DEUTSCHE BUNDESBANK
EUROSYSTEM

Discussion Paper

Deutsche Bundesbank
No 62/2020

GMM weighting matrices in cross-sectional asset pricing tests

Nora Laurinaityte
(Goethe University Frankfurt)

Christoph Meinerding
(Deutsche Bundesbank)

Christian Schlag
(Goethe University Frankfurt and Leibniz Institute for Financial Research SAFE)

Julian Thimme
(Karlsruhe Institute of Technology)

Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0

Please address all orders in writing to: Deutsche Bundesbank,
Press and Public Relations Division, at the above address or via fax +49 69 9566-3077

Internet http://www.bundesbank.de

ISBN 978–3–95729–796–9 (Printversion)
ISBN 978–3–95729–797–6 (Internetversion)

Non-technical summary

Research Question

Differences in expected returns of stocks are typically explained by the fact that these returns covary to different degrees with a particular risk factor. The so-called market price of risk (MPR) quantifies the compensation (in terms of expected return) per unit of risk. In order to be able to estimate the MPR, the factor mean must be known in advance. In practice, however, this mean is often estimated jointly with the MPR. This procedure creates a trade-off between an estimate of the factor mean that is as precise as possible and a high explanatory power of the model for asset prices. We examine to which extent this trade-off can bias the estimation results in the sense that irrelevant factors appear relevant or relevant factors appear irrelevant.

Contribution

We analyze the abovementioned trade-off for a particular estimator, which is frequently used in the literature due to its simplicity, and which is implemented via the generalized method of moments (GMM). We show that this estimator can incorrectly indicate a high explanatory power for a factor, i.e. a high cross-sectional R^2, if too little weight is put on a precise identification of the factor mean. In this case, the estimator suggests that the examined model approximates the true data-generating process for stock returns well, although this is not the case. We reveal the origins of this bias theoretically, gauge its size using simulations and document its practical relevance with a representative empirical example.

Results

We show that the GMM weighting matrix, which controls the abovementioned trade-off, is of first-order importance. We identify two major problems. First, irrelevant factors are considered relevant, i.e. the estimated MPR is large although the true MPR is zero. Depending on the choice of the weighting matrix, the estimated cross-sectional R^2 can take any value between 0 (the true value) and 1 (the maximum value). Second, weak and irrelevant factors can crowd out strong and truly relevant factors when estimating a multi-factor model. The estimated MPRs for irrelevant factors are then incorrectly large, while for relevant factors they are small and indistinguishable from zero. Both problems generally apply to all kinds of factors. They occur regardless of the sample size and are the more severe the larger the number of securities in the estimation.

Nichttechnische Zusammenfassung

Fragestellung

Unterschiede in erwarteten Renditen von Aktien werden in der Regel dadurch erklärt, dass diese Renditen in unterschiedlichem Maße von einem bestimmten Risikofaktor abhängen. Der sogenannte Marktpreis des Risikos (MPR) quantifiziert dabei die Kompensation (in Form von erwarteter Rendite) pro Einheit des eingegangenen Risikos. Um den MPR schätzen zu können, muss der Mittelwert des Faktors bekannt sein. In der Praxis wird dieser Mittelwert aber häufig erst gemeinsam mit dem MPR geschätzt. Dadurch entsteht allerdings ein Trade-off zwischen einer möglichst präzisen Schätzung des Faktormittelwerts und einem hohen Erklärungsgehalt des Modells für Wertpapierpreise. Wir untersuchen die Frage, inwieweit dieser Trade-off die Schätzergebnisse in dem Sinne verfälschen kann, dass irrelevante Faktoren als relevant oder relevante Faktoren als irrelevant erscheinen.

Beitrag

Wir analysieren den genannten Trade-off für einen speziellen, aufgrund seiner Einfachheit in der Literatur häufig eingesetzten Schätzer, der mittels der verallgemeinerten Momentenmethode (GMM) implementiert wird. Wir zeigen, dass dieser Schätzer fälschlicherweise einen hohen Erklärungsgehalt des betrachteten Faktors für die Renditen, also ein hohes Querschnitts-R^2 anzeigen kann, wenn zu wenig Gewicht auf eine genaue Identifikation des Faktormittelwerts gelegt wird. Der Schätzer suggeriert also in diesem Fall, dass das untersuchte Modell den wahren datenerzeugenden Prozess für Aktienrenditen gut approximiert, obwohl dies nicht der Fall ist. Wir leiten die Ursprünge dieses Problems theoretisch her, analysieren das Ausmaß des Problems mithilfe von Simulationen und dokumentieren die praktische Bedeutung des Problems anhand eines repräsentativen empirischen Beispiels.

