Journal of Banking and Finance 162 (2024) 107123

Contents lists available at ScienceDirect

Journal of Banking and Finance

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

GMM weighting matrices in cross-sectional asset pricing tests ☆

Nora Laurinaityte [a], Christoph Meinerding [b,*], Christian Schlag [c], Julian Thimme [d]

[a] Lithuanian Green Finance Institute at INVEGA and Vilnius University, Sauletekio avenue 9, 10222 Vilnius, Lithuania
[b] Deutsche Bundesbank, Research Centre, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt, Germany
[c] Goethe University Frankfurt and Leibniz Institute for Financial Research SAFE, Theodor-W.-Adorno-Platz 3, 60323 Frankfurt, Germany
[d] Karlsruhe Institute of Technology, Kaiserstrasse 12, 76131 Karlsruhe, Germany

ARTICLE INFO

JEL classification:
G00
G12
C21
C13

Keywords:
Asset pricing
Cross-section of expected returns

ABSTRACT

When estimating misspecified linear factor models for the cross-section of expected returns using GMM, the explanatory power of these models can be spuriously high when the estimated factor means are allowed to deviate substantially from the sample averages. In fact, by shifting the weights on the moment conditions, any level of cross-sectional fit can be attained. The mathematically correct global minimum of the GMM objective function can be obtained at a parameter vector that is far from the true parameters of the data-generating process. This property is not restricted to small samples, but rather holds in population. It is a feature of the GMM estimation design and applies to both strong and weak factors, as well as to all types of test assets.

1. Introduction

A multitude of tools have been devised to assess factor models' ability to explain cross-sectional variation in expected asset returns. Besides the Fama-MacBeth two-pass regression approach and its extensions, tests of factor models using the generalized method of moments (GMM) are popular, especially in the presence of nontradable factors motivated from macrofinancial equilibrium asset pricing models. In this paper, we show that a prominent GMM estimator is extremely sensitive to the choice of the weighting matrix if the tested asset pricing model is misspecified. It can produce biased parameter estimates and strongly inflated model performance statistics, leading applied researchers to falsely conclude that their model is helpful in understanding asset returns. As a consequence, we recommend against using this estimator. Researchers should instead rely on alternative misspecification-robust inference procedures. If they still wish to use the approach criticized here, we discuss several ways to detect and alleviate the problem within the presented framework in Section 4.

The estimator was proposed by Cochrane (2005), first prominently used by Yogo (2006), and afterwards adopted by many researchers, for example Darrat et al. (2011), Grammig et al. (2009), Maio and Santa-Clara (2012), Maio (2013a,b), Lioui and Maio (2014), Tedongap (2015), Da et al. (2016), and Chen and Lu (2017). It also features as a benchmark estimator in Penaranda and Sentana (2015) and Manresa et al. (2023) and in the literature on currency excess returns and carry trades (see, e.g., Menkhoff et al. (2012), Menkhoff et al. (2013), Della Corte et al. (2016a), Della Corte et al. (2016b), Bekaert and Panayotov (2020), Della Corte et al. (2022)). The estimator employs two sets of moment conditions: The first is given by the pricing errors of the test assets, formally $0 = E[R^e - R^e(F-\mu)\lambda] = E[R^e\{1-(F-\mu)\lambda\}]$. Since the term

☆ The authors would like to thank Geert Bekaert, Winston Dou, Andrei Goncalves, Campbell Harvey, Stephan Jank, Malte Knüppel, Emanuel Moench, Andreas Neuhierl, Nick Roussanov, Jantje Soenksen, participants of the MFA Conference 2020, the Econometric Society Winter Meeting 2019, the SGF Conference 2018 in Zurich, the Frontiers of Factor Investing Conference 2018 in Lancaster and seminar participants at KIT, Goethe University Frankfurt and the Deutsche Bundesbank for their valuable comments and suggestions. A previous version of this paper was entitled "Elephants and the cross-section of expected returns." This work represents the authors' personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank or the Eurosystem. Nora Laurinaityte and Christian Schlag gratefully acknowledge research and financial support from the Research Center SAFE, funded by the State of Hesse's initiative for research LOEWE, and later from the Leibniz Institute for Financial Research SAFE. The MATLAB files for replication are available at www.julianthimme.de/code.

* Corresponding author.
E-mail addresses: nora.laurinaityte@gmail.com (N. Laurinaityte), christoph.meinerding@bundesbank.de (C. Meinerding), schlag@finance.uni-frankfurt.de (C. Schlag), thimme@kit.edu (J. Thimme).

https://doi.org/10.1016/j.jbankfin.2024.107123
Received 29 November 2021; Accepted 27 February 2024
Available online 6 March 2024

in curly brackets can be interpreted as a (normalized) stochastic discount factor, the estimator is sometimes referred to as the *centered SDF approach*. These moment conditions imply that expected excess returns must be linear in the covariance between excess returns R^e and candidate factors F. However, the market price of risk (MPR) λ is not identifiable from this condition alone. Instead, it can only be identified jointly with the set of factor means μ. A second set of moment conditions of the form $0 = E[F - \mu]$ is added, recognizing sampling variation from estimating factor means.

Within the GMM framework, a common approach is to minimize the quadratic form $g_T' W g_T$, where g_T denotes the vector of sample averages of the moment conditions and W is a weighting matrix, assigning relative importance to (combinations of) the different moment conditions. We show that too low a weight on the second set of the above moment conditions leads to a weakly identified model, resulting in an overall incorrect inference. The wedge between the parameters minimizing the GMM objective function and those governing the true data-generating process can be substantial, even for standard choices of the weighting matrix. Consequently, many results in the papers mentioned above are contaminated by this bias. Intuitively, minimizing the quadratic form $g_T' W g_T$ leads to a situation in which μ is not identified by the moment condition $0 = E[F - \mu]$ alone, but the estimator also seeks to minimize the pricing errors through the choice of μ. When estimating a misspecified model, the cost of not matching $0 = E[F - \mu]$ can be lower in relative terms than the cost of having high pricing errors.

It is important to precisely pinpoint the extent of the problem. Specifically, a related, but different approach is presented by Burnside (2011). He discusses solutions of $a_T g_T = 0$ for generalized weighting schemes a_T, which are more flexible than the common “least squares” approach. The particular scheme a_T that he favors fixes the estimate of μ to the sample average of the factor and thus avoids the bias that we describe. Many papers in the literature that report GMM estimates for factor models only outline the moment conditions and do not explicitly state whether they use the least squares weighting scheme (which is susceptible to the bias) or the generalized Burnside (2011) scheme. For instance, the literature on currency excess returns and carry trades cited above employs a range of variations of these approaches, also in combination with (what seem to be) ad hoc fixes to the problem outlined in our paper, e.g., exogenously specified extreme weighting matrices.

In theory, the choice of the weighting scheme is innocuous, as long as the model is correctly specified. In applications, however, researchers always have to deal with misspecified models. In light of this, they do not make any statement as to whether a model is correct, i.e., whether the factor exposures fully explain the cross-sectional variation in expected returns. Instead, they analyze what fraction of this variation is explained by the model. In more technical terms, with a misspecified model there is no vector of parameters that sets all GMM moment conditions *identically equal to zero*. This means that Assumption 2.4 in Hansen (1982) is violated in the sense that not even in population are all moment conditions equal to zero. The minimum of the GMM objective function $g_T' W g_T$ is greater than zero and obtained at a parameter vector that can be far from the parameters of the true data-generating process. We show in a controlled simulation environment that the estimator misestimates the factor means in favor of matching the Euler equations.

The resulting bias can materialize in two ways: First, unpriced factors can appear priced, i.e., the estimated market price of risk (MPR) can be large (and significant) despite the true MPR being close to zero. The corresponding estimated pricing performance can be heavily overstated in the sense that pricing errors are small and cross-sectional R^2's are large. We show that the estimated R^2 can take on any value between the true cross-sectional R^2 and 1, depending on the choice of the weighting matrix. Importantly, the bias is substantial even for strong factors, i.e., factors that are strongly correlated with the returns of the test assets in the time series. Second, when estimating a linear factor model with multiple factors, relatively weaker and unpriced factors can “drive out” stronger priced factors. More precisely, MPR estimates of weaker factors can be large and significant, while those of stronger priced factors are biased downwards, i.e., small and insignificant.

Importantly, “relatively weaker” does not mean that any of the factors are necessarily weak or useless. There is a large amount of literature dealing with the econometric details of cross-sectional asset pricing with weak factors, e.g., Kan and Zhang (1999a,b), Stock and Wright (2000), Kleibergen (2009), Gospodinov et al. (2014), Kleibergen and Zhan (2015, 2020), Bryzgalova (2016), Gospodinov et al. (2019) or Burnside (2016). However, this literature is unrelated to the problem documented in our work, which applies to all factors irrespective of their strength.

Some papers argue that broadening the set of test portfolios mitigates some econometric issues of asset pricing tests (see, e.g., Lewellen et al., 2010). Interestingly, the problem discussed in our paper is exacerbated when the number of test assets is increased. If we add more test assets, ceteris paribus the overall weight on the Euler equation moment conditions in the GMM objective function increases relative to the moment conditions identifying the factor means (whose number does not change). Therefore, the estimator prioritizes the matching of the Euler equations more strongly at the cost of not matching the factor means.

Generally, parameter identification in GMM settings is not just a feature of the factor (Kan and Zhang, 1999a) or of weakly correlated test assets (Giglio et al., 2022), but also of the weighting matrix applied in the estimation. This general point has not been a particular point of focus in the literature. An exception is the application to continuously updated estimators in Gospodinov et al. (2017). We start our analysis with prespecified weighting matrices but show that the problem also carries over to endogenous ways of selecting the weights.

Finally, what is the practical relevance of the issue highlighted in this paper? Essentially, all prominent cross-sectional asset pricing tests have been subject to critique, highlighting shortcomings in particular situations. As a consequence, empirical asset pricing papers rarely show results from only one cross-sectional test to make a case for their preferred factor model. Clearly, running several tests – be it two-pass regressions, GMM, or portfolio sorts – makes biases in one of these methods more easily detectable. Still, the GMM estimator discussed here represents a popular choice, as recently highlighted by Manresa et al. (2023). Browsing through the literature, we find that it is often applied – even in isolation – in papers that empirically investigate equilibrium asset pricing models (or linearized versions thereof) featuring nontraded factors.

