We increase the number of time series observations while keeping the cross-sectional sample size fixed and examine how often consumption-based risk premia have empty, unbounded, or bounded confidence sets. This experiment shows that when the number of time series observations is small, we find mostly unbounded 95% confidence sets that become bounded and eventually empty as the number of time series observations increases. Because the consumption-based factors are less correlated with the

returns on the test assets, a larger number of time series observations are needed to render the confidence sets mostly bounded or empty compared to when we use the FF factors. Empty confidence sets are indicative of misspecification, so a data-generating process (DGP) that does not accord with factor pricing can generate all of our test results for the Kroencke (2017) and Lettau and Ludvigson (2001) data.

We further extend our tests to nonlinear asset pricing models that are analyzed using their SDFs. The same identification and time series versus cross-section sample size issues that arise when considering linear SDFs also hold in nonlinear SDFs estimated using GMM. We therefore use the extension of the GRS-FAR test for nonlinear settings, namely, the GMM-Anderson-Rubin (GMM-AR) statistic proposed by Stock and Wright (2000). The asymptotic distribution of this statistic is insensitive to the identification of the parameters of the SDF but requires a much larger time series than cross-section sample size. For the Kroencke (2017) data with a time series dimension equal to 55, this implies that we can use just a few test assets. When we do so using a constant relative risk aversion (CRRA) SDF, we find unbounded 95% confidence sets for the relative risk aversion parameter.

Overall, the asset pricing tests proposed in this paper are robust in the sense that they remain reliable when the risk factors are not of satisfactory quality and the time series length is not large. These robust tests are therefore particularly helpful in gauging consumption-based asset pricing models, for which the traditional tests lead to conflicting empirical findings. The unbounded confidence sets resulting from the robust tests help reconcile the conflicting findings in the existing literature. The unboundedness can be attributed to both the unsatisfactory quality and limited length of the consumption series proposed for asset pricing. To achieve the favorable bounded and tight confidence sets, we therefore need consumption measures of better statistical quality that have longer time series, or we need to impose further restrictions such as no zero-beta returns, while the asset pricing models also need to be correctly specified. We present empirical and numerical findings to further illustrate these points.

The rest of the paper is organized as follows. In Section I, we describe the identification condition for the risk premia in linear asset pricing models and propose our identification and risk premia tests for the model with a single risk factor. Section II extends our analysis to linear multifactor models. Section III discusses nonlinear asset pricing models that are analyzed with GMM using their SDFs. Finally, Section IV concludes. Technical details and additional material are relegated to the Internet Appendix.¹

¹ The Internet Appendix is available in the online version of this article on *The Journal of Finance* website.

I. Risk Premia from Linear Factor Models

A. Identification of Risk Premia

Premia on risk factors, such as consumption growth or the FF factors, are commonly estimated using the beta representation of asset returns (see, e.g., Breeden, Gibbons, and Litzenberger (1989)):

$$E(R_{i,t}) = \lambda_0 + \beta'_i \lambda_f, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (1)$$

where λ_0 is the zero-beta return, λ_f is the k -dimensional vector of risk premia, β_i is a k -dimensional vector that results from the regression of returns of the i^{th} asset on the k risk factors, and $R_{i,t}$ is the return on the i^{th} asset at time t .

The $N \times k$ -dimensional matrix β of the individual assets' betas,

$$\beta = \begin{pmatrix} \beta'_1 \\ \vdots \\ \beta'_N \end{pmatrix}, \quad (2)$$

results from the covariances between the asset returns and risk factors,

$$\beta = V_{Rf} V_{ff}^{-1}, \quad (3)$$

where V_{Rf} is the $N \times k$ -dimensional covariance matrix of asset returns and risk factors and V_{ff} is the $k \times k$ -dimensional covariance matrix of risk factors, so the β matrix is generally identified unless, for example, there are multicollinearity issues between the risk factors. The moment equation (1) identifies λ_0 and λ_f using the variation in the betas over the cross-section of asset returns.

To further articulate the identification of the risk premia, we assume that the moment equation also applies to the returns on any repackaged portfolio of assets, which can be characterized by an N -dimensional weight vector w whose elements add up to one, $w' \iota_N = 1$, with ι_N the N -dimensional vector of ones. The return on the repackaged portfolio is

$$r_t = w' R_t, \quad (4)$$

with $R_t = (R_{1,t} \dots R_{N,t})'$ the combined N -dimensional vector of asset returns at time t . Since

$$E(R_t) = \iota_N \lambda_0 + \beta \lambda_f, \quad (5)$$

the expected return on the repackaged portfolio is

$$E(r_t) = w' \iota_N \lambda_0 + w' \beta \lambda_f = w' (\iota_N \vdots \beta) \begin{pmatrix} \lambda_0 \\ \lambda_f \end{pmatrix}. \quad (6)$$

We cannot identify λ_0 and λ_f if all of the expected returns on the repackaged portfolios depend on them in the same manner. Put differently, not all weight vectors w should lead to the same value of the row vector

$$w' (\iota_N \vdots \beta). \quad (7)$$

This motivates the so-called full rank condition on the $N \times (k+1)$ -dimensional matrix $(\iota_N : \beta)$. To show that the risk premia are not identified when this full rank condition does not hold, we consider the setting with a single risk factor, so $k = 1$. A lower rank value of $(\iota_N : \beta)$ then implies that all elements of β are identical and proportional to the vector of ones, that is, $\beta = \iota_N \cdot d$, with d a constant that can be zero, and

$$E(r_t) = w' \iota_N \lambda_0 + d w' \iota_N \lambda_f = \lambda_0 + d \lambda_f, \quad (8)$$

so the expected returns do not vary over λ_0 and λ_f for different repackaged portfolios. Hence, we cannot identify the risk premia from the cross-section of asset returns. When the number of risk factors exceeds one and the matrix $(\iota_N : \beta)$ has a lower rank value, a similar argument can be made to show that there is not enough variation in the cross-section of asset returns to identify all risk premia.²

A full rank value of the matrix $(\iota_N : \beta)$ implies that there is enough variation in the cross-section of asset returns to identify the risk premia. Examples of lower rank values of the matrix are settings in which the betas are zero or identical over the different assets. Both of these cases might be considered unlikely ex ante, but the observed values of the asset returns and risk factors in the data can be such that the resulting betas are close to zero or close to identical over the different assets in a statistical sense. This turns out to be the case for many of the risk factors put forward in the literature. Inference based on traditional estimation methods as in the FM two-pass procedure is then not reliable and easily leads to erroneous conclusions.

A.1. Full Rank Condition for Misspecified Models

It is worth noting that the full rank condition on $(\iota_N : \beta)$ is not limited to (1). When expected asset returns are not fully spanned by the betas of the specified risk factors, the moment equation (1) no longer holds, which leads to the modified version

$$E(R_{i,t}) = \lambda_0 + \beta_i' \lambda_f + \alpha_i, \quad (9)$$

where α_i reflects the asset-specific misspecification error. If $\alpha_i = 0$, (9) reduces to (1).

The parameters (λ_0, λ_f) in (9) are commonly defined as the coefficients that result from projecting $E(R_t)$ on $(\iota_N : \beta)$. Since (9) is just a linear regression, it assumes that the α_i s are uncorrelated with the β_i s and that their mean over

² When there are k risk factors and $(\iota_N : \beta)$ has a lower rank value, there exists a nonzero $(k+1)$ -dimensional vector g such that $(\iota_N : \beta)g = 0$. The expected asset returns of a repackaged portfolio are then $E(r_t) = w' E(R_t) = w' \iota_N \lambda_0 + w' \beta \lambda_f = w' (\iota_N : \beta) (g : G) \begin{pmatrix} \lambda_0 \\ \lambda_f \end{pmatrix} = w' (\iota_N : \beta) G \lambda_f^*$, with $\begin{pmatrix} \lambda_0 \\ \lambda_f \end{pmatrix} = (g : G) \begin{pmatrix} \lambda_0^* \\ \lambda_f^* \end{pmatrix}$ and G a $(k+1) \times k$ -dimensional matrix such that $(g : G)$ is invertible. The expected return of the repackaged portfolio does not depend on λ_0^* , so we cannot identify it from the cross-section. Since λ_0 and λ_f are functions of λ_0^* and λ_f^* , we cannot identify them either.

the cross-section equals zero. To avoid multicollinearity, the full rank condition on $(\iota_N : \beta)$ is still needed, however, to identify the risk premia in this regression. Put differently, irrespective of whether researchers use (1) or (9) to infer risk premia, a generic requirement is that the $(\iota_N : \beta)$ matrix has full rank value.

A.2. Full Rank Condition with No Zero-Beta Return

Another modification of (1) occurs when we use excess returns for R_{it} , so the zero-beta return λ_0 can be left out, or put differently, set to zero; see, for example, Savov (2011). The factor pricing moment condition then reads

$$E(R_{i,t}) = \beta_i' \lambda_f. \quad (10)$$

Without the zero-beta return, the identification condition for λ_f in (10) is that β is a full rank matrix. In the case of a single factor, the risk premium is then identified when β is unequal to zero, so identical β s for the different assets is allowed. In contrast, with the zero-beta return included in (1), identification requires a full rank value of $(\iota_N : \beta)$, so β is not allowed to be constant over the different assets. Removing the zero-beta return therefore improves the identification of the risk premia.

Our proposed identification test, defined below, examines the full rank conditions for (1), (9), and (10), which are needed to identify the risk premia in each of these specifications. Building on that test, we propose another test that enables us to construct confidence sets for the risk premia in (1) and (10) that, unlike confidence sets that result from traditional test procedures, remain valid when the full rank conditions are questionable. This test also indicates the existence of misspecification, as described in (9). Because of the similarity of the rank identification conditions for (1), (9), and (10), in our subsequent discussion we focus on the rank condition for (1) and just mention rank conditions for (9) and (10) in passing.

Before providing our tests, we first briefly discuss the commonly used FM two-pass procedure and several consumption measures, for which the rank value of $(\iota_N : \beta)$ turns out to be generally problematic, to highlight the issues.

B. FM Two-Pass Procedure

To estimate risk premia, Fama and MacBeth (1973) propose a two-pass procedure. A simplified version of this procedure, justified by Shanken (1992), can be described as follows.

- (1) Estimate β in the linear regression model

$$R_t = c + \beta f_t + u_t, \quad t = 1, \dots, T, \quad (11)$$

where f_t is the k -dimensional vector of risk factors, c is an N -dimensional vector containing the constant terms, and u_t is an N -dimensional vector of errors, using the least squares estimator

$$\hat{\beta} = \sum_{t=1}^T \bar{R}_t \bar{f}_t' \left(\sum_{t=1}^T \bar{f}_t \bar{f}_t' \right)^{-1}, \quad (12)$$

where $\bar{f}_t = f_t - \bar{f}$, $\bar{f} = \frac{1}{T} \sum_{t=1}^T f_t$, $\bar{R}_t = R_t - \bar{R}$, and $\bar{R} = \frac{1}{T} \sum_{t=1}^T R_t$.

- (2) Regress the asset returns at time t , R_t , on the N -dimensional vector of ones ι_N and the estimated β to obtain a time series of estimates of the zero-beta return and the risk premia:

$$\begin{pmatrix} \hat{\lambda}_{0,t} \\ \hat{\lambda}_{f,t} \end{pmatrix} = [\iota_N' : \hat{\beta}' (\iota_N' : \hat{\beta})]^{-1} (\iota_N' : \hat{\beta})' R_t. \quad (13)$$

Use the average value of the zero-beta return and risk premia as the estimator of λ_0 and λ_f .³

$$\begin{pmatrix} \hat{\lambda}_0 \\ \hat{\lambda}_f \end{pmatrix} = \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \hat{\lambda}_{0,t} \\ \hat{\lambda}_{f,t} \end{pmatrix}. \quad (14)$$

The full rank value assumption of $(\iota_N' : \beta)$ that identifies the risk premia is identical to the absence of any multicollinearity in the (population) second-pass regression in (13). The FM two-pass procedure uses an ordinary (unweighted) least squares regression in the second pass. Alternatively, in the second pass, we can conduct a generalized least squares (GLS) regression, which uses the (inverse of the residual) covariance matrix from the first-pass time series regression as a weight matrix. The resulting GLS risk premia estimators are invariant to the repackaging used previously to motivate the identification condition for the risk premia. The full rank condition is therefore straightforward for the GLS risk premia estimator, that is, when it does not hold, the (population) GLS risk premia estimator becomes infeasible for an appropriately repackaged set of portfolios.⁴ Since the GLS risk premia estimator is invariant to repackaging, it is then infeasible in general if the full rank condition fails.