Ergebnisse

Wir zeigen, dass die GMM-Gewichtungsmatrix, die den genannten Trade-off steuert, für den untersuchten Schätzer von herausragender Bedeutung ist. Wir identifizieren dabei zwei Hauptprobleme. Erstens werden in Wahrheit irrelevante Faktoren für relevant gehalten, d.h. der geschätzte MPR ist groß, obwohl der wahre MPR null ist. Das geschätzte Querschnitts-R^2 kann dabei je nach Wahl der Gewichtungsmatrix jeden beliebigen Wert zwischen 0 (dem wahren Wert) und 1 (dem Maximalwert) annehmen. Zweitens können schwache und irrelevante Faktoren starke und in Wahrheit relevante

Faktoren verdrängen, wenn ein Modell mit mehreren Faktoren geschätzt wird. Fälschlicherweise sind die geschätzten MPR für irrelevante Faktoren dann groß, während sie für relevante Faktoren klein und nicht von null unterscheidbar sind. Beide Probleme betreffen grundsätzlich alle Arten von Faktoren. Sie treten unabhängig von der Stichprobengröße auf und sind umso schwerwiegender, je mehr Wertpapiere in der Schätzung verwendet werden.

Deutsche Bundesbank Discussion Paper No 62/2020

GMM Weighting Matrices in Cross-Sectional Asset Pricing Tests

Nora Laurinaityte[§] Christoph Meinerding[†] Christian Schlag[‡] Julian Thimme[*]

This version: November 4, 2020

This paper contains charts in color. Use color printer for best results.

Abstract: Cross-sectional asset pricing tests with GMM can generate spuriously high explanatory power for factor models when the moment conditions are specified such that they allow the estimated factor means to substantially deviate from the observed sample averages. In fact, by shifting the weights on the moment conditions, any level of cross-sectional fit can be attained. This property is a feature of the GMM estimation design and applies to strong as well as weak factors, and to all sample sizes and test assets. We reveal the origins of this bias theoretically, gauge its size using simulations, and document its relevance empirically.

Keywords: Asset pricing, cross-section of expected returns, GMM, factor zoo

JEL classification: G00, G12, C21, C13

[§] Goethe University Frankfurt, Theodor-W.-Adorno-Platz 3, 60323 Frankfurt, Germany. E-mail: nora.laurinaityte@hof.uni-frankfurt.de, phone +49 69 798 33738

[†] Deutsche Bundesbank, Research Centre, Wilhelm-Epstein-Straße 14, 60431 Frankfurt, Germany. E-mail: christoph.meinerding@bundesbank.de, phone +49 69 9566 7359

[‡] Goethe University Frankfurt and Leibniz Institute for Financial Research SAFE, Theodor-W.-Adorno-Platz 3, 60323 Frankfurt, Germany. E-mail: schlag@finance.uni-frankfurt.de, phone +49 69 798 33699

[*] Karlsruhe Institute of Technology, Blücherstraße 17, 76185 Karlsruhe, Germany. E-mail: thimme@kit.edu, phone +49 721 681 41380

The authors would like to thank Geert Bekaert, Andrei Goncalves, Campbell Harvey, Stephan Jank, Emanuel Moench, Andreas Neuhierl, Jantje Soenksen, participants of the virtual MFA Meeting 2020, the SGF conference 2018 in Zurich, the Frontiers of Factor Investing conference 2018 in Lancaster and seminar participants at Goethe University Frankfurt and Deutsche Bundesbank for 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. Laurinaityte and Schlag gratefully acknowledge research and financial support from the Research Center SAFE, funded by the State of Hessen initiative for research LOEWE. The Matlab files for replication are available on `www.julianthimme.de/code`.