The popularity of the GMM estimator in this strand of the literature may result from the fact that it is very lean, compared to, for example, Fama-MacBeth regressions. It prescribes one moment condition per factor to identify the factor means μ and one moment condition per test asset to identify the factor risk premia λ. To give an example, when estimating a three factor model with 25 test assets, the approach discussed in our paper employs 28 moment conditions to estimate six parameters. In particular, it does not require the separate identification of factor exposures (“betas”). In the above example, the Fama and MacBeth (1973) or Cochrane (2005) approach would estimate 103 parameters (25 “alphas”, 3×25 “betas”, and three “lambdas”) with 125 moment conditions (100 exogeneity conditions corresponding to OLS time series regressions and 25 conditions corresponding to cross-sectional regressions (see Cochrane (2005), Section 12.2)). Keeping the number of moment conditions small, relative to the sample size, is crucial to obtain reliable estimates in GMM settings (see, e.g., Newey and Windmeijer, 1994). Moreover, Fama-MacBeth regressions come with several other empirical challenges, which have been discussed extensively in the literature.

Our paper is structured as follows. Section 2 explains the bias of the estimator and also points out why it is different from the approach discussed in Burnside (2011). We then analyze the extent of the bias in a controlled environment using simulated data. Section 3 considers several examples with real data. We start by considering the Fama and

French (1993) three-factor model and augment it with an obviously meaningless factor. We then study prominent examples from the asset pricing literature: the models of Yogo (2006), Parker and Julliard (2005), Kroencke (2017), He et al. (2017), and Maio and Santa-Clara (2012). Some papers (rather indirectly) point towards the issue discussed in our paper, e.g., Parker and Julliard (2005), Savov (2011), or Delikouras and Kostakis (2019). Instead of providing an in-depth analysis of the fundamental problem that we are focusing on, they suggest ad hoc procedures to robustify their empirical findings, such as fixing betas to their OLS counterparts. Other papers do not seem to be aware of the problem. The most prominent example is Yogo (2006), which we investigate in greater detail.[1]

Finally, Section 4 provides suggestions for applied researchers on how to address the issue discussed here. As emphasized above, we recommend avoiding the use of the estimator discussed here and instead relying on alternatives. Kroencke and Thimme (2023) discuss several such alternative estimators and their properties in small samples. Should researchers still wish to use the GMM estimator discussed here, we elaborate on how they can diagnose and mitigate the bias. In short, one should always report point estimates for the factor means and compare them to estimates from other methods. For traded factors, it is crucial to consider that the MPR and the factor means are identical in theory.

Our paper is accompanied by an online appendix. Section A of this appendix discusses and illustrates the bias theoretically for several configurations of factors and test assets. Section B provides details on the simulation study, while Section C discusses papers in the literature that use the estimation approach we criticize. We do not provide replications of these papers because replication code is not available. Still, the results in these papers should be interpreted with caution in light of our findings. Section D provides further information regarding the results of Yogo (2006). The replication files for our paper can be downloaded from www.julianthimme.de/code.

2. Cross-sectional regressions with GMM

2.1. The problem

Unconditional linear asset pricing models imply that the expected excess return on asset $i = 1, \dots, n$ is proportional to the covariances of its excess return with a group of risk factors F_j $(j = 1, \dots, k)$:

$$E[R_i^e] = \sum_{j=1}^{k} Cov(R_i^e, F_j)\lambda_j^* \quad \Longleftrightarrow \quad 0 = E\left[R_i^e - \sum_{j=1}^{k} R_i^e(F_j - E[F_j])\lambda_j^*\right]. \tag{1}$$

Since the true factor means $E[F_j]$ are typically unknown, a common approach is to estimate the (normalized) MPRs λ_j^* and the factor means jointly using a GMM estimator with the moment conditions[2]

$$g_T(\lambda, \mu) = E_T\left[\begin{array}{ll} R_i^e - \sum_{j=1}^{k} R_i^e(F_j - \mu_j)\lambda_j, & i = 1, \dots, n \\ F_j - \mu_j, & j = 1, \dots, k \end{array}\right]. \tag{2}$$

In the first n moment conditions, the returns can be factored out. The conditions are thus equivalent to $E_T[R_i^e(1 - \sum_{j=1}^{k}(F_j - \mu_j)\lambda_j)] = 0$, which represents an Euler equation with stochastic discount factor $(1 - \sum_{j=1}^{k}(F_j - \mu_j)\lambda_j)$. Therefore, this estimator is sometimes referred to as the *centered SDF approach* (see Penaranda and Sentana, 2015).

Whenever we have more test assets than factors, i.e. $n > k$, the system is overidentified and we have to select a weighting scheme, which is specified as a $2k \times (n + k)$ matrix a_T. The parameter estimates are then given by the solution of $a_T g_T(\lambda, \mu) = 0$. The most common weighting schemes are of the type $a_T = \frac{\partial g_T'}{\partial \theta} W$, where $\theta = (\lambda, \mu)'$ is the vector of parameters, and W is referred to as the weighting matrix. With this weighting scheme, the GMM estimation boils down to minimizing the quadratic form $g_T' W g_T$.

Our paper is concerned with the impact of the weighting matrix W on the point estimates of λ and μ in the context of this least squares GMM approach. We discuss an alternative choice of the weighting scheme a_T, not corresponding to least squares, in Section 2.2. Importantly, in the knife-edge case where the model is correctly specified (which is obviously never true in empirical applications), i.e., when $g_T(\lambda, \mu) \to 0$ as $T \to \infty$ for the unique parameter vector $(\lambda, \mu) = (\lambda^*, E[F])$, the choice of the weighting scheme does not matter for the point estimates.

Assume now that the tested factor model is misspecified, such that expected returns are not proportional to the factor exposures, i.e., there are pricing errors e_i $(i = 1, \dots, n)$, given as

$$e_i = E\left[R_i^e - \sum_{j=1}^{k} R_i^e(F_j - E[F_j])\lambda_j^*\right], \tag{3}$$

which is nonzero for at least one test asset. For example, there could be an omitted factor F_{om} and $e_i = E[R_i^e(F_{om} - E[F_{om}])\lambda_{om}^*]$ with $\lambda_{om}^* \neq 0$. Then there is no parameter vector (λ, μ) setting all moment conditions in (2) equal to zero. In such situations, the parameter estimates depend on the weights assigned to the conditions in the estimation. Still, our goal is to estimate the parameters $\lambda = \lambda^*$ and $\mu = E[F]$ and report informative model performance statistics, for example $R^2 = 1 - \sum_i e_i^2 / \sum_i E[R_i^e]^2$ or $RMSE = \sqrt{\frac{1}{n}\sum_i e_i^2}$.

To illustrate the impact of the weighting matrix on estimates and model performance statistics, consider the extreme case where the weights on the moment conditions $E_T[F_j - \mu_j]$ are equal to zero, i.e., the lower right submatrix of W is equal to $0_{k \times k}$. The two sample moment conditions in Equation (2) then effectively reduce to only one:

$$\begin{aligned} g_T(\lambda, \mu) &= E_T[R_i^e - \sum_{j=1}^{k} R_i^e(F_j - \mu_j)\lambda_j] \\ &= E_T[R_i^e] - \sum_{j=1}^{k} E_T[R_i^e(F_j - E_T[F_j])]\lambda_j \\ &\quad - E_T[R_i^e] \sum_{j=1}^{k}(E_T[F_j] - \mu_j)\lambda_j. \end{aligned} \tag{4}$$

The reformulation in Equation (4) illustrates a simple way of minimizing the moment condition. The estimator can single out one factor j^*, set $\mu_{j^*} = E[F_{j^*}] - \lambda_{j^*}^{-1}$ and let λ_{j^*} become very small (basically equal to 0). Setting $\lambda_j = 0$ and μ_j to any arbitrary value for all other $j \neq j^*$ then makes $g_T(\lambda, \mu)$ approach zero, corresponding to the minimum of the GMM objective function. This is true for any choice of j^*: The second term on the right-hand side of Equation (4) goes to zero since all summands except one are equal to zero, and the summand for j^* is arbitrarily close to zero as λ_{j^*} becomes very small. Similarly, the third term converges to $E_T[R_i^e]$. As a consequence, the moment condition (4) converges to zero *for all test assets*.

In the following, we label such parameter estimates as *trivial solutions*. They are solutions of the optimization problem because they are global minima (one per candidate factor) of the GMM objective function. They are trivial in the sense that they set all moment conditions to zero, irrespective of the properties of the factors. The above argument

[1] Yogo (2006) also contains a non-linear model from which the linear model is derived by log-linearization. The non-linear model was recently criticized by Borri and Ragusa (2017). They find that the model in general has a hard time explaining the interest rate and the equity premium simultaneously. This result, however, is completely independent of the failure of the linear factor model that we document.

[2] Throughout the paper, we use the notation of Hansen (1982) in which a subscript T denotes the sample equivalent of a given moment.

holds irrespective of the strength of the factors (referring to the strength of the time series correlation of the factor with the test asset returns) and for priced and unpriced factors (referring to the cross-sectional correlation between expected returns and time series covariances).

Assigning a positive weight to the moment conditions $E_T[F_j - \mu_j]$ makes trivial solutions "costly", since $\mu_{j*} = E[F_{j*}] - \lambda_{j*}^{-1}$ with $\lambda_{j*} \to 0$ drives μ_{j*} far from $E[F_{j*}]$. If the weight goes to infinity, the estimates of μ_j are fixed to $E_T[F_j]$.[3] We refer to this solution as the *desired solution*. It is desired in the sense that we want μ_j to be close to $E_T[F_j]$ and to not give the estimator the additional degree of freedom to just match Euler equations and produce a trivial solution.