Inference using the FM two-pass estimator is based on an approximation of its distribution by a normal one. This approximation is precise in large samples when paired with a full rank value of the matrix $(\iota_N' : \beta)$, so the risk premia are identified (Shanken (1992)). It can be considerably off, however, in other

³ The estimator in (14) also directly results from (13) when, instead of the returns over a specific time period, we use their average.

⁴ Consider N repackaged assets with $\mathbf{r}_t = W R_t$ the N -dimensional vector of asset returns and W an invertible $N \times N$ -dimensional matrix with $W \iota_N = \iota_N$. When $(\iota_N' : \beta)$ has a lower rank value, we can specify W such that the first column of $W(\iota_N' : \beta)$ equals the N -dimensional vector of zeros. Since the GLS risk premia estimator is invariant to the specification of W , it is then clear that $(\iota_N' : \beta)$ has to be of full rank for the (population) GLS risk premia estimator.

empirically relevant settings. The first authors to emphasize this were Kan and Zhang (1999a), who show that the normal approximation is incorrect in a useless factor setting where β equals zero and there is misspecification, so the moment equation (1) does not hold. When β equals zero, $(\iota_N : \beta)$ has a lower rank value, and thus one of the conditions for the validity of the normal approximation is violated. Kleibergen (2009) shows that the normal approximation is also off when $(\iota_N : \beta)$ is relatively close to having a reduced rank value. This so-called weak factor setting is empirically relevant and applies to many macro-risk factors proposed in the literature, such as the different consumption growth measures. To overcome the deficiencies of the FM two-pass estimator, Kleibergen (2009) proposes a number of test procedures that remain accurate when $(\iota_N : \beta)$ is close to or has a lower rank value. These test procedures become similar to tests using the FM two-pass estimator in large samples with a full rank value of $(\iota_N : \beta)$.

Inference using the FM two-pass estimator is sensitive not only to potential lower rank values of the matrix $(\iota_N : \beta)$ but also to the cross-section sample size N relative to the time series sample size T , that is, the “limited T versus large N ” problem. The larger N , the more elements β has, so the larger the estimation error it contributes to the FM two-pass estimator. The number of elements of the covariance matrix estimator increases even more rapidly, which further affects the accuracy of the standard errors of the FM two-pass estimator. We next show how identification of the risk premia can be tested using the rank value of $(\iota_N : \beta)$ while allowing N and T not to be too far apart, which is typical for many applications involving macroeconomic factors. We do so by using an extension of the GRS test. We also show how we can use the GRS statistic to test the risk premia and misspecification without imposing the rank condition. Before we do so, we briefly discuss the data that we use to emphasize the empirical relevance of our results.

We consider five consumption measures from the asset pricing literature: (i) the consumption expenditure on nondurable goods and services reported by NIPA (“Reported”), (ii) the three-year consumption measure of Parker and Julliard (2005) (“P-J”), (iii) the fourth-quarter to fourth-quarter consumption measure of Jagannathan and Wang (2007) (“Q4-Q4”), (iv) the garbage measure of Savov (2011) (“Garbage”), and (v) the unfiltered NIPA consumption measure of Kroencke (2017) (“Unfiltered”).

To facilitate comparison, we use the same data as in Savov (2011) and Kroencke (2017). The annual garbage series is from 1960 to 2007 as in Savov (2011), while the other four consumption measures are from 1960 to 2014 as in Kroencke (2017). Figure 1 presents the consumption growth rates of the five measures, jointly with the excess market return, which we also use as a risk factor for comparison.

Figure 1 shows that all of the consumption growth rates are much less volatile than the market return, although some growth rates appear slightly more volatile than others. This is also reflected by the large standard deviation of the market return, 16.89, compared to the much smaller standard deviations of

Figure 1. Consumption growth and market return. This figure plots the market return (solid black), consumption growth based on the “Reported” NIPA measure (dashed), “P-J” (dashed o), “Q4-Q4” (dotted), “Garbage” (dashed diamond), and “Unfiltered” (dashed +). The garbage consumption growth is from Savov (2011). The other series are from Kroencke (2017) for the period 1960 to 2014 (yearly data).

consumption growth, which are in the 1.30 (“Reported”) to 3.06 (“P-J”) range.⁵ Overall, Figure 1 indicates that the behavior of the various consumption growth rates differs substantially from that of the market return.

Macroeconomic data are often available only at yearly or quarterly frequencies, so their number of observations is typically rather small; for example, here $T = 55$. To adequately represent the diversity of asset returns, it is also common to use a rather large number of assets. Kroencke (2017) uses 30 portfolios sorted by size, value, and investment in addition to the market portfolio, so $N = 31$. Many other studies use the well-known 25 Fama-French size and book-to-market sorted portfolios. The time series sample size T is then often not much larger than the cross-section sample size N . Asymptotic approximations for the distributions of two-pass estimators can then be problematic for two reasons:

- (1) The rank condition for identification of the risk premia might (almost) fail.
- (2) The “limited T versus large N ” problem (e.g., when $N = 31$, there are about 500 elements to be estimated in the covariance matrix of asset returns, but $T = 55$ is not sufficient to do so consistently).

We therefore extend the so-called identification pre-tests and identification-robust statistics designed to address the first problem to allow for settings in which T is just slightly larger than N , as we show next.

C. One Factor: Rank Test and its Application

We respecify the linear factor model and the accompanying moment equation such that the full rank identification condition on $(\iota_N : \beta)$ becomes less involved

⁵ These standard errors are scaled by 100, as reported in Kroencke (2017).

to test. We do so by subtracting the N^{th} equation from the first $(N - 1)$ equations in the linear regression model in (11) and from the moment equation in (5), which become

$$\begin{aligned}\mathbf{R}_t &= \boldsymbol{\alpha} + \mathbf{B}f_t + \mathbf{u}_t, \\ E(\mathbf{R}_t) &= \mathbf{B}\lambda_f,\end{aligned}\tag{15}$$

with $\mathbf{R}_t = (R_{1t} \dots R_{(N-1)t})' - \iota_{N-1}R_{Nt}$, $\mathbf{B} = (\beta_1 \dots \beta_{N-1})' - \iota_{N-1}\beta'_N$, $\boldsymbol{\alpha} = (c_1 \dots c_{N-1})' - \iota_{N-1}c_N$, and $\mathbf{u}_t = (u_{1t} \dots u_{(N-1)t})' - \iota_{N-1}u_{Nt}$. The full rank condition on $(\iota_N : \beta)$ has now become identical to a full rank condition on $\mathbf{B}$. The test statistics that we propose below are invariant with respect to the asset return being subtracted. The proof of this invariance is provided in the Internet Appendix.

In models with excess asset returns and λ_0 set to zero, we do not remove the N^{th} equation. The linear factor model and factor pricing moment condition then read

$$\begin{aligned}R_t &= c + \beta f_t + u_t, \\ E(R_t) &= \beta\lambda_f,\end{aligned}\tag{16}$$

and differ from (15) only with respect to the specification of the asset returns, since $\mathbf{R}_t$ in (15) has the N^{th} asset return subtracted while R_t in (16) is just the unaltered N -dimensional vector of (excess) asset returns. Since the specifications of (15) and (16) are almost identical except for the vector of asset returns, our proposed tests extend straightforwardly to linear factor pricing without a zero-beta return. For example, for the one-factor model with no zero-beta return, the rank test just tests whether $H_0 : \beta = 0$. We therefore refrain from explicitly stating these extensions here, although we do so in the Internet Appendix.

Section I.A.1 shows that the full rank condition on $(\iota_N : \beta)$ also applies to specifications in which the expected asset returns are not fully spanned by the betas of the specified risk factors as in (9). Our rank test thus also applies to such models.

C.1. *F-Statistic for Testing Identification under Small T*

When there is just one risk factor so $k = 1$, a lower rank value of $\mathbf{B}$ implies a zero value of $\mathbf{B}$. Testing for identification of the risk premia is then identical to testing $H_0 : \mathbf{B} = 0$ in the linear factor model (15). This hypothesis is commonly tested using the Wald statistic⁶

$$W(\mathbf{B} = 0) = T \hat{Q}_{FF} \hat{\mathbf{B}}' \hat{\Sigma}^{-1} \hat{\mathbf{B}},\tag{17}$$

⁶ The Wald statistic also results from the difference between the sum of squared residuals that result from estimating (11) with a constant value of β_i , $i = 1, \dots, N$, imposed and unconstrained estimation.

with $\hat{\mathbf{B}}$ the least squares estimator of $\mathbf{B}$, $\hat{\mathbf{B}} = \sum_{t=1}^T \bar{\mathbf{R}}_t \bar{f}_t' (\sum_{t=1}^T \bar{f}_t \bar{f}_t')^{-1}$, $\bar{\mathbf{R}}_t = \mathbf{R}_t - \bar{\mathbf{R}}$, $\bar{\mathbf{R}} = \frac{1}{T} \sum_{t=1}^T \mathbf{R}_t$, $\hat{\mathbf{Q}}_{FF}$ the sample variance of f_t , $\hat{\mathbf{Q}}_{FF} = \frac{1}{T} \sum_{t=1}^T \bar{f}_t \bar{f}_t'$, and $\hat{\Sigma}$ the residual sample covariance matrix, $\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T (\bar{\mathbf{R}}_t - \hat{\mathbf{B}} \bar{f}_t)(\bar{\mathbf{R}}_t - \hat{\mathbf{B}} \bar{f}_t)'$. Under the usual i.i.d. finite variance conditions (see, e.g., Shanken (1992)), the Wald statistic converges under H_0 to a χ_{N-1}^2 distributed random variable when the sample size T goes to infinity:

$$W(\mathbf{B} = \mathbf{0}) \xrightarrow{d} \chi_{N-1}^2, \quad \text{as } T \rightarrow \infty. \quad (18)$$

However, in settings in which T is not much bigger than N , the χ_{N-1}^2 distribution does not provide an accurate approximation of the distribution of the Wald statistic. This has motivated, for example, GRS to propose their statistic to test $\alpha = 0$ in the linear model (15), which has an exact F -distribution when we assume normal i.i.d. errors and fixed factors. This assumption on the errors is often not considered overly restrictive for yearly and quarterly data given the improved accuracy of the test procedure it provides; see also Campbell, Lo, and MacKinlay (1997).^{7,8} A similar F -statistic, which has an F -distribution under the same conditions as the GRS statistic, can also be used to test $H_0 : \mathbf{B} = \mathbf{0}$:

$$F(\mathbf{B} = \mathbf{0}) = \frac{T - N}{N - 1} \hat{\mathbf{Q}}_{FF} \hat{\mathbf{B}}' \hat{\Sigma}^{-1} \hat{\mathbf{B}} \sim F(N - 1, T - N). \quad (19)$$

When T is not much larger than N , the asymptotic approximation used for the distribution of the Wald statistic is inaccurate, so the Wald rank test becomes unreliable while the F -test in (19) remains accurate. In contrast, when T is much larger than N , the F -test in (19) is identical to the Wald-test in (18).

To illustrate the performance of the two rank tests described above, we present their simulated sizes at the 5% level for DGPs calibrated to the data used in Kroencke (2017). Figure 2 shows that the F -test performs well under various values of N and T : its actual size is close to the nominal 5%. In contrast, the Wald test severely overrejects the null (e.g., in Panel A it rejects the null with a probability around 90%, instead of the nominal 5%) when T is small, while it is equivalent to the F -test when T is large. Overall, Figure 2 indicates that the F -test is preferred in empirical settings where T is not large. These results are all in line with the simulation results reported in GRS.

⁷ For each of the 31 test assets considered, a Jarque-Bera normality test (see Jarque and Bera (1987)) does not reject at the 5% level that the residuals from (15) are normally distributed for many different consumption measures.

⁸ Normality is also assumed by Kan and Zhou (2004), who derive the exact distribution of the Hansen-Jagannathan statistic for model specification. Anatolyev and Gospodinov (2011) adjust the Anderson-Rubin and overidentification tests from the instrumental variable regression literature to make them applicable in settings where N is large. We leave further generalizations (such as relaxing the normality and homoskedasticity assumptions) of our proposed tests for large N to future research.

Figure 2. Rejection frequencies of $H_0 : \mathbf{B} = 0$ at the 5% level using Wald and F tests. This figure plots the sizes of rank tests (Wald (solid), F (dashed)) used to test $H_0 : \mathbf{B} = 0$ at the 5% level, where $\mathbf{B}$ is the centered version of β (i.e., subtract the N^{th} element). The data are generated from the linear factor model $R_t = \iota_N \lambda_0 + \beta(f_t + \lambda_f) + u_t$, where β is set to $\iota_N \cdot d$, with $d \in [-3, 3]$. λ_0, λ_f , and the variances of f_t and u_t are calibrated using $N = 31$ or the first $N = 5$ test assets in Kroencke (2017) and the garbage consumption growth measure of Savov (2011). The number of Monte Carlo replications is 5,000. Similar simulation outcomes for multifactor models are presented in the Internet Appendix.