1 Introduction

Cross-sectional asset pricing tests with GMM can generate spuriously low pricing errors and high cross-sectional R^2's, which make the tested models appear close to the true data-generating process when they should actually be rejected. We reveal the origins of this bias theoretically, gauge its size in a controlled environment using Monte Carlo simulations, and document its relevance with an empirical example.

We do not criticize the use of GMM for cross-sectional asset pricing per se, but a particular design of a GMM estimator which has been used by several researchers, e.g., Yogo (2006), Dhume (2010), Darrat, Li, and Park (2011), Maio and Santa-Clara (2012), Maio (2013), Lioui and Maio (2014), Da, Yang, and Yun (2016), and Chen and Lu (2017). This estimator uses two types of moment conditions. The first, $0 = E\left[R_i^e - R_i^e(F-\mu)\lambda\right]$, states that expected excess returns must be linear in the covariance between excess returns R^e and candidate factors F. The second moment condition, $0 = E[F-\mu]$, identifies the factor mean μ.

This estimation approach is popular because it is very lean. 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, but does not require the identification of factor exposures (betas). The alternative approach would comprise time series regressions to estimate betas and a regression of average returns on estimated betas, either in two steps (see Fama and MacBeth, 1973) or in a single step via GMM (see, e.g., Cochrane, 2005). To give an example, when estimating a three factor model with 25 test assets, the latter approach estimates 103 parameters with 125 moment conditions. The lean approach discussed in our paper implies 28 moment conditions for six parameters. Asymptotically, the two approaches must deliver the same estimates of the factor risk premia, as long as the asset pricing model is correctly specified, but keeping the number of moment conditions small, relative to the sample size, is crucial to obtain reliable estimates in GMM (see, e.g., Newey and Windmeijer, 1994).

In applications, asset pricing models are always misspecified, i.e., the moment conditions never hold exactly. Hence, the GMM weighting matrix has an impact on the parameter estimates. We show that standard choices for that matrix, such as the identity matrix or the "optimal" weighting matrix, estimated via iterated or continuously updating GMM, lead to biased parameter estimates and inflated model performance statistics. In a nutshell, too small a weight on the second moment condition can lead to an imprecise estimate of μ in favor of an improved cross-sectional fit, i.e., smaller pricing errors. This problem applies to weak and strong factors alike, it occurs for all sample sizes, and it is the more severe, the larger the number of test assets.

We identify two major problems: First, unpriced factors look priced, i.e., the estimated market price of risk (MPR) is large and significant, although the true MPR is close to zero. This scenario represents a type I error, in the sense that the null MPR = 0 is falsely rejected. The corresponding estimated pricing performance can be heavily inflated, in the sense that pricing errors seem small and cross-sectional R^2's seem large. We show that the estimated R^2 can take every 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 test asset returns in the time series.

Second, when estimating a linear factor model with multiple factors, weak and unpriced factors can "drive out" strong and priced factors. More precisely, MPR estimates of weak factors are large and significant, while those of strong and priced factors are small and insignificant. The latter finding represents a type II error, in the sense that the null MPR = 0 is not rejected although it actually should be rejected.

We exemplify these issues using a quarterly post war sample of the usual 25 size and book-to-

Table 1: **Market prices of risks for the elephant model**

	A: High weight on μ			B: Unit weight on μ			C: Low weight on μ		
MKT	3.10		3.43	3.07		0.38	0.50		-0.01
	(2.70)		(2.88)	(3.34)		(0.46)	(5.85)		(-0.09)
Elephants		595.71	0.03		101.25	128.50		5.90	7.64
		(0.98)	(0.00)		(2.93)	(5.81)		(3.24)	(5.88)
SMB			0.32			-1.72			-0.10
			(0.18)			(-1.19)			(-1.25)
HML			5.76			-0.47			-0.07
			(3.52)			(-0.39)			(-1.05)
RMSE (in %)	3.81	6.16	2.02	3.78	1.26	1.04	1.71	0.17	0.13
R^2 (in %)	-52.96	-213.03	56.64	-49.04	91.05	94.00	96.46	99.97	99.98

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) t-statistics (in parentheses) are calculated using a Bartlett kernel.

market sorted portfolios[1] as test assets and the following set of factors: (i) the market factor, a strong but basically unpriced factor (see Fama and French, 1992), (ii) SMB and HML, which are strong and priced, 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.[2] The elephant factor is a weak factor, as its time series correlation with returns is basically zero for all test assets.