2.2. Finite weights

As a simple starting point, we analyze weighting matrices W_x of the form

$$W_x = \begin{pmatrix} I_n & 0 \\ 0 & 10^x I_k \end{pmatrix} = \mathrm{diag}(1, \dots, 1, 10^x, \dots, 10^x). \quad (5)$$

The two extreme cases $x \to -\infty$ and $x \to \infty$, which lead to trivial and desired solutions, respectively, have been discussed in Section 2.1 above. For simplicity, assume that there is a single factor F. All formulas presented in the following naturally generalize to the case of multiple factors. The gradient $\frac{\partial g_T}{\partial \theta}$ with $\theta = (\lambda, \mu)$ is given by

$$\frac{\partial g_T'}{\partial \theta} = \begin{pmatrix} E_T[R_1^e(\mu - F)] & \dots & E_T[R_n^e(\mu - F)] & 0 \\ E[R_1^e]\lambda & \dots & E_T[R_n^e]\lambda & -1 \end{pmatrix}. \quad (6)$$

Given the choice of the weighting matrix W_x in (5), the estimator thus solves the two non-linear equations

$$\sum_{i=1}^{n} E_T[R_i^e(F-\mu)] \cdot E_T[R_i^e - R_i^e(F-\mu)\lambda)] = 0 \quad (7)$$

$$\left[\sum_{i=1}^{n} E_T[R_i^e]\lambda \cdot E_T[R_i^e(1-(F-\mu)\lambda)]\right] - 10^x(E_T[F]-\mu) = 0 \quad (8)$$

for λ and μ. For the two scenarios $x \to \pm\infty$, the solutions were already discussed in Section 2.1 above. The solutions for finite weights x lie in between these two extreme cases. We will analyze them in detail in Section 2.3.

An alternative to the least squares weighting scheme $a_T = \frac{\partial g_T'}{\partial \theta} W$ is discussed by Burnside (2011, Section 2.2 of the online appendix). He suggests to use

$$a_T = \begin{pmatrix} E_T[R_1^e(\mu - F)] & \dots & E_T[R_n^e(\mu - F)] & 0 \\ 0 & \dots & 0 & -1 \end{pmatrix}.$$

This boils down to setting the lower left submatrix of (6) equal to zero. With this choice, Equation (8) simplifies to $E_T[F] - \mu = 0$. The estimator thus effectively estimates μ *separately* from λ. It sets $\hat{\mu} = E_T[F]$ and then selects the value of λ that solves Equation (7), given $\hat{\mu}$. In this sense, it can be interpreted as a two-stage approach and, indeed, Burnside (2011) shows that the point estimates are identical to those of a classic Fama and MacBeth (1973) two-stage regression.

In contrast, the least squares GMM approach that we analyze takes the pricing errors into account when estimating μ. It is this decisive feature that creates the tension between minimizing pricing errors and estimating factor means. The GMM estimator searches for parameters that minimize all moment conditions jointly, and the pricing errors can also provide information about the factor means. The weighting matrix W, which determines the relative importance of pricing errors and factor means, then determines whether the solution is closer to a trivial or to a desired solution. The literature on currency excess returns and carry trades applies the moment conditions outlined in this paper extensively. However, some authors in this literature do not state explicitly whether they use the standard least squares weighting scheme or the enhanced weighting scheme of Burnside (2011). Our paper tries to clear up the resulting confusion and shows what can go wrong when the least squares approach is not applied carefully. As we document in detail in Section 2.3 below, even for seemingly innocuous choices of the weighting matrix, the estimator yields point estimates that can be far from the parameters of the true data-generating process.

2.3. Controlled environment

To analyze the behavior of the estimator for positive and finite weights on the moment conditions $F_j - \mu_j$, we specify a data-generating process and draw finite samples of factors and test asset returns. The factors $F_1, \dots, F_k$ are drawn independently from normal distributions with means and standard deviations that are all set to 1 percent per quarter. The data generating process for excess returns is given as

$$R_{i,t}^e = E[R_{i,t}^e] + \sum_{j=1}^{k} \beta_{i,j}(F_{j,t} - E[F_{j,t}]) + \sigma_i \varepsilon_{i,t}, \quad (9)$$

where the ε's are independent from one another and from the factors and are also i.i.d. normally distributed with means of 0 and standard deviations (σ_i) of 1 percent.

Unless stated otherwise, we simulate 25 return time series with 600 observations each. The sample size corresponds to a standard monthly post-war sample, but we also analyze the impact of the sample size in Section 2.5. Even though both the factor time series and the ε's are sampled from independent distributions, we subsequently orthogonalize them to ensure that they are perfectly orthogonal even in our small sample. Since we are not interested in standard error estimates or small sample properties, we draw only one single sample and then rescale this one sample to make sure that the sampled data have the same in-sample moments as the data in population.[4]

The cross-sectional variation in expected returns $E[R_{i,t}^e]$ is modeled as follows. We first draw a true factor exposure $b_{i,j}$ of the return of asset i to factor j from a normal distribution with a mean and standard deviation of 1 percent. Importantly, the true exposures are assumed to be constant over time, i.e., they are only drawn once before we simulate the factor and return time series. The vectors of true factor exposures $(b_{i,j})_{i=1,\dots,n}$ for the different factors j are supposed to be orthogonal. After drawing the vectors of factor exposures and orthogonalizing them, we again rescale them, such that they all have a mean and standard deviation of 1 percent in sample. Finally, to allow for model misspecification and varying degrees of explanatory power of the factor exposures in our setup, we set

$$E[R_{i,t}^e] = \sum_{j=1}^{k} (r_j b_{i,j} + \sqrt{1 - r_j^2}\, e_{i,j}), \quad (10)$$

where the parameters r_j are chosen between 0 and 1 and the $e_{i,j}$ have exactly the same properties as the $b_{i,j}$, but are orthogonal to them and have a cross-sectional mean of zero. Appendix B.1 shows that the true cross-sectional R^2 is then given by $\frac{1}{k}\sum_{j=1}^{k} r_j^2$. Our design allows us to analyze perfectly priced factors (by setting $r = 1$), perfectly unpriced factors (by setting $r = 0$), and everything in between.

We also want to distinguish between weak and strong factors. To this end, we introduce parameters $s_j \in [0, 1]$ that control the time series correlation between factors and returns. Appendix B.1 provides details on

[3] This choice of μ is equivalent to using the moment condition $0 = E\left[R_i^e - R_i^e(F - \bar{F})\lambda\right]$, where $\bar{F}$ denotes the time series average of the factor. This alternative GMM estimator has also been suggested in the literature; see, e.g., the detailed discussion in Ferson (2019), p. 180, p. 220, or pp. 224ff.

[4] We refrain from labeling this as a Monte Carlo simulation because, in such an experiment, one would draw a large number of small samples with sampling errors to study the small sample properties of estimators.

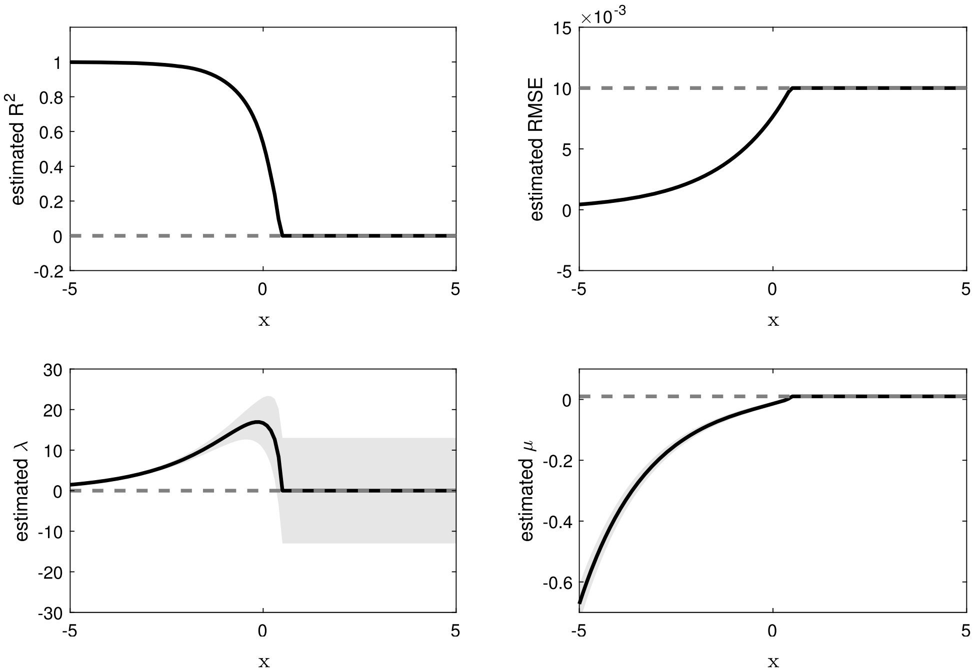

Fig. 1. A single strong and perfectly unpriced factor. We apply a GMM estimation with the moment conditions in Equation (2) and the weighting matrix in Equation (5). The figure shows estimated R^2 and RMSE and the point estimates of λ and μ, together with 95% confidence bounds as functions of x, the log weight on the moment condition that identifies the factor mean.

how we set these parameters and how the true MPRs can be calculated as functions of s_j, r_j, and $b_{i,j}$. Our parameterization guarantees that the volatility of excess returns is equal to 0.06 for all assets. This implies an annual return volatility of 20.78% for all test assets.

In the following, we analyze the behavior of the GMM estimator using data generated by the time series and cross-sectional model introduced above. We keep all parameters fixed, with the exception of s_j, r_j and the GMM weighting matrix W. The resulting framework is very flexible and allows us to vary the strength of the factor (strong vs. weak), the true cross-sectional R^2 (priced vs. unpriced factors), and the sample size along all three dimensions (length of time series, number of test assets, and number of factors).

2.4. Prespecified diagonal weighting matrices

Single factor models: We start our analysis with weighting matrices of the form $W_x = \text{diag}(1, \ldots, 1, 10^x, \ldots, 10^x)$, as introduced in Equation (5), and assuming that there is a single factor that is strong and perfectly unpriced. The latter means that expected test asset returns are orthogonal to the factor exposures, with the result that the true cross-sectional R^2 is equal to zero. The true MPR λ is thus equal to zero as well.