C.2. Estimation and Tests for Consumption Growth Measures

Table I reports the first-pass betas for the five different consumption growth measures and the market return using the data from Kroencke (2017). The table also reports a goodness of fit measure and the p -values of the F -test for the significance of $\mathbf{B}$. Following Kroencke (2017), we use as test assets 30 portfolios sorted by size, value, and investment, plus the market portfolio, so $N = 31$. The values for β in Table I exhibit substantial differences across the different consumption measures. In particular, the estimated β from reported NIPA consumption appears much smaller overall than its counterparts from alternative consumption measures such as the garbage measure in Savov (2011) or the unfiltered NIPA consumption in Kroencke (2017). This is also reflected by the pseudo- R^2 reported in Table I, which captures the percentage of the total variation of asset returns that is explained by the risk factor in each model. For comparison, Table I also reports the estimated β when using the market return “ R_m ” as the single risk factor, as in the classical capital asset pricing model. The large t -statistics and pseudo- R^2 resulting from “ R_m ” indicate that, compared to the various consumption growth rates (“Reported,” “P-J,” “Q4-Q4,”

Table I
 β with 31 Portfolio Returns

The test assets are the 30 portfolios sorted by size, value, and investment, plus the market portfolio, over 1960 to 2014 (yearly data) taken from Kroencke (2017). The pseudo- R^2 is a goodness of fit measure that captures the percentage of the variation of asset returns that is explained by the risk factor; see Kleibergen and Zhan (2015). The reported p -value is based on the F -test of $H_0 : rank(\iota_N : \beta) = 1$ (or equivalently, $\mathbf{B} = 0$). To gauge consumption growth ("Reported," "P-J," "Q4-Q4," "Garbage," and "Unfiltered"), we also consider the return of the market portfolio (" R_m ") as the single risk factor. In this scenario, the market portfolio is excluded from the test assets, so the remaining 30 portfolios sorted by size, value, and investment are used as test assets.

	Reported		P-J		Q4-Q4		Garbage		Unfiltered		R_m	
	$\hat{\beta}$	t -Stat	$\hat{\beta}$	t -Stat	$\hat{\beta}$	t -Stat	$\hat{\beta}$	t -Stat	$\hat{\beta}$	t -Stat	$\hat{\beta}$	t -Stat
(1)	-1.30	-0.41	1.19	0.93	4.18	1.51	4.54	3.33	5.40	3.81	1.37	8.54
(2)	-0.23	-0.09	1.14	1.06	4.42	1.93	4.06	3.52	4.70	3.98	1.22	10.06
(3)	0.14	0.06	0.99	1.06	4.15	2.09	3.74	3.86	3.80	3.64	1.10	11.19
(4)	-0.52	-0.25	0.73	0.86	3.46	1.92	3.66	4.30	3.33	3.50	1.02	12.15
(5)	-0.23	-0.11	0.73	0.85	3.74	2.06	3.62	4.30	3.33	3.44	1.07	14.49
(6)	-0.84	-0.43	0.58	0.72	3.01	1.75	3.42	4.39	3.10	3.43	1.03	16.95
(7)	-0.85	-0.44	0.49	0.63	2.73	1.63	3.51	4.90	2.95	3.32	1.01	18.06
(8)	-0.59	-0.33	0.39	0.53	2.69	1.69	3.06	4.55	2.85	3.41	0.95	16.49
(9)	-0.12	-0.07	0.30	0.43	2.75	1.87	2.61	3.98	2.48	3.16	0.91	20.95
(10)	0.34	0.20	0.64	0.92	2.97	2.01	3.14	4.53	2.60	3.29	0.94	25.53
(11)	-0.95	-0.46	0.21	0.25	2.25	1.22	3.58	4.29	2.85	2.91	1.07	15.59
(12)	-0.44	-0.25	0.60	0.81	2.77	1.77	3.15	4.22	2.59	3.09	0.98	24.07
(13)	-0.63	-0.38	0.52	0.78	2.49	1.73	2.95	4.38	2.22	2.84	0.89	21.78
(14)	-0.56	-0.31	0.07	0.10	2.26	1.39	2.51	3.10	2.04	2.28	0.95	15.29
(15)	0.88	0.53	0.46	0.67	3.14	2.19	2.75	3.89	2.25	2.85	0.85	14.11
(16)	0.73	0.45	0.58	0.88	3.04	2.19	2.86	4.49	2.62	3.56	0.83	15.45
(17)	0.96	0.56	0.71	1.03	3.50	2.40	2.93	4.53	2.81	3.61	0.84	12.05

(Continued)

Table I—Continued

	Reported		P-J		Q4-Q4		Garbage		Unfiltered		R_m	
	$\hat{\beta}$	t -Stat	$\hat{\beta}$	t -Stat	$\hat{\beta}$	t -Stat	$\hat{\beta}$	t -Stat	$\hat{\beta}$	t -Stat	$\hat{\beta}$	t -Stat
(18)	1.21	0.65	1.05	1.40	4.27	2.72	3.17	4.35	3.33	4.02	0.91	11.90
(19)	1.23	0.67	1.09	1.47	3.97	2.56	2.86	3.65	2.96	3.53	0.84	9.88
(20)	1.77	0.84	1.32	1.56	5.22	2.98	2.93	3.26	4.21	4.67	0.99	10.40
(21)	-0.50	-0.22	0.58	0.61	3.57	1.78	3.85	4.30	3.88	3.73	1.15	13.40
(22)	-0.09	-0.05	0.66	0.88	3.14	1.98	3.00	4.10	3.15	3.82	0.93	14.19
(23)	0.37	0.23	0.54	0.81	2.84	2.02	2.94	4.49	2.27	2.96	0.84	15.29
(24)	0.44	0.31	0.54	0.92	2.96	2.40	2.84	5.30	2.25	3.38	0.73	14.27
(25)	-0.50	-0.32	0.31	0.49	2.11	1.57	2.69	5.05	2.22	3.10	0.80	16.67
(26)	1.04	0.62	0.79	1.16	3.70	2.60	3.09	4.44	2.54	3.25	0.90	20.33
(27)	1.10	0.66	0.98	1.46	3.71	2.63	2.90	4.13	2.74	3.60	0.88	17.81
(28)	0.03	0.01	0.71	0.98	3.18	2.07	3.15	4.32	3.07	3.83	0.96	20.55
(29)	-0.89	-0.41	0.17	0.19	2.44	1.26	3.54	3.82	2.70	2.59	1.14	16.92
(30)	-1.40	-0.58	0.03	0.03	2.10	0.97	3.92	3.95	2.92	2.52	1.25	15.86
(31)	0.23	0.13	0.72	1.00	3.32	2.21	3.21	4.74	2.95	3.72	1.00	—
Pseudo- R^2	0.003		0.015		0.067		0.270		0.182		0.778	
p -Value												
$\mathbf{B} = 0$	0.257		0.495		0.533		0.468		0.373		0.037	

Table II

Risk Premium λ_f with 31 Portfolio Returns

The test assets are the 30 portfolios sorted by size, value, and investment, plus the market portfolio, over 1960 to 2014 (yearly data) taken from Kroencke (2017). The estimate of λ_f and t -statistics result from the Fama-MacBeth (1973) two-pass procedure. The cross-sectional R^2 results from the regression of $\bar{R}$ on $(\iota_N : \hat{\beta})$. FACCHECK equals the percentage of the variation of the residuals explained by the three largest principal components. FACCHECK of the test assets is 93%. The GRS-FAR test, given in (23) and discussed in Sections I.D and I.E, uses a critical value from the F -distribution to construct the 95% confidence set (CS).

	Reported	P-J	Q4-Q4	Garbage	Unfiltered	R_m
Estimate of λ_f	0.37	4.42	2.00	1.71	2.04	4.60
FM t	0.73	3.04	2.98	1.21	2.18	1.08
Shanken t	0.70	1.83	1.81	1.05	1.75	1.05
R^2	0.027	0.589	0.640	0.150	0.671	0.126
FACCHECK	93%	93%	93%	90%	92%	72%
95% C.S. of λ_f						
GRS-FAR	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-23.9, 101.4)$

“Garbage,” and “Unfiltered”), the market factor is much more closely related to the test assets.

It is worth noting that the estimated betas in Table I exhibit two patterns. First, there are betas that are not significantly different from zero; see, for example, the “Reported” consumption measure. Second, some betas appear to be sizeable and significant but their elements are alike, as is the case for the “Garbage” and “Unfiltered” consumption measures. If β is zero, then the matrix $(\iota_N; \beta)$ does not have full rank and consequently $\mathbf{B} = \mathbf{0}$. Similarly, if β has little spread, so it is close to being spanned by ι_N , a nonzero value of $\mathbf{B}$ is also jeopardized. Therefore, the β estimates in Table I call into question whether $\mathbf{B}$ is different from zero for the various consumption measures.

We use the proposed F -test in (19) to evaluate whether $\mathbf{B}$ differs from zero for all cases in Table I. The resulting p -values for testing $H_0 : \mathbf{B} = \mathbf{0}$ are reported. The large p -values in Table I imply that we cannot reject the null at the 5% level, so the rank condition of the two-pass approach is jeopardized for all five consumption measures. Put differently, the large p -values suggest that the statistical quality of the consumption measures considered is not sufficiently satisfactory to render credibility to the FM two-pass procedure. In contrast, the small p -value under “ R_m ” implies that the identification rank condition is likely to be satisfied for the market factor.

Using the FM two-pass approach, we compute risk premia estimates for the different consumption growth measures. These are reported in Table II. Table II shows that all of the estimates of λ_f on consumption growth are positive, although their significance differs.⁹ However, our findings on β in Table I put the estimates and t -statistics on the consumption risk premium in

⁹ Our estimates of λ_f under the five consumption measures, as well as the Shanken (1992) t -statistics and cross-sectional R^2 , are identical to those in Kroencke (2017).

Figure 3. Power curves of the t -test. This figure plots power curves of the t -test with the Shanken (1992) correction. The null hypothesis is $\lambda_f = 2$. The significance level is 5%. The DGP is calibrated to the test assets in Kroencke (2017), with “Reported,” “P-J,” “Q4-Q4,” “Garbage,” “Unfiltered,” and “ R_m ” as the risk factors. The number of Monte Carlo replications is 5,000, with $N = 31$ and $T = 55$.

Table II under doubt, since the sizeable weak factor literature demonstrates that both the estimates and t -statistics can be spurious if the rank condition is jeopardized; see Kan and Zhang (1999a), and Kleibergen (2009).

To briefly illustrate the malfunction of the t -test on the risk premia, we compute its simulated power curve, which we plot in Figure 3. In particular, we calibrate the DGP to the six model specifications in Table II, with $N = 31$ and $T = 55$.¹⁰ Using the simulated data, we test $\lambda_f = 2$, which is close to the average of the five consumption risk premium estimates reported in Table II, using the t -test with the Shanken (1992) correction. Figure 3 shows that the t -test overrejects the null at the 5% level for all five consumption measures. In contrast, the t -test has size close to the nominal 5% when it is used to test the risk premium on the market factor “ R_m .” These findings are consistent with Table I, where all consumption measures indicate issues with the rank identification condition, which implies that the t -test becomes unreliable, while the market factor does not.

Table II also reports the cross-sectional R^2 and FACCHECK, which is a measure of the unexplained factor structure (with three factors) left in the residuals of the time series regression of the FM two-pass procedure,

$$FACCHECK = \frac{v_1 + v_2 + v_3}{v_1 + \dots + v_N}, \quad (20)$$

¹⁰ Similar findings under other choices of N and T are presented in the Internet Appendix, together with the detailed description of the DGP.

Table III
Risk Premium λ_f with 31 Portfolio Returns and No Zero-Beta Return

The test assets are the 30 portfolios sorted by size, value, and investment, plus the market portfolio, over 1960 to 2014 (yearly data) taken from Kroencke (2017). The estimate of λ_f results from the Fama-MacBeth (1973) two-pass procedure under $\lambda_0 = 0$. The t -statistic results from Newey-West standard errors with three lags, as in Savov (2011). The rank test p -value is based on the F -test of $H_0 : \beta = 0$. The GRS-FAR test uses a critical value from the F -distribution to construct the 95% confidence set. The Internet Appendix provides details of the rank and GRS-FAR tests without the zero-beta return.