We vary the log weight x on the moment condition identifying the factor mean, and we study the effects of this variation on the point estimates (i.e., the solutions to Equations (7) and (8)), the estimated R^2, RMSE, and 95% confidence bands. Fig. 1 shows these quantities for values of x between -5 and 5 for the case of a single strong and unpriced factor.

The figures confirm what was surmised in Section 2.1. We find that, once x exceeds a critical value (in this case around 0.45), the estimates are in line with the true values, i.e., an R^2 of almost zero, an RMSE that is equal to the cross-sectional standard deviation of expected returns, and point estimates $\hat{\lambda} = 0$ and $\hat{\mu} = 0.01$. This is in line with what we had labeled a desired solution in Section 2.1.

For $x \to -\infty$, the figure suggests that the estimated R^2 goes to 1, the RMSE goes to 0, and $\hat{\lambda}$ and $\hat{\mu}$ go to 0 and $-\infty$, respectively. This is in line with what we label as a trivial solution in Section 2.1.

For intermediate values of x below 0.45, we observe an estimated R^2 between 0 and 1 and large and significant λ estimates. For example, when using the identity matrix for weighting (i.e., $x = 0$), we estimate an R^2 of 0.53 and a λ of 16.67 with a t-statistic of 5.16.

To sum up, our analysis shows that low values of x lead to inflated R^2's and biased parameter estimates. Even seemingly innocuous choices of the weighting matrix, such as the identity matrix, can lead to such a result.

In the online appendix, we present and discuss results of further cases. Overall, the patterns are qualitatively very similar to those in Fig. 1. In Section B.2, we consider a factor that is strong and (imperfectly) priced (with a true cross-sectional R^2 of 0.5). In Section B.3, we then turn to the analysis of a weak factor.

Multifactor models: Our flexible setup allows us to analyze multifactor models with arbitrary combinations of weak, strong, priced and unpriced factors. Recall that there is always at least one factor j^* for which $\mu_{j^*} = E[F_{j^*}] - \lambda_{j^*}^{-1}$ and $\lambda_{j^*} \to 0$, i.e., the MPR is biased downwards when the estimator runs into a trivial solution (see Section 2.1). It is interesting to analyze *which* factor plays that role in situations with several factors. We exemplify the intuition in a case with one strong and priced factor (setting $r_1 = 1$ and $s_1 = \sqrt{0.9}$) and one rather weak and perfectly unpriced factor (setting $r_2 = 0$ and $s_2 = \sqrt{0.01}$). The true cross-sectional R^2 is equal to $0.5 = \frac{1}{2}(r_1^2 + r_2^2)$. Fig. 2 shows the usual statistics for this scenario.

The figure documents that, as before, the estimated R^2 and RMSE move to the true values as x increases. For log weights x above a certain threshold (in this case $x = 2.6$), deviations of $\hat{\mu}$ from the sample averages of the factors are so costly that the minimum of the GMM objective function is attained at the true parameter values. For values of x below that threshold, it is again cheaper to reduce the pricing errors at the cost of not matching the moment condition $E_T[F_j - \mu_j]$.

Moreover, we observe that the relatively weak and unpriced factor (in this case F_2) is selected as j^*. The term $E_T[R_i^e(F_2 - E_T[F_2])]$ as part of the second term on the right-hand side of Equation (4) is closer to zero anyway, relative to F_1. Thus, the objective function can be minimized with λ_2 being far from zero and the moment condition $E_T[\mu_2 - F_2]$ being effectively satisfied. With these choices of λ_2 and μ_2, the strong and priced factor F_1 is not needed to reduce the pricing er-

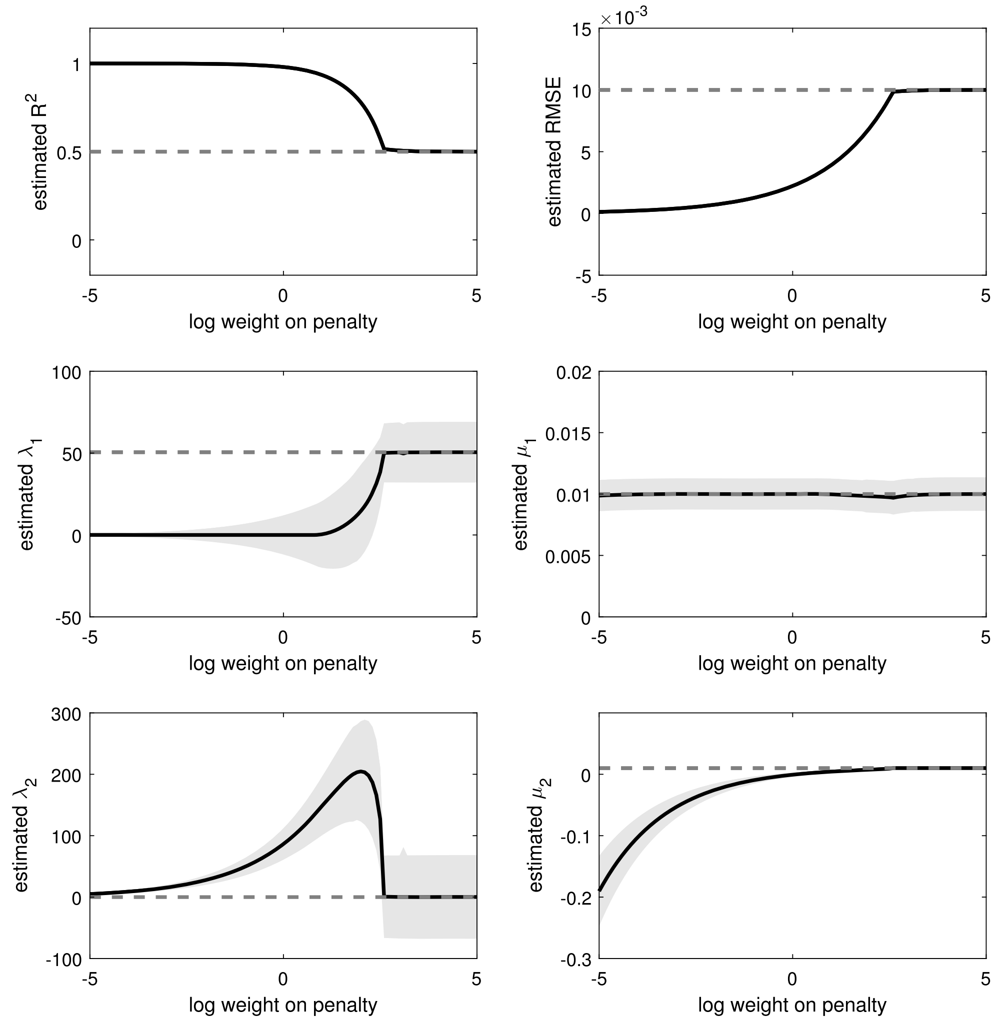

Fig. 2. One strong and priced and one weak and unpriced factor. We apply a GMM estimation with the moment conditions in Equation (2) and the weighting matrix in Equation (5). The figure shows estimated R^2 and RMSE and the point estimates of λ_1, λ_2, μ_1, and μ_2, together with 95% confidence bounds as functions of x, the log weight on the moment condition that identifies the factor mean. Returns are simulated according to Equations (9)-(10) with $r_1 = 1$, $r_2 = 0$, $s_1 = \sqrt{0.9}$, and $s_2 = \sqrt{0.01}$.

rors any further, so the impact of the covariances of returns with this factor is nullified by setting λ_1 to zero.

In numbers, when weighting the moment conditions with the identity matrix, the MPR λ_1 of the strong and priced factor F_1 is estimated at −2.03 (true value is 50) with a t-statistic of −0.32. The MPR λ_2 of the weaker and perfectly unpriced factor F_2 is estimated at 88.69 (true value is 0) with a t-statistic of 6.51. The estimated cross-sectional R^2 is equal to 97.96% (true value is 50%) and the estimated root mean squared pricing error is 0.22% (true value is 1%).

Taken together, we see that relatively weaker unpriced factors can drive out relatively stronger priced factors. More precisely, the MPR estimates of relatively weaker factors can be large and significant, while those of stronger priced factors become small and insignificant. As in the single factor case, the estimated cross-sectional R^2 can be inflated heavily, and we again observe this pattern even for innocuous choices of weights like the identity matrix.

Apart from this basic insight concerning the different roles of weak and strong factors, many other patterns are qualitatively similar to those in Fig. 1. A detailed discussion of the various cases is provided in Appendix B.5. There we also discuss the case of two equally strong factors. In contrast to the analysis presented in Fig. 2 showing that weaker factors drive out stronger factors, we cannot conclude that unpriced factors drive out priced factors if they are comparable in strength. Still, parameter estimates are biased and R^2's are inflated when the weight on the moment conditions for the factor means is low.

2.5. *The impact of sample size*

The results presented above refer to 25 test assets simulated for 600 months, corresponding to the post-war samples that have long been widely used in asset pricing. Some papers warn that using too few test assets can lead to problems with cross-sectional tests. Lewellen et al. (2010) argue that the strong factor structure in size and book-to-market sorted portfolios makes spurious factors appear to be priced if they happen to be correlated with one of the Fama and French (1993) factors. They recommend using a broader set of equity portfolios. Ang et al. (2020) argue that portfolio construction destroys information by reducing the cross-sectional variation in betas and recommend the use of individual stocks instead. In contrast, Bekaert and De Santis (2021) warn that asset or portfolio returns can be highly correlated (especially in the case of corporate bond returns, which they consider), which can lead to extreme portfolio weights when testing mean-variance efficiency.

Interestingly, increasing the number of test assets makes the problem discussed in our paper even more severe. When we add more test assets but keep the parameter x in the weighting matrix (5) fixed, the

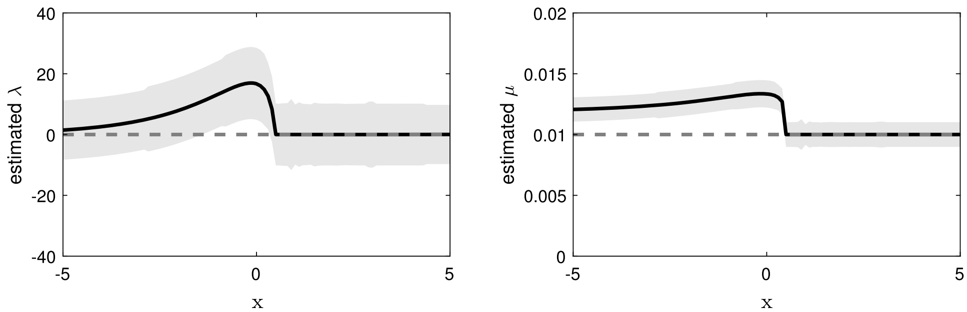

Fig. 3. Using the inverse of the estimated covariance matrix for weighting. We apply an iterated GMM estimation with the moment conditions in Equation (2) and use the inverse of the covariance matrix of the moment conditions, estimated using the point estimates from the previous stage, as the weighting matrix. In the first stage GMM estimation, we use a diagonal matrix of the form $W = \text{diag}(1, \ldots, 1, 10^x)$ for weighting. The figure shows the point estimates of λ and μ, together with 95% confidence bounds as functions of x, the log weight on the moment condition that identifies the factor mean in the first stage.

overall weight on the Euler equation moment conditions in the GMM objective function increases relative to the moment conditions identifying the factor means (whose number does not change). Appendix B.4 provides simulation results for different sample sizes. With the exact setting from Section 2.3 and $W = I_{n+k}$, we show that our estimates equal the true data-generating parameters if the number of test assets is equal to only 5. In contrast, when the number of test assets exceeds 200, the estimator produces biased point estimates and highly inflated model performance statistics, with a cross-sectional R^2 close to 1. To circumvent this issue, one would have to rescale the weight on the moment conditions $E_T[F - \mu]$ in such a way that the relative importance of these moment conditions is kept constant across sample sizes.

For completeness, we also consider the impact of the length of the time series T. Not surprisingly, the length of the time series does not matter at all, since the bias we document is not a small sample property.

2.6. Endogenous weighting matrices

In applications, weighting matrices are often chosen endogenously, for instance by inverting a prior estimate of the covariance matrix of error terms, to account for the fact that less volatile moments provide more information on the parameters to be estimated. With that in mind, we now analyze the behavior of the GMM estimator for alternative endogenous specifications of W.

As a starting point, we stick to the diagonal structure of W but use the inverses of the variances of the moment conditions as weights, as is done, e.g., by Yogo (2006). Following Equation (2), these variances are simply given by the variances of the test asset returns and of the factor. When the true λ is equal to zero, the corresponding diagonal weighting matrix boils down to $\text{diag}(Var(R_1^e)^{-1}, \ldots, Var(R_n^e)^{-1}, Var(F)^{-1})$. When the true λ is different from zero, the variances of the pricing errors increase only slightly.

The extent of the problem discussed here thus essentially depends on the volatility differential between factors and test asset returns. If the pricing factor F is about as volatile as the test asset returns (e.g., if the factor is traded or a factor-mimicking portfolio), we basically once again find ourselves with the case of an identity matrix as weighting matrix with all the consequences described in Section 2.4. To assess the severity when the pricing factor and the test asset returns have very different volatilities (e.g., when analyzing low-volatility macro factors), we can simply scale the weight on the pricing factor up or down. The true pricing performance of linear factor models is invariant to affine linear transformations of the factors. But for the GMM estimation, multiplying a factor F_j by a scalar m is tantamount to multiplying the weight on the moment condition $\mu_j - F_j$ by $1/m^2$. The quantitative analysis presented above gives an idea of how strongly such a scaling can affect the estimation results. Specifically, when the weighting matrix is based on inverses of variances of moment conditions, estimates for low-volatility macro factors (like consumption growth) should be affected less than those for traded factors.

Instead of imposing a prespecified (diagonal) structure on the weighting matrix, researchers often try to estimate the covariance matrix of the moment conditions using an iterated approach. Parameter estimates in the current stage allow us to estimate the covariance matrix whose inverse then serves as the weighting matrix for the next stage. The algorithm either runs through a prespecified number of iterations or until the parameter estimates show signs of "convergence".

We use the same simulated strong and perfectly unpriced factor as in Section 2.4 and run an iterated GMM estimation. In the first step of the iteration, we use the weighting matrix $W = \text{diag}(1, \ldots, 1, 10^x)$. Using the point estimates from this first stage, we estimate the covariance matrix of the moment conditions. We calculate the inverse of this estimated covariance matrix and use it as the weighting matrix in the next stage. We iterate the procedure until convergence, i.e., until the point estimates from the current iteration stage are very close to those of the previous stage.[5]

Fig. 3 shows the point estimates of λ and μ, together with 95% confidence intervals, as functions of the log weight x on the moment condition identifying the factor means in the first stage GMM estimation. If the iteration always converged to the true parameters, the point estimates would be independent of the weight x in the first stage.

The figure shows that this is not the case. In fact, the pattern in point estimates is remarkably similar to the one from Fig. 1. Note that the results in Fig. 1 are, by construction, equal to the results in the first stage of the estimation analyzed here. We again find that there is a critical value for x, which is slightly greater than zero. For log weights above that critical value, the point estimates of λ are equal to zero, the true value. Starting with a log weight below the critical value leads to biased parameter estimates. For these values, the μ estimates are biased as well.

Fig. 3 suggests that biased first stage estimates carry over to later stages in the iteration. This is not surprising. Suppose the log weight x in the first stage is very low. As argued in Section 2.1 above, we then end up with parameter estimates in this case that are close to a trivial

[5] As an alternative to iterated GMM, we also investigate the variant known as "continuously updating GMM" (CUGMM) proposed by Hansen et al. (1996). Instead of estimating the covariance matrix in an iterated fashion and hoping that it converges to the true covariance matrix, this procedure considers the estimated covariance matrix as a function of the true parameters and directly minimizes the GMM objective function $g_T(\lambda,\mu)' W(\lambda,\mu) g_T(\lambda,\mu)$, where $W(\lambda,\mu)$ is the inverse of $\widehat{Cov(g(\lambda,\mu))}$. We run CUGMM estimations on our simulated data. We find that the algorithm runs into areas of the parameter space where the estimated covariance matrix of the moment conditions cannot be inverted and the algorithm does not converge to a solution. The global infimum of the modified GMM objective function is at the trivial solution, where the estimated covariance matrix of the moment conditions is singular.

solution in the first stage estimation, meaning that $\hat{\mu} \approx E_T[F] - \frac{1}{\hat{\lambda}}$ and $\hat{\lambda} \approx 0$. With these parameters, the first n moment conditions of the form $R_i^e - R_i^e(F-\mu)\lambda$ approximately simplify to $R_i^e - R_i^e$. So not only is the sample average of these moment conditions virtually equal to zero, so too is each single observation. Thus, the variances of the first n moment conditions and their covariances with the factors are close to zero. The covariance matrix of the moment conditions after the first stage can be approximately written as

$$\widehat{Cov(g)} \approx \begin{pmatrix} 0_{n\times n} & 0_{n\times 1} \\ 0_{1\times n} & \widehat{Var(F)} \end{pmatrix},$$

i.e., $\widehat{Cov(g)}$ is (close to) singular. The estimated variance of the pricing errors is close to zero, and this makes these moments appear much more informative about the parameters to be estimated than the moment conditions identifying the factor means. This leads to biased parameter estimates, and, consequently, biased estimates of the covariance matrix, in the subsequent stages.

3. Examples

We now turn to a set of popular empirical examples that highlight the practical relevance of the bias outlined in our paper. Section 3.1 considers the CAPM, the Fama and French (1993) three-factor model, and the number of Asian elephants kept in zoos worldwide (as an example of a weak factor). In Section 3.2, we apply the estimation design to a set of macroeconomic factors: durable consumption growth (Yogo, 2006), long-run consumption growth (Parker and Julliard, 2005), unfiltered consumption growth (Kroencke, 2017), intermediary leverage (He et al., 2017), and term and default spread (Maio and Santa-Clara, 2012). In these two subsections, we employ a one-stage GMM estimation with a fixed diagonal weighting matrix for the purpose of exposition.

Importantly, although we choose the factor models of Parker and Julliard (2005), Kroencke (2017), and He et al. (2017) for illustrative purposes here, we wish to emphasize that we are not leveling any criticism at the original papers, since these authors do not use the estimation design that we discuss here. This is not the case, however, for the model of Yogo (2006). In Section 3.3, we reconsider the two-stage GMM approach of Yogo (2006) in great detail. We start from the weighting matrix used in the original paper and then study several variations of it.

3.1. The Fama-French factors and Asian elephants

To exemplify the bias described in the previous sections, we use a quarterly post war sample comprising the standard 25 size and book-to-market sorted portfolios as test assets and the following set of factors: (i) the market factor, a strong but basically unpriced factor for this cross-section (see Fama and French, 1992), (ii) SMB and HML, which are strong and priced,[6] and (iii) an obviously economically meaningless factor, namely the log growth rate of the number of captive Asian elephants living in zoos around the world.[7] The elephant factor is a weak factor, as its time series correlation with returns is close to zero for all test assets.

Table 1 shows MPR estimates and model performance statistics for three factor models obtained with three different GMM weighting matrices. For the first column of each panel, we assign a high weight to the moment conditions identifying the factor means to make sure they are correctly estimated. The second column shows the analogous quantities, estimated using an identity matrix for weighting the moment conditions, and the respective third column shows them assigning a low weight to the moment conditions identifying the factor means.[8]

For the one-factor model (Panel A), a high weight on μ leads to a negative R^2 (the GMM estimation corresponds to a cross-sectional regression without intercept). The market factor has a significant MPR, but does not explain cross-sectional return variation. With a low weight on the condition for μ, however, the one-factor model exhibits the bias described in Section 2.4, namely an estimated cross-sectional R^2 close to one, pricing errors close to zero, and an MPR estimate statistically significantly different from zero with a t-statistic of 5.85.