	Reported	P-J	Q4-Q4	Garbage	Unfiltered	R_m
Estimate of λ_f	0.31	10.14	2.28	2.09	2.44	7.55
FM t	0.61	3.87	3.73	3.25	3.48	3.60
Rank test						
p -Value	0.269	0.531	0.273	0.006	0.211	0.000
95% C.S. of λ_f						
GRS-FAR	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-0.8, 7.8)$	$(-\infty, -5.8)$ $\cup (-0.6, \infty)$	$(4.7, 8.0)$

where $v_1 > v_2 > \dots > v_N$ are the characteristic roots of the covariance matrix in descending order; see Kleibergen and Zhan (2015).¹¹ Similar to the malfunction of the t -test, Kleibergen and Zhan (2015) show that the cross-sectional R^2 can also be spuriously large under weak or useless factors, when a strong factor structure is present in the residuals of the first-pass time series regression of the FM two-pass procedure. We find that FACCHECK for the original $N = 31$ test assets is 93%. This implies that 93% of the variation in the test assets is explained by the three largest principal components, that is, there is a strong factor structure. Table II, however, shows that the factor structure checks for the residuals are also around 90% to 93% under the five consumption growth rates, which implies that these consumption growth rates do not explain any of the main factors of the asset returns. This evidence shows that the large cross-sectional R^2 s in Table II do not have to indicate strong factor pricing. The difficulty in interpreting the cross-sectional R^2 s can also be inferred from the pseudo- R^2 s in Table I, which are small for some of the consumption growth series, for example, “P-J,” “Q4-Q4,” and “Unfiltered,” for which the cross-sectional R^2 s are large. The former (pseudo- R^2 in Table I) shows little explanatory power, which implies that the latter (cross-sectional R^2 in Table II) should be interpreted with caution; see Kleibergen and Zhan (2015).

While Table II reports estimation results for the risk premia when the zero-beta return is incorporated, Table III does so when the zero-beta return is no longer present. The first impression when comparing the values of the FM t -test across these tables is that the significance of the risk premia is greatly enhanced by removing the zero-beta return. However, in Table III the p -values from our rank test, which just tests $H_0 : \beta = 0$ since the zero-beta return is

¹¹ As explained in Kleibergen and Zhan (2015), this corresponds to using the trace norm of the covariance matrix as a measure of the total variation.

left out, show that this first impression is misleading. Kan and Zhang (1999a) emphasize that the FM t -test is not credible, regardless of whether the zero-beta return is included, if $\beta = 0$. Thus, except for the “Garbage” consumption measure, we cannot trust the FM t -test for the risk premia on any of the consumption measures.

D. GRS-FAR Test

The normal approximation of the distribution of the FM two-pass estimator is accurate when the sample size is large and $\mathbf{B}$ differs from zero. In this case, we could use the F -statistic in (19) to test the significance of $\mathbf{B}$ and use the normal distribution for the FM t -test if $\mathbf{B}$ is found to be significant. However, this procedure introduces pre-test bias for the consecutive FM t -test.¹² To avoid pre-test bias, we propose testing a hypothesis on the risk premia, say $H_0 : \lambda_f = \lambda_{f,0}$, using another extension of the GRS statistic, which we refer to as a GRS-FAR statistic because of its similarity to the Anderson-Rubin (1949) statistic in linear instrumental variables regression.

When the model is correctly specified and $H_0 : \lambda_f = \lambda_{f,0}$ holds, the factor model in (15) can be respecified as

$$\begin{aligned}\mathbf{R}_t &= \boldsymbol{\alpha}(\lambda_{f,0}) + \mathbf{B}(\bar{f}_t + \lambda_{f,0}) + \mathbf{u}_t, \\ E(\mathbf{R}_t) &= \mathbf{B}\lambda_{f,0},\end{aligned}\tag{21}$$

where $\bar{f}_t = f_t - \mu_f$ is the demeaned factor and the sample counterpart of μ_f is $\bar{f} = \frac{1}{T} \sum_{t=1}^T f_t$. The null hypothesis H_0 then implies that

$$H_0 : \boldsymbol{\alpha}(\lambda_{f,0}) = \boldsymbol{\alpha} - \mathbf{B}(\lambda_{f,0} - \mu_f) = \mathbf{0}.\tag{22}$$

We can now test H_0 by conducting a GRS test for the significance of $\boldsymbol{\alpha}(\lambda_{f,0})$ in (21). We do so just by running (time series) regressions of the different elements of $\mathbf{R}_t$ on a constant term and $\bar{f}_t + \lambda_{f,0}$, where $\lambda_{f,0}$ is specified by the null hypothesis, $H_0 : \lambda_f = \lambda_{f,0}$. We then test the (joint) significance of the constant terms in the $N - 1$ different regressions at the prespecified value of

¹² When testing the significance of $\mathbf{B}$, we could use a larger critical value to avoid this. For example, Stock and Yogo (2005) suggest using a larger critical value for the first-stage F -statistic, which is used as a pre-test for identification of the two-stage least squares regression in the linear instrumental variables regression model. This rule of thumb, however, does not address the “limited T versus large N ” problem, so we would have to increase it further. The limited simulation study conducted in Figure 3 shows that we could use the p -value from testing $\mathbf{B} = \mathbf{0}$ by the F -test as a pre-test measure, as suggests that a tiny p -value might indicate that the normal approximation of the FM t -test would do. A much more extensive simulation study, however, would be needed to determine an appropriate threshold. We also note that such a pre-test makes the FM t -test less powerful than our GRS-based test since a significant value of the F -test already shows that the confidence set from the GRS-based test is bounded, which implies that meaningful inference is possible, while a much larger value of the F -test would be needed to render the FM t -test reliable.

$\lambda_{f,0}$ using the GRS test. We state its resulting specification in two (identical) ways:

$$\begin{aligned} \text{GRS-FAR}(\lambda_{f,0}) &= \frac{T}{1 - \lambda'_{f,0} \hat{Q}(\lambda_{f,0})^{-1} \lambda_{f,0}} \hat{\alpha}(\lambda_{f,0})' \hat{\Sigma}^{-1} \hat{\alpha}(\lambda_{f,0}) \\ &= \frac{T}{1 + \lambda'_{f,0} \hat{Q}_{FF}^{-1} \lambda_{f,0}} (\bar{\mathbf{R}} - \hat{\mathbf{B}} \lambda_{f,0})' \hat{\Sigma}^{-1} (\bar{\mathbf{R}} - \hat{\mathbf{B}} \lambda_{f,0}), \end{aligned} \quad (23)$$

where

$$\hat{\alpha}(\lambda_{f,0}) = \bar{\mathbf{R}} - \tilde{\mathbf{B}}(\lambda_{f,0}) \lambda_{f,0} \quad (24)$$

and $\tilde{\mathbf{B}}(\lambda_{f,0}) = \sum_{t=1}^T \mathbf{R}_t (\bar{f}_t + \lambda_{f,0})' (\sum_{t=1}^T (\bar{f}_t + \lambda_{f,0}) (\bar{f}_t + \lambda_{f,0})')^{-1}$ is the least squares estimator of $\mathbf{B}$ in (21) with H_0 imposed, so that $\alpha(\lambda_{f,0})$ is restricted to zero, and $\hat{Q}(\lambda_{f,0}) = \frac{1}{T} \sum_{t=1}^T (\bar{f}_t + \lambda_{f,0}) (\bar{f}_t + \lambda_{f,0})'$. The first specification of the GRS-FAR statistic stated in (23) closely resembles the GRS statistic and is identical to the FAR statistic proposed in Kleibergen (2009), while the second one is identical to the Hotelling-type statistic proposed in Beaulieu, Dufour, and Khalaf (2013). A proof of the equivalence of both expressions is given in Kleibergen, Kong, and Zhan (2019). Khalaf and Schaller (2016) provide a “split-sample” extension of the GRS-FAR test.

Since it is a GRS test, the distribution of the GRS-FAR statistic does not depend on $\mathbf{B}$. Under $H_0 : \lambda_f = \lambda_{f,0}$, i.i.d normal errors, and fixed factors, it is (proportional to) an F -distributed random variable,

$$\frac{T - k - N + 1}{T(N - 1)} \times \text{GRS-FAR}(\lambda_{f,0}) \sim F(N - 1, T - k - N + 1), \quad (25)$$

while under i.i.d errors, it converges to a χ^2_{N-1} distributed random variable when the sample size T goes to infinity,

$$\text{GRS-FAR}(\lambda_{f,0}) \xrightarrow{d} \chi^2_{N-1}, \quad \text{as } T \rightarrow \infty. \quad (26)$$

Identical to the rank test, the use of critical values that result from the F -distribution leads to appropriate sizes when T is not much larger than N , while χ^2 critical values then lead to large size distortions, as we show by simulation below. When T is much larger than N , it is immaterial which critical values are used and both lead to correct sizes that do not depend on $\mathbf{B}$. The FM t -test can still be unreliable, however, in these settings when $\mathbf{B}$ is close or equal to zero.

The GRS-FAR statistic jointly tests whether all factor pricing conditions in (1) hold at $H_0 : \lambda_f = \lambda_{f,0}$. This is different from the FM t -test, which does not test whether expected asset returns are fully spanned by the betas with risk premia equal to their hypothesized value, $\lambda_{f,0}$, but tests whether expected asset returns lie in the direction of the betas of the appropriate risk factors as in (9). In contrast, a rejection of $H_0 : \lambda_f = \lambda_{f,0}$ using the GRS-FAR test can result (i) when factor pricing occurs at a value of λ_f other than the hypothesized one,

$\lambda_{f,0}$, (ii) or when there is no value of the risk premia at which factor pricing occurs, so expected asset returns are not fully spanned by the betas as in (9). The latter cannot occur using the FM t -test, since it does not test all factor pricing conditions.

The $100 \times (1 - \alpha)\%$ confidence set for λ_f (denoted by $CS_{\lambda_f}(\alpha)$ below) that results from the GRS-FAR test consists of all values of $\lambda_{f,0}$ for which the GRS-FAR test does not reject using the $100 \times (1 - \alpha)\%$ critical value that results from the F -distribution:

$$CS_{\lambda_f}(\alpha) = \left\{ \lambda_{f,0} : \frac{T - k - N + 1}{T(N - 1)} \times \text{GRS-FAR}(\lambda_{f,0}) \leq F_{\alpha}(N - 1, T - k - N + 1) \right\}, \quad (27)$$

where $F_{\alpha}(N_1, N_2)$ is the upper α^{th} quantile of the $F(N_1, N_2)$ distribution. Since GRS-FAR($\lambda_{f,0}$) is not a quadratic function of λ_f , it cannot directly be inverted, so the confidence set in (27) does not have the usual expression of an estimator plus or minus a multiple of the standard error. Instead, we have to specify a grid of values for $\lambda_{f,0}$ and compute the GRS-FAR statistic for each value of $\lambda_{f,0}$ on the grid to determine if it is less than the appropriate critical value so $\lambda_{f,0}$ is part of the confidence set. Alternatively, we can use Dufour and Taamouti (2005), who provide a closed-form expression of the confidence set (see also the Internet Appendix).

Specifically, the confidence set in (27) can have three distinct shapes.

- (1) Bounded and convex: There is a closed compact set of values of $\lambda_{f,0}$ for which the test statistic $\frac{T-k-N+1}{T(N-1)} \times \text{GRS-FAR}(\lambda_{f,0})$ does not exceed the critical value.
- (2) Unbounded: This occurs either when there are no values of $\lambda_{f,0}$ for which the test statistic exceeds the critical value (unbounded and convex), or when there is a bounded set of values of $\lambda_{f,0}$ for which the test statistic exceeds the critical value (unbounded and disjoint).
- (3) Empty: The test statistic exceeds the critical value for all values of $\lambda_{f,0}$.

Bounded and convex confidence sets occur when the risk premia are well identified and the expected asset returns are spanned by the β s, so the moment equation in (21) holds. Unbounded confidence sets are indicative of a small $\mathbf{B}$. When $\mathbf{B}$ and $E(\mathbf{R}_t)$ are both equal to zero, all values of the risk premia λ_f make the moment condition hold. Hence, if we test $H_0 : \lambda_f = \lambda_{f,0}$ at a very large, possibly infinite, value of $\lambda_{f,0}$ using a size-correct test at, say, the 5% significance level, it would reject 5% of the time.¹³ In this case, we mostly do not reject the hypothesis of an infinite value of λ_f , that is, we obtain an unbounded 95% confidence set. Dufour (1997, Theorems 3.3 and 3.6) formally proves this argument and shows that any size-correct procedure used to test parameters that can be nonidentified must have a positive probability of producing an

¹³ For infinite values of the hypothesized risk premium parameter, $\lambda_{f,0}$, the GRS-FAR statistic in (23) equals the F -statistic (19) testing if $\mathbf{B}$ equals zero.