The numbers in square brackets show test statistics for t-tests of the hypotheses $\lambda = 0$ and $\mu = \bar{F}$. As expected, μ_{MKT} is significantly different from the sample average of the market factor for low values of x. Note that this t-test is equivalent to testing if the moment condition identifying the factor mean is equal to zero. In terms of Fig. 1, it is again equivalent to checking whether the confidence interval around the μ-estimate contains the sample average of the factor. This can serve as an important plausibility check.

The results for the Fama-French three-factor model are similar, except for overall pricing performance being high already with a high weight on the factor mean moments. Going from a high to a unit weight improves pricing performance only slightly. Further lowering the weight again increases the R^2 to almost 1 and decreases the pricing errors. At the same time, the coefficient estimates shrink towards zero.

For the multi-factor model featuring the elephant factor, we see a similar pattern for the cross-sectional R^2, but an additional effect for the MPR estimates. Not only does the elephant factor become significant, it also drives out all three Fama-French factors, suggesting that just the elephant factor is enough to explain the cross-section of expected returns when the weight on μ is set low enough. At the same time, the estimated μ significantly differs from the sample average of the elephant factor, while the other factor mean estimates are still close to their respective sample averages. This pattern is perfectly in line with the findings presented in Section 2.4 above. Weak and unpriced factors can drive out strong and priced factors when the weight on the moment conditions identifying the factor means is (too) low.

3.2. Macroeconomic factors

We now look at five prominent macroeconomic factor models from the more recent literature. In a departure from the approach taken in the original papers, we standardize the representations and the estimation technique for the sake of comparability. More precisely, we consider linear factor models (i.e., we linearize the stochastic discount factors from the original papers), do not impose any constraints on the parameters, and, of course, always use the estimation technique discussed here. Consequently, the findings in this subsection do not overturn any of the results from the original papers, and the numbers reported below should not be compared to them. In this subsection, we always impose a diagonal weighting matrix with varying weights on the moments identifying the factor means.

Table 2 reports point estimates, standard errors, and model performance statistics (mean absolute errors and cross-sectional R^2). Panel A shows results for the three-factor model suggested by Yogo (2006), featuring the log growth rate of consumption of nondurable goods and services (hereinafter denoted by F_1), the log growth rate of consumption of durable goods (F_2), and the log return on the aggregate stock market (F_3) as factors. For Panel B, we linearize the stochastic discount factor representation introduced by Parker and Julliard (2005). The factors are the log growth rate of consumption of nondurables and services

[6] Both these factors and the test asset returns were downloaded from Kenneth French's webpage; see http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[7] The data are available at http://www.asianelephant.net/database.htm. We thank the creators of this website, Jonas Livet and Torsten Jahn, for making these data publicly available.

[8] Here and in the following subsection, "high", "unit" and "low" mean $x = 5$, $x = 0$, and $x = -5$ in the notation of Equation (5) in Section 2.

Table 1
Market prices of risks for strong and weak factors.

	A: CAPM			B: Fama French 3			C: FF3 + elephants		
	Weight on μ			Weight on μ			Weight on μ		
	High	Unit	Low	High	Unit	Low	High	Unit	Low
λ_{MKT}	3.10	3.07	0.50	3.43	3.37	0.48	3.43	0.38	−0.00
	(0.94)	(0.92)	(0.09)	(1.18)	(1.15)	(0.13)	(1.19)	(0.82)	(0.05)
	[3.29]	[3.34]	[5.85]	[2.91]	[2.94]	[3.52]	[2.88]	[0.46]	[−0.09]
λ_{SMB}				0.32	0.31	−0.01	0.32	−1.72	−0.10
				(1.62)	(1.60)	(0.22)	(1.74)	(1.44)	(0.08)
				[0.20]	[0.19]	[−0.03]	[0.18]	[−1.19]	[−1.25]
λ_{HML}				5.76	5.69	0.93	5.76	−0.47	−0.07
				(1.68)	(1.63)	(0.15)	(1.64)	(1.21)	(0.07)
				[3.42]	[3.49]	[6.34]	[3.52]	[−0.39]	[−1.05]
$\lambda_{Elephants}$							0.01	128.50	7.64
							(47.58)	(22.11)	(1.30)
							[0.00]	[5.81]	[5.87]
μ_{MKT}	1.47	1.04	−169.15	1.47	1.34	−36.45	1.47	1.47	1.48
	(0.55)	(0.56)	(27.98)	(0.55)	(0.55)	(5.75)	(0.54)	(0.54)	(0.54)
	[−0.00]	[−0.77]	[−6.10]	[−0.00]	[−0.24]	[−6.60]	[−0.00]	[−0.00]	[0.01]
μ_{SMB}				0.41	0.40	0.96	0.41	0.42	0.59
				(0.34)	(0.34)	(0.37)	(0.34)	(0.36)	(0.37)
				[−0.00]	[−0.04]	[1.46]	[−0.00]	[0.03]	[0.49]
μ_{HML}				0.98	0.75	−73.21	0.98	0.98	1.10
				(0.36)	(0.37)	(11.08)	(0.36)	(0.36)	(0.36)
				[−0.00]	[−0.61]	[−6.70]	[−0.00]	[0.01]	[0.34]
$\mu_{Elephants}$							0.48	−0.25	−12.65
							(0.06)	(0.13)	(2.24)
							[−0.00]	[−5.41]	[−5.86]
RMSE (%)	3.81	3.78	1.71	2.02	2.00	0.83	2.02	1.04	0.13
R^2 (%)	−52.96	−49.04	96.46	56.64	58.12	99.15	56.64	94.00	99.98

The table reports estimates of λ and μ from a GMM estimation using the moment conditions $E[R_i^e - R_i^e(F-\mu)\lambda] = 0$ and $E[F-\mu] = 0$ along with three different weighting matrices. Heteroskedasticity and autocorrelation-consistent (HAC) standard errors (in parentheses) are calculated using a Bartlett kernel. Test statistics of t-tests of the hypotheses $\lambda = 0$ and $\mu = \bar{F}$ are shown in brackets.

over the following 12 quarters (F_1), downloaded from the NIPA tables, and the log risk-free rate between time $t+1$ and $t+12$ (F_2). The model of Kroencke (2017) is presented in Panel C. Factor F_1 in the linearized model we consider here is the log growth rate of unfiltered consumption, using the definition provided in Kroencke (2017).[9] For the results in Panels A-C, we use quarterly excess returns of the standard 25 size and book-to-market-sorted portfolios from 1951:Q1 to 2001:Q4 as test assets.[10] Panel D refers to the intermediary asset pricing model proposed by He et al. (2017). Factor F_1 is the intermediary capital risk factor, and factor F_2 is the market factor.[11] Panel E refers to a model suggested by Hahn and Lee (2006) and studied by Maio and Santa-Clara (2012). Factor F_1 is the default spread, F_2 is the term spread, and F_3 is again the market factor.[12] The data for Panels D and E are monthly and range from 1970 to November 2018.

The results are very similar across all panels. With a high weight on the moment conditions identifying the factor means, we find that the pricing performances are moderate (the cross-sectional R^2 is sometimes even negative), while the factor means are matched very well. Only very few MPR estimates are significantly different from zero in this situation, e.g., the one for the consumption factor in Kroencke (2017) and the intermediary factor proposed by He et al. (2017). With decreasing weights on the factor means, pricing performances increase at the cost of not matching the moments identifying the factor means. For instance, the results in Panel A suggest that F_2, the growth in durable consumption, explains the cross-section of expected returns almost entirely. At the same time, we find that the factor means of the two consumption-based factors are significantly different from their sample averages. This pattern is evident for the unit weighting matrix (with an R^2 of 95.26%, close to the value reported in the original paper of Yogo (2006)), and even more so when the weight on the factor means is low.

3.3. The optimal weighting matrix

Finally, we explore the model of Yogo (2006) in more detail by following the approach from the original paper regarding the estimation of the weighting matrix in a two-stage GMM approach. The sample, the factors, and the test assets are exactly as described above in Section 3.2.

Yogo performs a two-stage GMM estimation, where the weighting matrix in the second stage is the inverse of the covariance matrix estimated in the first stage (Section D of the online appendix provides details). The weighting matrix $W^{(1)}$ in the first stage is obtained using some initial values for the parameters λ and μ. More specifically, $W^{(1)}$ is chosen as

[9] The unfiltered consumption factor is downloaded from Tim Kroencke's website at https://sites.google.com/site/kroencketim/.

[10] Both the test asset returns and the consumption data are taken from the supplementary material to Yogo (2006), kindly provided on Motohiro Yogo's website at https://sites.google.com/site/motohiroyogo/.

[11] The intermediary factor is downloaded from Asaf Manela's website at https://apps.olin.wustl.edu/faculty/manela/data.html.

[12] The default spread is defined as the spread between Moody's Seasoned Baa Corporate Bond Yield and the Federal Funds Rate. The term spread is defined as the spread between the 10-year US Treasury yield and the Federal Funds Rate. Both time series are downloaded from the Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org.

Table 2
GMM — various models.

	A: Yogo			B: Parker/Julliard			C: Kroencke			D: He/Kelly/Manela			E: Maio/Santa-Clara		
	Weight on μ			Weight on μ			Weight on μ			Weight on μ			Weight on μ		
	High	Unit	Low	High	Unit	Low	High	Unit	Low	High	Unit	Low	High	Unit	Low
λ_1	335.15	−49.08	−16.50	75.43	45.03	3.13	106.79	45.64	2.90	8.39	8.42	1.71	−116.80	−97.64	−6.84
	(212.11)	(69.09)	(21.05)	(47.85)	(13.78)	(0.51)	(63.18)	(14.32)	(0.72)	(3.58)	(3.51)	(0.23)	(97.77)	(41.98)	(1.90)
	[1.58]	[−0.71]	[−0.78]	[1.58]	[3.27]	[6.15]	[1.69]	[3.19]	[4.05]	[2.34]	[2.51]	[7.61]	[−1.19]	[−2.33]	[−3.60]
λ_2	−177.56	154.18	44.02	−9.50	−4.99	−0.29				−6.08	−6.14	−1.60	191.68	139.35	9.17
	(165.86)	(26.27)	(15.69)	(10.30)	(6.84)	(0.51)				(4.29)	(4.23)	(0.32)	(126.50)	(46.36)	(2.19)
	[−1.07]	[5.87]	[2.81]	[−0.92]	[−0.73]	[−0.56]				[−1.42]	[−1.51]	[−5.03]	[1.52]	[3.01]	[4.18]
λ_3	−4.96	2.10	0.61										−1.38	−1.28	−0.09
	(5.17)	(1.24)	(0.45)										(2.86)	(1.56)	(0.09)
	[−0.96]	[1.69]	[1.37]										[−0.48]	[−0.82]	[−1.00]
μ_1	0.51	0.71	0.51	6.20	5.27	−24.23	0.54	−0.79	−33.07	0.09	0.42	−28.22	3.42	3.61	8.54
	(0.05)	(0.06)	(0.29)	(0.32)	(0.40)	(4.88)	(0.07)	(0.30)	(8.15)	(0.29)	(0.51)	(4.25)	(0.14)	(0.16)	(1.25)
	[−0.01]	[2.99]	[−0.02]	[−0.00]	[−2.33]	[−6.23]	[−0.00]	[−4.36]	[−4.12]	[−0.00]	[−0.51]	[−6.66]	[0.00]	[1.24]	[4.11]
μ_2	0.92	0.31	−1.38	15.81	15.92	18.61				0.55	0.14	26.92	1.20	0.92	−5.68
	(0.06)	(0.16)	(0.76)	(0.93)	(0.93)	(1.02)				(0.19)	(0.23)	(4.01)	(0.12)	(0.14)	(1.64)
	[0.00]	[−3.87]	[−3.00]	[0.00]	[0.11]	[2.75]				[0.00]	[0.51]	[6.58]	[−0.00]	[−2.00]	[−4.20]
μ_3	1.88	1.87	0.56										0.55	0.055	0.62
	(0.56)	(0.56)	(0.56)										0.19)	(0.19)	(0.19)
	[0.00]	[−0.01]	[−2.35]										0.00]	[0.01]	[0.37
MAE (%)	0.41	0.10	0.03	0.39	0.23	0.02	0.68	0.26	0.02	0.15	0.14	0.02	0.13	0.07	0.00
R^2 (%)	1.17	95.26	99.62	16.47	71.77	99.87	−110.27	65.68	99.86	10.16	11.80	98.29	17.08	79.65	99.92

The table reports estimates of λ from a GMM estimation using the moment conditions $E\left[R_i^e - R_i^e(F-\mu)\lambda\right] = 0$ and $E[F-\mu] = 0$ along with three different weighting matrices. Heteroskedasticity and autocorrelation-consistent (HAC) standard errors (in parentheses) are calculated using a Bartlett kernel. Test statistics of t-tests of the hypotheses $\lambda = 0$ and $\mu = \bar{F}$ are shown in brackets.

Table 3
Parameter estimates for different weighting matrices.

	First stage results					Second stage results		
	$x=-4$	$x=0$	$x=4$	FMB		$x=-4$	$x=0$	$x=4$
λ_1	−5.45	15.54	335.58	278.74	λ_1	−5.29	17.90	140.03
	(26.05)	(113.72)	(234.99)	(76.00)		(3.27)	(31.28)	(43.22)
	[−0.21]	[0.14]	[1.43]	[3.67]		[−1.62]	[0.57]	[3.24]
λ_2	17.79	152.41	−181.43	−103.76	λ_2	19.20	170.57	−137.94
	(43.87)	(39.41)	(272.88)	(64.47)		(1.79)	(15.56)	(53.54)
	[0.41]	[3.87]	[−0.66]	[−1.61]		[10.74]	[10.96]	[−2.58]
λ_3	0.23	1.00	−5.00	3.13	λ_3	0.20	0.66	−2.08
	(1.35)	(2.70)	(6.63)	(0.96)		(0.08)	(0.85)	(1.31)
	[0.17]	[0.37]	[−0.75]	[3.26]		[2.54]	[0.78]	[−1.58]
μ_1	0.25	0.48	0.51	0.51	μ_1	0.40	0.53	0.56
	(0.36)	(0.05)	(0.05)	(0.05)		(0.03)	(0.04)	(0.04)
	[−0.71]	[−0.65]	[0.00]			[−3.40]	[0.50]	[1.00]
μ_2	−4.78	0.30	0.92	0.92	μ_2	−4.29	0.28	0.94
	(15.71)	(0.25)	(0.12)	(0.07)		(0.49)	(0.10)	(0.08)
	[−0.36]	[−2.42]	[0.01]			[−10.56]	[−6.12]	[0.27]
μ_3	−0.74	1.60	1.88	1.88	μ_3	0.86	1.64	1.59
	(3.79)	(0.56)	(0.56)	(0.55)		(0.44)	(0.47)	(0.45)
	[−0.69]	[−0.51]	[0.00]			[−2.30]	[−0.51]	[−0.64]
MAE (%)	0.01	0.12	0.42	0.42	J	7.83	23.17	29.11
R^2 (%)	99.94	93.46	1.25	0.95	p-val	1.00	0.39	0.14

The table presents the results for the durable consumption model of Yogo (2006) for different choices of the pre-estimation weighting matrix. The sub-block related to the three factor means is multiplied by w. "FMB" denotes a standard Fama-MacBeth two-stage regressions. The μ_i in this last column are the sample averages of the three factors and not part of the regressions. Heteroskedasticity and autocorrelation-consistent (HAC) standard errors (in parentheses) are calculated using a Bartlett kernel. Test statistics of t-tests of the hypotheses $\lambda = 0$ and $\mu = \bar{F}$ are shown in brackets. Details regarding the estimation of the model are presented in Section D of the online appendix.

$$W^{(1)} = \begin{pmatrix} \det(\widehat{\Omega}^{(1)}_{1,\dots,25})^{-\frac{1}{25}} I_{25} & 0 \\ 0 & (\widehat{\Omega}^{(1)}_{26,\dots,28})^{-1} \end{pmatrix},$$

where $\widehat{\Omega}^{(1)}$ is the estimate of the 28×28-covariance matrix of the moment conditions, given these initial values. $\widehat{\Omega}^{(1)}_{i,\dots,j}$ denotes block sub-matrices of $\widehat{\Omega}^{(1)}$ and I_{25} is the 25-dimensional identity matrix. To study the impact of the weights for the factor mean moment conditions on the estimation results, we multiply the lower block $(\widehat{\Omega}^{(1)}_{26,\dots,28})^{-1}$ in $W^{(1)}$ by a factor 10^x, varying x between −4 and 4.

The results are shown in Table 3. Just as in the cases discussed above, the pricing performance of the model varies dramatically with x. With $x = -4$, the pricing error is only 0.012% and the R^2 is 0.999. Moreover, the first stage estimates of the factor means differ dramatically from the sample averages $\bar{F}_1 = 0.513$, $\bar{F}_2 = 0.915$, and $\bar{F}_3 = 1.880$, because the weights on these moment conditions are very low. Increasing x leads to an increase in mean absolute pricing errors and a decrease in R^2. The second column ($x = 0$) is identical to the results reported in Yogo (2006). It is also identical to the first column in Table D.1 in the appendix, which shows the results for several further specifications. For $x = 4$, the R^2 is down to 0.013. At the same time, the factor mean estimates are basically equal to the sample averages, at least in the first stage regression. Overall, the results are qualitatively remarkably similar to the ones reported in Panel A of Table 2, although the construction of the weighting matrix is much more convoluted.

We also compare the point estimates and the cross-sectional R^2 to the results of a standard Fama-MacBeth two-pass regression in the column labeled "FMB" in Table 3. The R^2 there is very close to the one for $x = 4$. There are no degrees of freedom for the factor means in a Fama-MacBeth regression (the values for μ_1, μ_2, μ_3 reported in this column are just the sample averages of the factors). It is thus similar to the case $x = 4$ in which the additional parameters μ_i are also basically fixed to the sample averages of the factors.

Finally, it is also interesting to look at the point estimates from the second stage of the GMM estimation in order to assess the difference between a prespecified and an "efficient" weighting matrix. The estimation using the latter can be interpreted as a GLS approach, in contrast to using a prespecified and diagonal weighting matrix, which would correspond to OLS. Comparing the point estimates to those from the first stage, we see that the imprecision in the factor means estimation carries over from the first to the second stage. The estimated mean growth rate of durable consumption in our replication of Yogo (2006) is only 0.28, which is very low compared to the sample average of 0.92, and even further away from this value than the estimate of 0.30 in the first stage. However, as pricing performance diminishes with our adjustments to the procedure, the estimate of the mean growth rate of durable consumption comes closer to its sample average in the second stage as well. As already pointed out, this suggests that the weight on the moment condition that pins down the factor mean of durable consumption growth is too small in the original implementation by Yogo (2006).

The wide range of estimates in Table 3, particularly when compared to the Fama-MacBeth results, suggests that the estimates with $x = 4$ may be regarded as close to the true parameters, while those obtained with $x = 0$ should be regarded as suspicious. However, in contrast to the controlled simulation exercise in Section 2, we do not know the true factor means. The R^2 of the durable goods model may in fact be close to 1. However, this would require the true mean growth rate of durable consumption to be 0.3 percentage points, while the sample average equals 0.915 percentage points. Such a discrepancy would in turn imply that the true covariance of durable consumption growth with test asset returns is much higher than the sample covariance, i.e., that the factor is in fact much more important than it seems in sample. The appeal of GMM is that it trades off information from different moment conditions, so one is tempted to justify the high R^2 by arguing that this outcome is the most likely, given that the pricing error time series seem more informative (including about the factor means) than the factor time series.

On the other hand, our extensive analysis in Section 2 shows that the GMM estimator tends to favor low pricing errors and thus produces biased estimates. From this point of view, imagine that the true mean growth rate of durable consumption was exactly equal to the sample average. In this case, our analysis predicts exactly the pattern that we find in Table 3, and this pattern strongly suggests that the durable goods model does not do a very good job of explaining the cross-section of expected returns.