Figure 4. Power curves of the GRS-FAR Test. This figure plots power curves of the GRS-FAR test with χ^2 critical values (solid), F critical values (dashed), and the t -test (dashed diamond) with the Shanken (1992) correction. The null hypothesis is $\lambda_f = 2$. The significance level is 5%. The DGP is calibrated to $N = 31$ or the first $N = 5$ test assets in Kroencke (2017) and the garbage consumption growth in Savov (2011). The number of Monte Carlo replications is 5,000.

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 nonidentified. Empty confidence sets occur when the model is misspecified, so there is no value of λ_f for which the moment condition holds. In this case, the expected asset returns are not spanned by the β s as in (9). Misspecification can also lead to unbounded confidence sets that exclude a bounded set of values of the risk premia when $\mathbf{B}$ is small.

We conduct a simulation exercise to show the relevance of the GRS-FAR test with F critical values. The DGP is calibrated to the test assets in Kroencke (2017) and the garbage consumption growth in Savov (2011).¹⁴ The resulting power curves using F and χ^2 critical values and for the FM t -test testing $H_0 : \lambda_f = 2$ are presented in Figure 4. The figure shows that the GRS-FAR test using F critical values performs well for various values of N and T . In particular, in the settings of Figure 4, Panel A with $N = 31$ and $T = 55$ calibrated to Kroencke (2017), this test has actual size close to the nominal 5% level. In contrast, the GRS-FAR test with χ^2 critical values severely overrejects the null when T is small, although it becomes equivalent to the GRS-FAR test with

¹⁴ We obtain similar results when we calibrate to other consumption measures.

Figure 5. p -Values of testing the risk premium by the GRS-FAR test with 31 portfolio returns. The test assets, taken from Kroencke (2017), comprise the 30 portfolios sorted by size, value, and investment, plus the market portfolio, in 1960 to 2014 (yearly data). The null hypothesis is that λ_f is equal to the value on the horizontal axis. The corresponding p -values of the GRS-FAR test (using F critical values (dashed)) are plotted, together with the 5% line (solid).

F critical values when T becomes large. Similar to Figure 3, Figure 4 also shows that the FM t -test overrejects the null when T is small, although its performance improves as T increases while it has less or comparable power to the GRS-FAR test. Overall, Figure 4 shows that it is appealing to use the GRS-FAR test with F critical values to conduct inference on risk premia, especially when the length of the time series is small and/or the statistical quality of risk factors is potentially weak.

E. Inference on Consumption Risk Premia using the GRS-FAR Test

E.1. Including a Zero-Beta Return

The bottom row of Table II shows the 95% confidence sets for the risk premium on the different consumption measures and the market return that result from the GRS-FAR test. The 95% GRS-FAR confidence sets result from the p -value plots of the GRS-FAR test for testing λ_f equal to the value on the horizontal axis in Figure 5. A p -value larger than 0.05 implies that the hypothesized value of λ_f on the horizontal axis is not rejected at the 5% significance level, so the 95% confidence sets result from the intersection of the p -value plot with the 5% line. For the different consumption growth measures, the p -value plots lie above the 5% line, so the 95% confidence sets for their risk premia are all unbounded, while the confidence set for the risk premium on the market return is bounded. This shows that little information for the risk premium on the consumption growth measures is contained in the data.

Recall that Table I shows that we cannot reject the hypothesis of a zero value of $\mathbf{B}$ at the 5% significance level for any of the consumption growth measures. The moment equation in (21) implies that when $\mathbf{B}$ equals zero, we cannot rule out that the risk premium λ_f is infinite. The unbounded 95% confidence sets for the consumption growth measures in Table II therefore naturally result from the significance tests on $\mathbf{B}$ reported in Table I. Since $\mathbf{B}$ is significant at the 5% level for the market return, the 95% confidence set for its risk premium is bounded albeit very wide. It suggests that the short time series used in Kroencke (2017) make it hard to draw meaningful conclusions about risk premia, even when we use statistically strong risk factors such as the market return.

E.2. No Zero-Beta Return

With no zero-beta return, we conduct a GRS-FAR test of $H_0 : \lambda_f = \lambda_{f,0}$ by evaluating whether $E(R_t) = \beta \lambda_{f,0}$ holds (see the Internet Appendix). The bottom row of Table III shows the 95% confidence sets for the risk premium on the different consumption measures and the market return that result from inverting the GRS-FAR test. These do not differ much from the 95% confidence sets that result from the GRS-FAR test reported in Table II and clearly show the identification issues of the risk premia that arise due to small values of β . Even for the risk premium on the “Garbage” consumption measure, the 95% confidence set that results from the GRS-FAR test is rather wide in Table III and much more so than that resulting from the FM t -test. This confidence set shows that a zero value of the risk premium on the “Garbage” consumption measure is not rejected at the 5% significance level using the GRS-FAR test.

The p -value of testing $H_0 : \beta = 0$ for the market return in Table III is tiny, so the 95% confidence sets of its risk premium that result from the FM t -test (3.44, 11.66) and the GRS-FAR test (4.7, 8.0) are comparable. Since the market return is a traded factor, its risk premium equals its mean return. The average market return equals 5.77 and its 95% confidence set is (1.31, 10.23), which is also comparable to the 95% confidence sets that result from the FM t -test and the GRS-FAR test.

In sum, compared with Table II, Table III shows that removing the zero-beta return helps mitigate the identification problem of the risk premia but does not fully resolve it. In particular, with the zero-beta return included, identification requires a full rank value of $(\iota_N : \beta)$, so β is not allowed to be constant over the different assets. However, when we remove the zero-beta return, we only need $\beta \neq 0$ to achieve identification in the single-factor model, so identical β s for the different assets is allowed. Nonetheless, the “limited T versus large N ” problem is not affected by the inclusion or exclusion of a zero-beta return, so even when we remove the zero-beta return by using excess returns, we can still end up with wide or even unbounded 95% confidence sets for the consumption risk premium, as shown in Table III.

F. Mimicking Portfolios and Alternative Test Assets

The wide and even unbounded confidence sets on the consumption risk premium reported above could induce researchers to consider two specific solutions to remedy it.

The first is to use mimicking portfolios of consumption growth instead of consumption growth itself, since Breeden (1979) and Huberman, Kandel, and Stambaugh (1987) establish that risk factors can be replaced by their mimicking portfolios for asset pricing tests. Ample empirical evidence further shows that replacing nontraded factors with their mimicking portfolios can yield larger betas with better spreads. One may therefore expect the betas from mimicking portfolios to satisfy the rank condition and thus improve inference on risk premia. However, when nontraded factors such as consumption growth are of poor statistical quality, Kleibergen and Zhan (2018) show that their mimicking portfolios are contaminated by nonnegligible estimation errors on the weights that occur during the construction of such portfolios. Specifically, when $\mathbf{B}$ is zero or close to it, these estimation errors can dominate the weights and make the FM two-pass t -test using the mimicking portfolios unreliable. Existing literature commonly ignores these errors, which leads to overstatement of the remedial effects of mimicking portfolios. When taking these errors into account, Kleibergen and Zhan (2018) show that the (GRS-) FAR test for risk factors and the corresponding mimicking portfolio AR (MPAR) test have similar power.¹⁵ Therefore, using mimicking portfolios for consumption growth does not necessarily improve inference on risk premia.

A second potential solution is to adopt alternative test assets, which may yield betas with more variation and thus improve identification of the risk premia. For example, one might want to add the risk-free asset, which has zero beta, to the 30 portfolios sorted by size, value, and investment. By construction, the resulting betas have more variation than those from the 31 portfolios used in Kroencke (2017). In addition, the cross-section dimension can be reduced if we use a smaller number of test assets, which decreases the N/T ratio and thus mitigates the “limited T versus large N ” problem. In the Internet Appendix, we document our empirical findings that result from using alternative test assets, such as the 10 industry portfolios or the 30 portfolios sorted by size, value, and investment augmented by the risk-free asset. We find that adopting alternative test assets may help improve identification of the risk premia, but does not fully resolve the fact that the limited length of the consumption growth series together with the small variation of their betas with test asset returns leads to wide or even unbounded confidence sets of the consumption risk premium for a variety of test assets.

II. Multifactor Models, Larger T , and Misspecification

Instead of using just a single risk factor, it is common to estimate linear factor models with multiple risk factors. The rank and GRS-FAR tests proposed

¹⁵ See their figure 2 for the comparison of FAR and MPAR tests in Kleibergen and Zhan (2018).

Table IV
Risk Premia λ_f in Multifactor Models

The test assets are the 30 portfolios sorted by size, value, and investment over 1960 to 2014 (yearly data) taken from Kroencke (2017). We use the reported NIPA measure for consumption growth Δc . The estimate of λ_f and t -statistics result from the Fama-MacBeth (1973) two-pass procedure. The cross-sectional R^2 results from the regression of $\bar{R}$ on $(\iota_N : \hat{\beta})$. The pseudo- R^2 is a goodness of fit measure that captures the percentage of the variation in asset returns that is explained by the risk factors. FACCHECK equals the percentage of the variation in the residuals explained by the three largest principal components. FACCHECK of the test assets is 93%. The (subset) GRS-FAR test uses critical values from the F -distribution. Plots of its resulting p -values are provided in the Internet Appendix. The rank test p -value is based on the F -test of $H_0 : \text{rank}(\mathbf{B}) = k - 1$.

	Two Factors, $N = 30$		Three Factors, $N = 30$		
	R_m	Δc	R_m	SMB	HML
Estimate of λ_f	9.90	1.43	2.74	3.17	3.78
FM t	2.04	2.91	0.71	1.68	1.90
Shanken t	1.38	1.91	0.68	1.67	1.87
R^2		0.373		0.858	
Pseudo- R^2		0.781		0.893	
FACCHECK		72%		49%	
95% C.S. of λ_f					
GRS-FAR	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$
Rank test					
p -Value		0.611		0.090	

above for a single risk factor naturally extend to such settings. In particular, the (subset) GRS-FAR test allows us to construct confidence sets for the risk premium on each risk factor in a multifactor model (see the subset FAR test in Kleibergen (2009)), while the rank test evaluates the rank value of $\mathbf{B}$, which is necessary to identify the risk premia. For ease of exposition, we provide the expressions of the rank and GRS-FAR statistics and their F bounding distributions in the Internet Appendix, focusing on our empirical findings here.

A. Multifactor Models

Table IV reports results for two multifactor models: a two-factor model with the market factor R_m and consumption growth Δc as risk factors, and the Fama and French (FF, 1993) three-factor model using R_m , SMB (small minus big), and HML (high minus low) as risk factors. The first model is also used in, for example, Savov (2011). We use the reported NIPA measure for consumption. Since R_m is a risk factor in both models, we exclude it from the 31 test assets from Kroencke (2017), so $N = 30$.

Table IV shows that the p -value for the rank test is above 5% for both multifactor models. Thus, for both models, we cannot reject a lower rank value of $\mathbf{B}$ and no identification of the risk premia at the 5% level. This is especially striking for the FF factors since their pseudo- R^2 shows that they explain about

90% of the variation of the returns in the 30 test assets. Because of the insignificant values of the rank tests, we have to interpret the FM two-pass estimates and t -tests with caution. The insignificant values of the rank tests also explain why the 95% confidence sets that result from the (subset) GRS-FAR test (see the Internet Appendix for its implementation) are unbounded for all of the risk premia.

Tables I and IV show that even for risk factors with strong explanatory power for the returns on the test assets, as shown by the pseudo- R^2 for the market return and the FF factors, the rank tests for a lower rank value of $\mathbf{B}$ are just barely significant or insignificant at the 5% level, yielding either very large or unbounded 95% confidence sets for the risk premia. The results basically show that since N is large, the estimation error resulting from the considerable number of components that are estimated for the covariance matrix is too large to make a meaningful statement concerning any risk premia. To make such a statement, we need more time series observations since $T = 55$ and $N = 31$ or 30 leads to the “limited T versus large N ” problem that we discuss above. Since yearly macroeconomic data typically cover a similar time span, it is not likely that their risk premia can be accurately determined. To determine the presence of risk premia for macroeconomic data, we thus need more frequent observations, for example, from quarterly or monthly observations.

B. Larger T

To increase the number of time series observations T , we next consider the quarterly consumption data used in Lettau and Ludvigson (2001). Their data start in the third quarter of 1963 and end in the third quarter of 1998, so $T = 141$. The test assets in Lettau and Ludvigson (2001) are the $N = 25$ size and book-to-market sorted FF portfolios. These data allow us to examine factor pricing with T substantially larger than N .

Table V reports results for two models also studied in Lettau and Ludvigson (2001): the conditional consumption capital asset pricing model with Δc (quarterly consumption growth), cay (lagged consumption-wealth ratio), and their interaction $\Delta c \times cay$ as the three factors, jointly with the Fama-French three-factor model. Our estimates of the risk premia in Table V are identical to those in Lettau and Ludvigson (2001). The significant t -statistic for the interaction term $\Delta c \times cay$ and the large cross-sectional R^2 are used in Lettau and Ludvigson (2001) as empirical evidence in favor of the conditional consumption capital asset pricing model.

However, the pseudo- R^2 and the FACCHECK statistic in Table V show that the three factors— Δc , cay , and $\Delta c \times cay$ —have just minor explanatory power for the test asset returns. This is corroborated by the large p -value of the rank test, which shows that we cannot reject the hypothesis that the risk premia are not identified at the 5% significance level. The (subset) GRS-FAR test thus yields unbounded 95% confidence sets for each of the risk premia in

Table V
Risk Premia λ_f in Multifactor Models with a Large T

The test assets are the 25 Fama-French portfolios over 1963Q3 to 1998Q3 (quarterly data) taken from Lettau and Ludvigson (2001). The estimate of λ_f and the t -statistics result from the Fama-MacBeth (1973) two-pass procedure. The cross-sectional R^2 results from the regression of $\bar{R}$ on $(\iota_N : \hat{\beta})$. The pseudo- R^2 is a goodness of fit measure that captures the percentage of the variation in asset returns that is explained by the risk factors. FACCHECK equals the percentage of the variation in the residuals explained by the three largest principal components. FACCHECK of the test assets is 96%. The (subset) GRS-FAR test uses critical values from the F -distribution. Plots of its resulting p -values are provided in the Internet Appendix. The rank test p -value is based on the F -test of $H_0 : \text{rank}(\mathbf{B}) = k - 1$.

	Three Factors, $N = 25$			Three Factors, $N = 25$		
	Δc	cay	$\Delta c \times \text{cay}$	R_m	SMB	HML
Estimate of λ_f	0.02	−0.13	0.06	1.33	0.47	1.46
FM t	0.20	−0.43	3.12	0.83	0.94	3.24
Shanken t	0.15	−0.31	2.25	0.78	0.94	3.22
R^2		0.698			0.803	
Pseudo- R^2		0.105			0.946	
FACCHECK		95%			38%	
95% C.I. of λ_f						
GRS-FAR	$(-\infty, \infty)$	$(-\infty, \infty)$	$(-\infty, \infty)$	$\emptyset$	$\emptyset$	$\emptyset$
Rank test						
p -Value		0.274			0.030	

the conditional consumption capital asset pricing model; see also Kleibergen (2009).¹⁶

In contrast, the FF factors are statistically strong, as reflected by the pseudo- R^2 , FACCHECK, and the significant rank test in Table V. However, the 95% confidence sets for the risk premia on these three factors resulting from the (subset) GRS-FAR test are empty, which indicates that the average asset returns are not sufficiently spanned by the betas, despite their large cross-sectional R^2 , so there are no values of λ_f that satisfy the beta representation in (1) to the extent that the GRS-FAR test statistic is not significant. Thus, the three-factor model is misspecified since it fails to fully explain the cross-section of expected asset returns.¹⁷ In the case of misspecification, the factor pricing restrictions no longer fully apply, so we need to adapt our inference methods, adopting misspecification-robust methods; see, for example, Kan, Robotti, and Shanken (2013) and Gospodinov, Kan, and Robotti (2014, 2018).

¹⁶ Similarly, we get unbounded confidence sets for risk premia in the conditional consumption capital asset pricing model when using the quarterly data in Parker and Julliard (2005).

¹⁷ We note that the statistics in table 8 of Lettau and Ludvigson (2001) also show that factor pricing is rejected at the 5% significance level when using the FF factors or when using the consumption-based factors.

C. Why Do Such Different Confidence Sets Occur?

When the time series sample size T is not much larger than the cross-section sample size N , as for the Kroencke (2017) data, the 95% confidence sets for the risk premia are unbounded for different settings of the risk premia so they do not seem to conflict with one another. For the Lettau-Ludvigson (2001) data, however, the 95% confidence sets are unbounded for one set of risk premia while they are empty for another set of risk premia. The cross-section sample sizes N are roughly the same for the Kroencke (2017) and Lettau-Ludvigson (2001) data, which raises the question of why the confidence sets for the risk premia on the consumption-based factors remain unbounded when T increases while the confidence sets of the risk premia on the FF factors change dramatically from being unbounded for the Kroencke (2017) data with $T = 55$ to being empty for the Lettau-Ludvigson (2001) data with $T = 141$.

To understand why the different shapes of the 95% confidence sets for the risk premia resulting from the Kroencke (2017) and Lettau-Ludvigson (2001) data can occur, we conduct a simulation experiment using a DGP that is calibrated to the Lettau-Ludvigson (2001) data. We therefore jointly simulate the test asset returns, FF factors, consumption growth, (lagged) consumption-wealth ratio, and their interaction from a 31-dimensional joint normal distribution with the covariance matrix calibrated to that of the returns on the 25 size and book-to-market sorted FF portfolios, the three FF factors (R_m , SMB , and HML), consumption growth, the lagged consumption-wealth ratio, and their interaction resulting from the Lettau-Ludvigson (2001) data. We then construct the 95% confidence sets for the risk premia for different time series sample sizes T and various settings of the risk factors while keeping the cross-section sample size N fixed at 25, $N = 25$. We note that the mean of the simulated test asset returns is also calibrated to these data, so it is not constrained to lie in the span of the covariances between the test asset returns and the risk factors, that is, the DGP is misspecified and does not satisfy the factor pricing moment equation (5). Without loss of generality, the mean of the simulated risk factors is set to zero.

Figure 6 shows the frequency of occurrence of unbounded, bounded, and empty 95% confidence sets as a function of the time series sample size T when using two different sets of simulated risk factors and the 25 simulated size and book-to-market sorted FF portfolios as test assets. In Panel A, the risk factors used are the three FF factors. Panel A shows that for data sets with a small number of time series observations, the 95% confidence sets for the risk premia on the FF factors are often unbounded, which occurs less and less often when the number of time series observations increases. The frequency of occurrence of empty confidence is exactly the opposite. When the number of time series observations increases, the 95% confidence sets increasingly become empty and are always empty for large enough T . The latter results from the misspecification present in the DGP. The frequency of occurrence of bounded confidence sets is increasing for smaller numbers of time series observations and decreasing to zero for larger numbers since the DGP is misspecified.

Figure 6. Frequency of occurrence of empty, unbounded, and bounded 95% confidence sets for the risk premium from the (subset) GRS-FAR test as T increases. The figure plots the frequency of occurrence of empty (solid), unbounded (dashed diamond), and bounded (dashed) 95% confidence sets for the risk premium resulting from the (subset) GRS-FAR test as a function of T . The DGP is a 31-dimensional joint normal distribution whose covariance matrix is calibrated to that of the returns on the 25 size and book-to-market sorted FF portfolios, the three FF factors (R_m , SMB, and HML), consumption growth, the lagged consumption-wealth ratio, and their interaction resulting from the Lettau-Ludvigson (2001) data. The mean of the returns in the DGP is calibrated to the sample average, while the mean of the factors is set to zero. Panel A uses the three FF factors as risk factors in the estimated model and Panel B uses consumption growth, the lagged consumption-wealth ratio, and their interaction as risk factors. The test assets are the $N = 25$ simulated returns on the FF portfolios. The number of Monte Carlo replications is 5,000.

Panel B of Figure 6 shows the frequency of occurrence of unbounded, bounded, and empty 95% confidence sets for the risk premium when we use simulated consumption growth, the lagged consumption wealth ratio, and their interaction as risk factors for the same simulated test assets as used in Panel A. The patterns of the frequencies of occurrence of unbounded, bounded, and empty confidence sets are the same as in Panel A except that the decay and increase are considerably less pronounced.

The two panels in Figure 6 help explain the different shapes of the 95% confidence sets for the risk premium when using the Kroencke (2017) and the Lettau-Ludvigson (2001) data. Specifically, when using $T = 55$, as in the Kroencke (2017) data, Figure 6 shows that the 95% confidence sets are unbounded around 50% of the time when using the (simulated) FF risk factors for estimation and around 90% of the time when using the (simulated) consumption-based factors as risk factors. When using $T = 141$, as is the case for the Lettau-Ludvigson (2001) data, empty 95% confidence sets occur around 90% of the times when using the (simulated) FF factors as risk factors in the estimation and unbounded 95% confidence sets occur around 40% of the time when using the (simulated) consumption-based factors as risk factors.¹⁸ Figure 6 thus shows that the different confidence sets for the risk premium resulting from the Kroencke (2017)

¹⁸ We note that strength of identification of the risk premium might be stronger in the simulated data than in the real Lettau-Ludvigson (2001) data, which is one of the reasons we cannot use bootstrapping to conduct inference on the risk premium.

Figure 7. Frequency of occurrence of empty, unbounded, and bounded 95% confidence sets for the risk premium from the GRS-FAR test as T increases, single risk factor, misspecified models. This figure plots the frequency of occurrence of empty (solid), unbounded (dashed diamond), and bounded (dashed) 95% confidence sets for the risk premia resulting from the GRS-FAR test as a function of T . The DGP is a 31-dimensional joint normal distribution whose covariance matrix is calibrated to that of the returns on the 25 size and book-to-market sorted FF portfolios, the three FF factors (R_m , SMB, and HML), consumption growth, the lagged consumption-wealth ratio, and their interaction resulting from the Lettau-Ludvigson (2001) data. The mean of the returns in the DGP is calibrated to the sample average, while the mean of the factors is set to zero. Panel A uses the market return as the single risk factor in the estimated model and Panel B uses consumption growth as the single risk factor. The test assets are the $N = 25$ simulated returns on the FF portfolios. The number of Monte Carlo replications is 5,000.

and the Lettau-Ludvigson (2001) data are in line with a DGP that is calibrated to the Lettau-Ludvigson (2001) data. Using time series sample sizes beyond that used by Lettau and Ludvigson (2001), we also expect the 95% confidence sets that result for the risk premia on the consumption-based risk factors to become empty since even at $T = 141$, an empty 95% confidence set occurs more frequently than an unbounded one for the simulated data. Figure 6, Panel B, also shows that a bounded 95% confidence set is just a bit less likely than an empty or unbounded one when $T = 141$.

Instead of using three risk factors, Figure 7 shows the frequency of unbounded, bounded, and empty 95% confidence sets as a function of the time series sample size T when using a single risk factor in a simulation experiment where the DGP is identical to that in Figure 6. Panel A of Figure 7 uses the simulated market return as the single risk factor and Panel B uses consumption growth. It is interesting to observe that in Panel A, the 95% confidence set on the risk premium on the market return can still be unbounded for a small number of time series observations but is often bounded for numbers of time series observations comparable to those in Kroencke (2017). Table II shows that this 95% confidence set is bounded but very wide for the Kroencke (2017) data, which is in line with Figure 7, Panel A. Panel B of Figure 7 shows that for consumption growth used as a risk factor, the 95% confidence sets are mostly unbounded for such small time series sample sizes but become less so when the time series sample size increases: as the time series sample size

increases, the 95% confidence sets of the risk premium become more empty. This increase is more rapid when using the market return as a risk factor than when using consumption growth. The empty confidence sets result since the DGP does not satisfy the factor pricing moment equation.

When the time series sample size increases, the sample average of the test asset returns concentrates around its mean. When this mean is not in accordance with factor pricing as in our simulation experiment, we will eventually start to reject factor pricing, which results in an empty 95% confidence set for the risk premium. The time span over which we start to observe empty 95% confidence sets resulting from the GRS-FAR test then depends on how large the estimated covariance matrix (i.e., $\hat{\Sigma}$) is. When the risk factors explain the test asset returns well, as reflected by a large pseudo- R^2 , which is the case for the market return and FF factors, the covariance matrix is rather small so the GRS-FAR test statistic will be larger and we will start to observe empty confidence sets over much smaller time spans than when the risk factors do not explain much of the variation in the returns on the test assets, which is the case for the consumption-based factors. The estimated covariance matrix is then much larger, which results in smaller GRS-FAR statistics that are significant less often. It then takes much longer before we start to observe empty confidence sets, with this time span likely beyond our observed sample size. This explains why for the real data, we observe empty confidence sets for the risk factors with high explanatory power for the test asset returns but not for the risk factors with little explanatory power.