4. Discussion and conclusion

Cross-sectional asset pricing tests using GMM can generate spuriously high explanatory power for linear factor models as well as biased estimates for the market prices of factor risks. Tests based on simulated data show that any desired level of cross-sectional fit can be obtained by shifting the weights on the moment conditions. Our findings apply to all sample sizes, and we find that the larger the number of test assets, the more severe the problem. The only condition that needs to be met (and is always met in applications) is that the model is not perfectly specified, i.e., the true cross-sectional R^2 is less than one.

We revisit a number of macroeconomic factors suggested in the empirical asset pricing literature and find patterns that are strikingly similar to those from our simulation study. The prime example is the model of Yogo (2006), suggesting that growth in durable consumption is an important asset pricing factor. However, Yogo estimates a factor mean of around 0.3 percent per quarter, while the sample average factor mean is greater than 0.9 percent per quarter. This allows the estimator to spuriously match the Euler equations and heavily inflate model performance statistics. The same is likely true for other papers in the asset pricing literature, which we do not replicate, but which use the same estimation approach. A list of the papers we could identify is provided in Section C of the appendix.

How can we diagnose if the results in these and other papers are biased and pricing performance is spuriously high? A few straightforward suggestions can be made based on our analysis. Most easily, as mentioned by Parker and Julliard (2005), one can perform several GMM estimations with varying weights on the moment conditions that identify the factor means. The figures in our paper suggest that, if the point estimates are stable across weights in the neighborhood of a particular benchmark weight, we are either at the desired solution (which means that the parameter estimates are unbiased) or at a trivial solution. In the latter case, the estimated cross-sectional R^2 is close to one and there is one factor for which the estimated factor mean is close to its sample average minus the inverse of the estimated MPR.

Alternatively, we suggest straightforward hypothesis tests of the factor mean estimates like the ones discussed in Section 3.1. Typically, the estimates will be far from the sample averages of the factors if the pricing performance of the factors is inflated, and the hypothesis that the mean estimate is equal to the sample average will be rejected.

Estimating the factor means and the MPRs sequentially can serve as another plausibility check. One can demean the factors in a first step and then perform the GMM estimation with pricing error conditions only. The point estimates resulting from this procedure would equal the ones from the joint GMM estimation with a high enough weight on the factor mean conditions. The standard errors, however, would obviously be different, since the standard errors for the MPRs would disregard any estimation uncertainty about the factor means. Still, such a sequential procedure may serve as a benchmark for gaining a better understanding of the point estimates, irrespective of their significance.

Another way of checking the plausibility of the model performance statistics is to compare the results of the GMM estimation to the cross-sectional R^2 and the pricing errors from a Fama and MacBeth (1973) regression. Moreover, the point estimates of the MPRs should be compared to Fama-MacBeth estimates as well. Fama-MacBeth regressions effectively pin down the factor mean estimates to the sample averages of the factors and do not allow the pricing errors to have an impact on the estimate of the factor means. Alternatively, a GMM estimation with the weighting scheme discussed by Burnside (2011) and in Section 2.2 above produces the same point estimates as two-stage Fama-MacBeth regressions and, additionally, gives standard errors that account for the usual errors-in-variables problem of two-stage regressions.

In the case of a traded factor, the estimate of λ should not only be tested for being significantly different from zero. Instead, researchers should also check whether it is close to the sample average of the factor (divided by the factor variance).

CRediT authorship contribution statement

Nora Laurinaityte: Data curation, Investigation. **Christoph Meinerding:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing. **Christian Schlag:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing. **Julian Thimme:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing.

Data availability

Data will be made available on request.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jbankfin.2024.107123.

References

Ang, A., Liu, J., Schwarz, K., 2020. Using stocks or portfolios in tests of factor models. J. Financ. Quant. Anal. 55, 709–750.
Bekaert, G., De Santis, R.A., 2021. Risk and return in international corporate bond markets. J. Int. Financ. Mark. Inst. Money 72, 101338.
Bekaert, G., Panayotov, G., 2020. Good carry, bad carry. J. Financ. Quant. Anal. 55, 1063–1094.
Borri, N., Ragusa, G., 2017. Sensitivity, moment conditions and the risk-free rate in Yogo (2006). Crit. Finance Rev. 6, 381–393.
Bryzgalova, S., 2016. Spurious Factors in Linear Asset Pricing Models. Working Paper.
Burnside, C., 2011. The cross section of foreign currency risk premia and consumption growth risk: comment. Am. Econ. Rev. 101, 3456–3476.
Burnside, C., 2016. Identification and inference in linear stochastic discount factor models with excess returns. J. Financ. Econom. 14, 295–330.
Chen, Z., Lu, A., 2017. Seeing the unobservable from the invisible: the role of CO2 in measuring consumption risk. Rev. Finance 22, 977–1009.
Cochrane, J., 2005. Asset Pricing. Princeton University Press, New Jersey.
Da, Z., Yang, W., Yun, H., 2016. Household production and asset prices. Manag. Sci. 62, 387–409.
Darrat, A., Li, B., Park, J., 2011. Consumption-based CAPM models: international evidence. J. Bank. Finance 35, 2148–2157.
Delikouras, S., Kostakis, A., 2019. A single-factor consumption-based asset pricing model. J. Financ. Quant. Anal. 54, 789–827.
Della Corte, P., Ramadorai, T., Sarno, L., 2016a. Volatility risk premia and exchange rate predictability. J. Financ. Econ. 120, 21–40.
Della Corte, P., Riddiough, S.J., Sarno, L., 2016b. Currency premia and global imbalances. Rev. Financ. Stud. 29, 2161–2193.
Della Corte, P., Sarno, L., Schmeling, M., Wagner, C., 2022. Exchange rates and sovereign risk. Manag. Sci. 68, 5591–5617.
Fama, E., French, K., 1992. The cross-section of expected stock returns. J. Finance 47, 427–465.
Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 33, 3–56.
Fama, E., MacBeth, J., 1973. Risk, return, and equilibrium: empirical tests. J. Polit. Econ. 81, 607–636.
Ferson, W., 2019. Empirical Asset Pricing: Models and Methods. MIT Press.
Giglio, S., Xiu, D., Zhang, D., 2022. Test Assets and Weak Factors. Working Paper.
Gospodinov, N., Kan, R., Robotti, C., 2019. Too good to be true? Fallacies in evaluating risk factor models. J. Financ. Econ. 132, 451–471.
Gospodinov, P., Kan, R., Robotti, C., 2014. Misspecification-robust inference in linear asset pricing models with irrelevant risk factors. Rev. Financ. Stud. 27, 2139–2170.
Gospodinov, P., Kan, R., Robotti, C., 2017. Spurious inference in reduced-rank asset pricing models. Econometrica 85, 1613–1628.

Grammig, J., Schrimpf, A., Schuppli, M., 2009. Long-horizon consumption risk and the cross-section of returns: new tests and international evidence. Eur. J. Finance 15, 511–532.
Hahn, J., Lee, H., 2006. Yield spreads as alternative risk factors for size and book-to-market. J. Financ. Quant. Anal. 41, 245–269.
Hansen, L., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1084.
Hansen, L., Heaton, J., Yaron, A., 1996. Finite sample properties of some alternative GMM estimators. J. Bus. Econ. Stat. 14, 262–280.
He, Z., Kelly, B., Manela, A., 2017. Intermediary asset pricing: new evidence from many asset classes. J. Financ. Econ. 126, 1–35.
Kan, R., Zhang, C., 1999a. GMM tests of stochastic discount factor models with useless factors. J. Financ. Econ. 54, 103–127.
Kan, R., Zhang, C., 1999b. Two-pass tests of asset pricing models with useless factors. J. Finance 54, 203–235.
Kleibergen, F., 2009. Tests of risk premia in linear factor models. J. Econom. 149, 149–173.
Kleibergen, F., Zhan, Z., 2015. Unexplained factors and their effects on second-pass R-squared's. J. Econom. 189, 101–116.
Kleibergen, F., Zhan, Z., 2020. Robust inference for consumption-based asset pricing. J. Finance 75, 507–550.
Kroencke, T., 2017. Asset pricing without garbage. J. Finance 72, 47–98.
Kroencke, T.A., Thimme, J., 2023. A Skeptical Appraisal of Robust Asset Pricing Tests. Working Paper.
Lewellen, J., Nagel, S., Shanken, J., 2010. A skeptical appraisal of asset pricing tests. J. Financ. Econ. 96, 175–194.
Lioui, A., Maio, P., 2014. Interest rate risk and the cross-section of stock returns. J. Financ. Quant. Anal. 49, 483–511.
Maio, P., 2013a. Intertemporal CAPM with conditioning variables. Manag. Sci. 59, 122–141.
Maio, P., 2013b. Return decomposition and the intertemporal CAPM. J. Bank. Finance 37, 4958–4972.
Maio, P., Santa-Clara, P., 2012. Multifactor models and their consistency with the ICAPM. J. Financ. Econ. 106, 586–613.
Manresa, E., Peñaranda, F., Sentana, E., 2023. Empirical evaluation of overspecified asset pricing models. J. Financ. Econ. 147, 338–351.
Menkhoff, L., Sarno, L., Schmeling, M., Schrimpf, A., 2012. Carry trades and global foreign exchange volatility. J. Finance 67, 681–718.
Menkhoff, L., Sarno, L., Schmeling, M., Schrimpf, A., 2013. Currency Risk Premia and Macro Fundamentals. Working Paper.
Newey, W., Windmeijer, F., 1994. Generalized method of moments with many weak moment conditions. Econometrica 77, 687–719.
Parker, J., Julliard, C., 2005. Consumption risk and the cross section of expected returns. J. Polit. Econ. 113, 185–222.
Penaranda, F., Sentana, E., 2015. A unifying approach to the empirical evaluation of asset pricing models. Rev. Econ. Stat. 97, 412–435.
Savov, A., 2011. Asset pricing with garbage. J. Finance 66, 177–201.
Stock, J., Wright, J., 2000. GMM with weak identification. Econometrica 68, 1055–1096.
Tedongap, R., 2015. Consumption volatility and the cross-section of stock returns. Rev. Finance 19, 367–405.
Yogo, M., 2006. A consumption-based explanation of expected stock returns. J. Finance 61, 539–580.