D. Shapes of Confidence Sets for Correctly Specified Models

To further emphasize the importance of misspecification for the DGP used in Figures 6 and 7, we repeat the simulation experiment but now explicitly impose factor pricing, so $E(R_t) = \iota_N \lambda_0 + \beta \lambda_f$ for our simulated data. We do so by simulating from the joint normal distribution used in Figures 6 and 7, but the mean for the test asset returns is specified such that it satisfies the factor pricing condition.¹⁹

Figure 8 shows the frequency of occurrence of unbounded, bounded, and empty 95% confidence sets for the risk premium from the (subset) GRS-FAR test as a function of the time series sample size T for two sets of risk factors. Panel A shows the results for the simulated FF factors while Panel B does so for the consumption-based factors. Since factor pricing is imposed on the DGPs, empty confidence sets occur rarely. For smaller numbers of time series observations, we see that the occurrence of unbounded 95% confidence sets continue to exist for longer time spans when we use the consumption-based factors, since their explanatory power for the test asset returns is much lower

¹⁹ The risk premium used to make the factor pricing condition hold is the FM two-pass estimate that uses the estimated β s from the risk factors. The mean asset return is then obtained by premultiplying the estimated risk premium by the estimated β s. The DGP thus differs for the two sets of risk factors.

Figure 8. Frequency of occurrence of empty, unbounded, and bounded 95% confidence sets for the risk premia from the (subset) GRS-FAR test as T increases, three risk factors, factor pricing imposed on DGPs. This figure plots the frequency of occurrence of empty (solid), unbounded (dashed diamond), and bounded (dashed) 95% confidence sets for the risk premia resulting from the (subset) GRS-FAR test as a function of T . The DGP is a 31-dimensional joint normal distribution whose covariance matrix is calibrated to that of the returns on the 25 size and book-to-market sorted FF portfolios, the three FF factors (R_m , SMB, and HML), consumption growth, the lagged consumption-wealth ratio, and their interaction resulting from the Lettau-Ludvigson (2001) data. The factor pricing restriction is imposed on the mean of the test asset returns. Panel A uses the three FF factors as risk factors in the estimated model and Panel B uses consumption growth, the lagged consumption-wealth ratio, and their interaction as risk factors. The test assets are the $N = 25$ simulated returns on the FF portfolios. The number of Monte Carlo replications is 5,000.

compared to that of the FF factors. It is striking that we observe such confidence sets for the FF factors for smaller numbers of time periods. In all other cases, we obtain the favored bounded 95% confidence sets for the risk premia, which we eventually find only for sufficiently long time spans.

Figure 8 can explain the unbounded 95% confidence sets for the risk premia that we generally observe for the Kroencke (2017) data and for the Lettau-Ludvigson (2001) data when using the consumption-based factors. However, it cannot explain the empty 95% confidence sets for the risk premia that we obtain for the Lettau-Ludvigson (2001) data when using the FF factors as risk factors. What we observe is that factor pricing is a demanding restriction to impose on expected asset returns. Even small deviations from it will imply empty 95% confidence sets for the risk premia for sufficiently long time spans irrespective of whether the risk factors have high explanatory power for the test asset returns.

III. Relative Risk Aversion in Nonlinear Models

In addition to cross-sectional tests based on the linear beta representation, consumption measures are also commonly used in nonlinear asset pricing models. The same identification and time series compared to cross-section sample size issues that arise in linear models also hold in nonlinear models, for

Table VI
GMM Estimates of the Relative Rate of Risk Aversion γ

The moment condition for GMM is $E_t[\delta(\frac{C_{t+1}}{C_t})^{-\gamma} R_{m,t+1}^e] = 0$. The rank test of Wright (2003) tests no identification of γ . Its resulting p -value is reported. The GMM-AR test is from Stock and Wright (2000), for which we provide the expression of the test statistic in the Internet Appendix. The excess market return used for GMM estimation is from 1960 to 2014 (yearly data), as in Kroencke (2017).

	Reported	P-J	Q4-Q4	Garbage	Unfiltered
Estimate of γ	136.15	42.35	64.05	15.63	22.53
<i>SE</i>	52.33	22.87	39.61	8.38	11.98
Rank test, p -value	1.00	0.29	0.50	0.29	0.29
95% C.S. of γ , GMM-AR	(37.51, ∞)	(12.25, ∞)	(16.74, ∞)	(3.88, ∞)	(4.86, ∞)

which we provide further details in the Internet Appendix. We therefore use a straightforward extension of the GRS-FAR test, the so-called GMM-AR test (see Stock and Wright (2000)), to address nonlinear models. Our motivation here is similar to that in the linear case discussed above. In particular, because the t -test is not uniformly valid, we adopt a robust test (GRS-FAR for linear models and GMM-AR for nonlinear models) to reevaluate various consumption measures proposed for asset pricing.

Because of the nonlinearity, we cannot construct a baseline DGP for the returns on the test assets that implies a well-established finite-sample distribution for the GMM-AR statistic, as we did for the GRS-FAR test. We thus have to rely on the asymptotic approximation of the distribution of the GMM-AR statistic, which, as we show in the Internet Appendix, works well for our time spans of interest only when N is small. As a result, below we follow Savov (2011) and Kroencke (2017), and focus on small cross-section dimensions with $N = 1$ and $N = 2$.

A. Confidence Set for Relative Risk Aversion with One Test Asset

Both Savov (2011) and Kroencke (2017) consider the CRRA SDF with just one test asset, namely, the excess market return, so $N = 1$ and

$$E_t \left[\delta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} R_{m,t+1}^e \right] = 0. \quad (28)$$

For convenience, δ is set to 0.95 so the relative rate of risk aversion γ is just-identified. We estimate it using GMM with the sample moment of (28).

Table VI presents our GMM estimates of γ that result from minimizing the GMM objective function and their standard errors, which are identical to those reported in Kroencke (2017).²⁰ Savov (2011) and Kroencke (2017) use the relatively small estimates of γ to render credible their consumption measures

²⁰ The only exception is that under “Reported” in Table VI, we report the GMM estimate, while Kroencke (2017) reports the value of γ that minimizes the mean absolute error.

(“Garbage” and “Unfiltered,” respectively).²¹ However, if γ is weakly identified, the GMM estimates as well as their t -statistics are not reliable; see Stock and Wright (2000) and Kleibergen (2005).

To further illustrate that the conventional GMM t -test is not reliable under the various consumption measures, we adopt the test of Wright (2003) for testing the null that there is no identification of γ . In particular, we test whether the derivative of the moment function has reduced rank, that is, $E[\delta(\frac{C_{t+1}}{C_t})^{-\gamma} R_{m,t+1}^e \Delta c_{t+1}] = 0$, where $\Delta c_{t+1} = \ln(\frac{C_{t+1}}{C_t})$. The resulting p -values of this test are reported in Table VI. The large p -values imply that the null of no identification cannot be rejected. Consequently, the GMM t -test has to be interpreted with caution.

It is worth noting that violation of the rank condition for GMM is similar to the violation we discussed for the two-pass approach. In particular, the rank condition for identification in GMM corresponds to $E[\delta(\frac{C_{t+1}}{C_t})^{-\gamma} R_{m,t+1}^e \Delta c_{t+1}] \neq 0$. When δ is a fixed constant and $\frac{C_{t+1}}{C_t} \approx 1$, this rank condition is jeopardized due to the weak correlation between consumption growth and asset returns. Put differently, the weak correlation between consumption growth and asset returns threatens identification of both the risk premium in the conventional two-pass approach and the relative rate of risk aversion in GMM.

Instead of using the GMM estimates, we test $H_0 : \gamma = \gamma_0$ using the GMM-AR test, as its validity does not depend on the GMM rank condition. A $100 \times (1 - \alpha)\%$ confidence set for γ can be constructed as follows. If the GMM-AR test statistic exceeds the upper α quantile of the χ_1^2 distribution, reject $H_0 : \gamma = \gamma_0$ at the $100 \times \alpha\%$ level. The values of γ_0 that are not rejected constitute the $100 \times (1 - \alpha)\%$ confidence set for γ .²² The 95% confidence sets for γ for the different consumption measures are reported in the last row of Table VI. None of the consumption measures have an upper bound, so little information on the relative rate of risk aversion is available in the different consumption measures.

Since all 95% confidence sets for γ from the GMM-AR test are unbounded, there is no clear indication which consumption measure is preferred.²³ These findings are similar to the unbounded 95% confidence sets we reported for consumption risk premia in linear asset pricing models.

B. Joint Confidence Sets with Two Test Assets

Another commonly used moment condition for relative risk aversion is

$$E_t \left[\delta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} R_{f,t+1} \right] = 1, \quad (29)$$

²¹ They show that their estimated relative risk aversion γ implies a risk-free rate r_f that appears reasonable by using $r_f = -\log(\delta) + \gamma E[\log(C_{t+1}/C_t)] - \frac{\gamma^2}{2} \text{Var}[\log(C_{t+1}/C_t)]$.

²² Further details about the GMM-AR test are provided in the Internet Appendix.

²³ As a robustness check, we restrict the sample to the 1960 to 2006, as in Savov (2011). The resulting GMM-AR 95% confidence sets for γ are also unbounded. Similarly, we get unbounded 95% confidence sets for γ using the quarterly consumption growth measures from Parker and Julliard (2005) and Lettau and Ludvigson (2001).

Figure 9. Joint 95% confidence sets for (γ, δ) . In this figure, we test for values of the relative rate of risk aversion γ within the range of $[0, 300]$ and of the discount factor δ in $[0.5, 1]$ using the GMM-AR test from Stock and Wright (2000). See the Internet Appendix for a detailed description of the test statistic. The moment conditions are $E_t[\delta(\frac{C_{t+1}}{C_t})^{-\gamma} R_{m,t+1}^e] = 0$ and $E_t[\delta(\frac{C_{t+1}}{C_t})^{-\gamma} R_{f,t+1}] = 1$. The shaded areas constitute the 95% joint confidence sets for γ and δ . The excess market return and the gross risk-free rate used for the test are for the period 1960 to 2014 (yearly data).

where $R_{f,t+1}$ is the gross risk-free rate; see, for example, Savov (2011) and Kroencke (2017). Combining equations (28) and (29) provides two restrictions, so we can now estimate both γ and δ and no longer have to pin down δ to 0.95. We therefore consider (γ, δ) as a pair of parameters and use the GMM-AR test to conduct joint tests on them.

The joint two-dimensional confidence sets for γ and δ that result from the GMM-AR test are shown for the five consumption measures in Figure 9. The two-dimensional confidence sets consist of all values of (γ, δ) for which the GMM-AR statistic testing (28) and (29) at these specific values does not exceed the appropriate critical value of the χ_2^2 distribution. In Figure 9, we focus on γ in the $[0, 300]$ range and δ in the $[0.5, 1]$ range, since these ranges are most relevant in practice. The figure shows that all five consumption measures induce large γ s, as shown by the (shaded) 95% confidence sets, so none of these measures has strong informational content for the relative rate of risk aversion.

C. Testing Linear or Nonlinear SDFs: What is the Difference?

Because of the limited time spans of the series we consider, we refrain from using larger choices of N , such as $N = 5, 10$, or 25 , since the GMM-AR test is

unreliable for such settings.²⁴ This therefore raises the following question: if researchers use the same assets to test the risk premium in linear models and relative risk aversion in nonlinear models, what should we expect?

The linear and nonlinear SDF moment conditions specified in (5) and (28) allow the parameters of interest, such as the risk premia and relative risk aversion, to be estimated without needing to fully specify the DGP. Thus, tests of these moment conditions, for example, the GRS-FAR and GMM-AR tests, do not test the full specification of the DGP. DGPs therefore exist that match or approximately match both of the moment conditions from linear and nonlinear SDFs. We next show this for two DGPs. In the first, both the linear moment conditions (5) and the nonlinear CRRA moment conditions (see, e.g., Cochrane (2001)),

$$E_t \left[\delta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} (\iota_N + R_{t+1}) \right] = \iota_N, \quad (30)$$

hold. In the second, the nonlinear SDF moment condition (30) holds and the linear SDF moment condition (5) approximately holds.

- (1) The returns on the test assets and log-consumption growth are jointly normal distributed with mean asset returns such that both the linear SDF and nonlinear CRRA moment conditions (5) and (30) hold. The zero-beta return and risk premium for the linear SDF then imply that the relative risk aversion and discount factor in the CRRA SDF are (see the Internet Appendix, *DGP: normal returns*)

$$\gamma = \lambda_f / \text{var}(\Delta c_t), \quad \text{and} \quad \delta = e^{-\gamma^2 \text{var}(\Delta c_t)/2} / (1 + \lambda_0), \quad (31)$$

where Δc_t is log-consumption growth at time t .

- (2) The log-returns on the test assets and log-consumption growth are jointly normal distributed with the CRRA moment condition (30) holding. The linear SDF moment condition (5) then approximately holds with the same relationship between the risk premium and the relative risk aversion parameter as in (31) and the zero-beta return is equal to (see the Internet Appendix, *DGP: log-normal returns*)

$$\lambda_0 = -\ln(\delta) - \frac{\gamma^2 \text{var}(\Delta c_t)}{2}. \quad (32)$$

To further illustrate the above two cases, we compute power curves of the GRS-FAR and GMM-AR tests using DGPs that are calibrated to data from Kroencke (2017). Figure 10 uses a DGP that accords with setting 1 above, so both the linear and nonlinear SDF hold, while Figure 11 accords with setting 2

²⁴ The GMM-AR test is unreliable since the asymptotic distribution of the GMM-AR statistic does not provide an accurate approximation of its finite-sample one for such values of N in conjunction with the limited time spans. See the Internet Appendix.

Figure 10. Power curves of GRS-FAR that tests $H_0 : \lambda_f = 2$ and GMM-AR that tests $H_0 : \gamma = 2/\text{var}(\Delta c_t)$ at the 5% level, DGP: normal returns. This figure plots power curves of GMM-AR (solid) and GRS-FAR (dashed) at the 5% significance level. The moment condition imposed on the DGP is $E(R_t) = \iota_N \lambda_0 + \beta \lambda_f$. The null hypothesis is $H_0 : \lambda_f = 2$ for GRS-FAR and $H_0 : \gamma = 2/\text{var}(\Delta c_t)$ for GMM-AR. The horizontal axes show the value of λ_f used in the DGP (*DGP: normal returns* in the Internet Appendix). The DGP is calibrated to the test assets and the unfiltered consumption growth from Kroencke (2017). The $N = 5, 10$, or 25 test assets used for calibration are taken from the 31 portfolios in their order starting from the market portfolio. The number of Monte Carlo replications is 5,000.

above, so the nonlinear SDF holds while the linear SDF approximately holds. When using the GMM-AR test, both figures show the need for a large time span compared to the number of test assets for the test to be reliable in the sense that its rejection frequency under the null hypothesis corresponds to the significance level of the test. We use F critical values for the GRS-FAR test in Figures 10 and 11. Neither figure shows any overrejection under the null

Figure 11. Power curves of GMM-AR that tests $H_0 : \gamma = 15$ and GRS-FAR that tests $H_0 : \lambda_f = 15 \times \text{var}(\Delta c_t)$ at the 5% level. DGP: log-normal returns. This figure plots power curves of GMM-AR (solid) and GRS-FAR (dashed) at the 5% significance level. The moment condition imposed on the DGP is $E_t[\delta(\frac{C_{t+1}}{C_t})^{-\gamma}(\iota_N + R_{t+1})] = \iota_N$ with $\delta = 0.95$. The null hypothesis is $H_0 : \gamma = 15$ for GMM-AR and $H_0 : \lambda_f = 15 \times \text{var}(\Delta c_t)$ for GRS-FAR. The horizontal axes show the value of γ used in the DGP (*DGP: log-normal returns* in the Internet Appendix). The DGP is calibrated to the test assets and the unfiltered consumption growth from Kroencke (2017). The $N = 5, 10$, or 25 test assets used for calibration are taken from the 31 portfolios in their order starting from the market portfolio. The number of Monte Carlo replications is 5,000.

hypothesis, even in Figure 11 where the implied (approximately) linear factor model does not have normally distributed errors. This result indicates that use of (misspecified) F critical values for the GRS-FAR test can render it reliable in settings where the time span is not much larger than the number of test assets.

Figures 10 and 11 show that both the GMM-AR and GRS-FAR tests have rejection frequencies close to the significance level when the time span is much

larger than the number of test assets. They also reveal differences in discriminatory power of the two tests that are somewhat difficult to interpret because of the rescaling of the relative risk aversion and risk premium parameters needed to make the power curves comparable. The most important observation from these two figures is that there is little independent information in the GRS-FAR test of the linear SDF (5) compared to the GMM-AR test of the CRRA SDF (30) for the DGPs considered. Put differently, it is not surprising that researchers may get qualitatively similar findings when testing linear and nonlinear models with the same test assets using GRS-FAR and GMM-AR tests, respectively. Thus, nonrejection of a specific SDF by either one of the two tests does not imply that another SDF is rejected.²⁵

IV. Conclusions

The prototypical correlations between macro-risk factors and asset returns and the number of time series observations compared to the number of assets are such that traditional asset pricing tests on risk premia and other parameters of SDFs are not reliable. We propose GRS-based inference procedures for analyzing parameters of linear asset pricing models that can address such settings. We find 95% confidence sets that are generally very wide or even unbounded when the time series dimension is not much larger than the cross-section dimension and that become empty when the time series dimension is much larger. We reconcile these findings using a simulation experiment with a DGP in which there is misspecification. When expanding the time series dimension while keeping the sufficiently large number of cross-section observations fixed, the simulated 95% confidence sets move from being primarily unbounded for smaller numbers of time series observations to more bounded and eventually empty for larger numbers of time series observations.

For nonlinear consumption-based asset pricing models, we use a GMM test procedure that can address the small correlations but needs many more time series observations than the number of assets for the asymptotic approximation of its finite-sample distribution to be accurate. Hence, we typically have to restrict attention to just a small number of assets to obtain reliable inference. As in the case of the risk premia for linear factor models when the number of time series observations is small, we find unbounded 95% confidence sets primarily for the parameters of nonlinear asset pricing models like the relative rate of risk aversion.

The empirical findings of this paper show that meaningful inference is infeasible if the data are too noisy and the time series dimension is too short compared to the cross-section dimension. With such limited information in the data, we cannot conclude whether the consumption growth factor explains part of the variation in the cross-section of expected asset returns, nor can we reject the possibility that consumption growth explains all of the variation.

²⁵ Since the GMM misspecification J -statistic equals the GMM-AR and GRS-FAR tests at their minimal value, this holds for the GMM J -test as well.

Practitioners, however, can use our proposed rank test to determine whether there is sufficient information in the data to identify risk premia. If this rank test rejects no identification of the risk premia, then our proposed GRS-FAR test provides an informative bounded confidence set for the risk premia. The GRS-FAR test is a weak identification-robust test procedure that extends the weak identification-robust Anderson-Rubin test to a finite-sample setting in asset pricing models. In Kleibergen, Kong, and Zhan (2019), extensions of other weak identification-robust tests, like, the Lagrange multiplier and likelihood ratio tests (see, e.g., Kleibergen (2002, 2005) and Moreira (2003)) to this setting are discussed.

Overall, our findings call for the use of identification-robust inference methods in empirical asset pricing studies. Together with consumption measures of better quality and more time series observations, these robust methods will provide a more reliable assessment of consumption-based asset pricing models in future research.

Initial submission: December 8, 2017; Accepted: March 22, 2019
 Editors: Stefan Nagel, Philip Bond, Amit Seru, and Wei Xiong

REFERENCES

- Anatolyev, Stanislav, and Nikolay Gospodinov, 2011, Specification testing in models with many instruments, *Econometric Theory* 27, 427–441.
- Anderson, Theodore W., and Herman Rubin, 1949, Estimation of the parameters of a single equation in a complete system of stochastic equations, *Annals of Mathematical Statistics* 20, 46–63.
- Beaulieu, Marie-Claude, Jean-Marie Dufour, and Lynda Khalaf, 2013, Identification-robust estimation and testing of the zero-beta CAPM, *Review of Economic Studies* 80, 892–924.
- Bekker, Paul A., 1994, Alternative approximations to the distributions of instrumental variable estimators, *Econometrica* 62, 657–681.
- Breeden, Douglas T., 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, *Journal of Financial Economics* 7, 265–296.
- Breeden, Douglas T., Michael R. Gibbons, and Robert H. Litzenberger, 1989, Empirical tests of the consumption-oriented CAPM, *Journal of Finance* 44, 231–262.
- Campbell, John Y., Andrew Wen-Chuan Lo, and Archie Craig MacKinlay, 1997, *The Econometrics of Financial Markets*, volume 2 (Princeton University Press, Princeton, NJ).
- Cochrane, John H., 2001, *Asset Pricing* (Princeton University Press, Princeton, NJ).
- Dufour, Jean-Marie, 1997, Some impossibility theorems in econometrics with applications to structural and dynamic models, *Econometrica* 65, 1365–1387.
- Dufour, Jean-Marie, and Mohamed Taamouti, 2005, Projection-based statistical inference in linear structural models with possibly weak instruments, *Econometrica* 73, 1351–1365.
- Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.
- Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.
- Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.
- Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2014, Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors, *Review of Financial Studies* 27, 2139–2170.
- Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2017, Spurious inference in reduced-rank asset-pricing models, *Econometrica* 85, 1613–1628.

- Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2018, Asymptotic variance approximations for invariant estimators in uncertain asset-pricing models, *Econometric Reviews* 37, 695–718.
- Hansen, Lars Peter, 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–1054.
- Huberman, Gur, Shmuel Kandel, and Robert F. Stambaugh, 1987, Mimicking portfolios and exact arbitrage pricing, *Journal of Finance* 42, 1–9.
- Jagannathan, Ravi, and Yong Wang, 2007, Lazy investors, discretionary consumption, and the cross-section of stock returns, *Journal of Finance* 62, 1623–1661.
- Jarque, Carlos M., and Anil K. Bera, 1987, A test for normality of observations and regression residuals, *International Statistical Review / Revue Internationale de Statistique* 55, 163–172.
- Kan, Raymond, Cesare Robotti, and Jay Shanken, 2013, Pricing model performance and the two-pass cross-sectional regression methodology, *Journal of Finance* 68, 2617–2649.
- Kan, Raymond, and Chu Zhang, 1999a, Two-pass tests of asset pricing models with useless factors, *Journal of Finance* 54, 203–235.
- Kan, Raymond, and Chu Zhang, 1999b, GMM tests of stochastic discount factor models with useless factors, *Journal of Financial Economics* 54, 103–127.
- Kan, Raymond, and Guofu Zhou, 2004, Hansen-Jagannathan distance: Geometry and exact distribution, Working paper, University of Toronto.
- Khalaf, Lynda, and Huntley Schaller, 2016, Identification and inference in two-pass asset pricing models, *Journal of Economic Dynamics and Control* 70, 165–177.
- Kleibergen, Frank, 2002, Pivotal statistics for testing structural parameters in instrumental variables regression, *Econometrica* 70, 1781–1803.
- Kleibergen, Frank, 2005, Testing parameters in GMM without assuming that they are identified, *Econometrica* 73, 1103–1123.
- Kleibergen, Frank, 2009, Tests of risk premia in linear factor models, *Journal of Econometrics* 149, 149–173.
- Kleibergen, Frank, Lingwei Kong, and Zhaoguo Zhan, 2019, Identification-robust testing of risk premia in finite samples, Working paper, University of Amsterdam.
- Kleibergen, Frank, and Zhaoguo Zhan, 2015, Unexplained factors and their effects on second pass R-squared's, *Journal of Econometrics* 189, 101–116.
- Kleibergen, Frank, and Zhaoguo Zhan, 2018, Identification-robust inference on risk premia of mimicking portfolios of nontraded factors, *Journal of Financial Econometrics* 16, 155–190.
- Kroencke, Tim A., 2017, Asset pricing without garbage, *Journal of Finance* 72, 47–98.
- Lettau, Martin, and Sydney Ludvigson, 2001, Resurrecting the (C) CAPM: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238–1287.
- Lewellen, Jonathan, Stefan Nagel, and Jay Shanken, 2010, A skeptical appraisal of asset pricing tests, *Journal of Financial Economics* 96, 175–194.
- Moreira, Marcelo J., 2003, A conditional likelihood ratio test for structural models, *Econometrica* 71, 1027–1048.
- Newey, Whitney K., and Frank Windmeijer, 2009, GMM with many weak moment conditions, *Econometrica* 77, 687–719.
- Parker, Jonathan A., and Christian Julliard, 2005, Consumption risk and the cross section of expected returns, *Journal of Political Economy* 113, 185–222.
- Savov, Alexi, 2011, Asset pricing with garbage, *Journal of Finance* 66, 177–201.
- Shanken, Jay, 1992, On the estimation of beta-pricing models, *Review of Financial Studies* 5, 1–33.
- Stock, James H., and Jonathan H. Wright, 2000, GMM with weak identification, *Econometrica* 68, 1055–1096.
- Stock, James H., and Motohiro Yogo, 2005, Testing for weak instruments in Linear IV regression, in D. W. Andrews, and J. H. Stock, eds.: *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, 80–108 (Cambridge University Press, Cambridge).
- Wright, Jonathan H., 2003, Detecting lack of identification in GMM, *Econometric Theory* 19, 322–330.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: Internet Appendix.
Replication code.