Table 1 shows MPR estimates and model performance statistics for three different GMM weighting matrices. For Panel A, we assign the moment conditions that identify the factor means a high weight to make sure it is correctly estimated. The results are the same as those from a simple two-stage Fama and MacBeth (1973) regression. The model with the market return as the only factor has a negative R^2, despite the significant point estimate (which is due to the fact that we run a cross-sectional regression without intercept). The elephant factor is insignificant and does not explain cross-sectional return variation. HML betas are strongly positively related to average returns in the cross-section, so that the model with four factors has a cross-sectional R^2 of around 57%.

Panel B shows the same statistics, estimated with an identity matrix for weighting the moment conditions. The estimated cross-sectional R^2 of the elephant model is 91%, and the MPR estimate is statistically significantly different from zero with a t-statistic of 2.93. Moreover, the elephant factor drives out all three Fama-French factors. The cross-sectional R^2 is basically the same for the one-factor and the four-factor model (91% and 94%), suggesting that just the elephant factor is enough to explain the cross-section of expected returns.

For Panel C, we assign a lower than unit weight on the moment conditions that identify the factor means. The one-factor model featuring the strong market factor now shows a similar behavior as the elephant factor in Panel B. The estimated cross-sectional R^2 is close to 1 and

[1] Although the example focuses on the Fama-French test portfolios and a postwar sample, we show that this artifact emerges for all sets of test assets and all sample sizes. It is, thus, different from the discussions on "lucky factors" (see Harvey, Liu, and Zhou, 2016) and factor structure in the Fama-French portfolios (see Lewellen, Nagel, and Shanken, 2010).

[2] 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.

the pricing errors are close to zero.

The key insight to understand these results is that the first moment condition

$$0 = E\left[R_i^e - R_i^e(F-\mu)\lambda\right] \quad = \quad E[R_i^e] - E[R_i^e]E[F-\mu]\lambda - Cov(R_i^e, F)\lambda$$

has trivial solutions. These are characterized by the relation $\mu \approx E[F] - \frac{1}{\lambda}$ between μ and λ and by λ being close to zero. Of course, with a small λ, the difference between μ and $E[F]$ is large, implying a high value for the second moment condition $E[F-\mu]$. Still, the parameter estimates end up in such a trivial solution if the moment $E[F-\mu]$ is assigned a small weight. Our paper elaborates on various manifestations of this rationale theoretically, empirically, and via Monte Carlo simulations.

The controlled environment of our simulation study, in which we know the true parameters, is useful to gauge the size of the bias. In applications, however, we do not know the true factor mean and the true market price of risk, so there is no clear benchmark to which the estimates should be compared. Still, the results presented in our paper give rise to a "smell test". One advantage of GMM estimators is that they trade off the uncertainty about different moments, here the uncertainty about the true factor means and the expected test asset returns. Choosing the weights in a prespecified weighting matrix corresponds to taking a stand on this trade-off. Our results help to diagnose whether we give the factor mean too much liberty in favor of a good cross-sectional fit.

There are alternatives to a prespecified weighting matrix. In general, one can have the algorithm put the highest weights on the most informative moments, by choosing the inverse of the covariance matrix of the moment conditions. We show in a controlled environment that using the true covariance matrix can mitigate the problem. However, in applications, the true covariance matrix is unknown and needs to be estimated. We show that an iterated GMM estimation does not converge to the true covariance matrix and the point estimates depend on the weighting matrix in the first iteration. Continuously updating GMM, i.e. optimizing over moments and weights simultaneously, does not converge because the estimated covariance matrix is close to singular around the trivial solution described above.

Our paper is structured as follows. Section 2 reviews the related literature. In Section 3, we explain the bias theoretically for strong and weak, priced and unpriced factors, which can be considered limiting cases. In Section 4, we conduct a detailed simulation analysis for the full range of cases in between. While these two sections focus on prespecified weighting matrices, Section 5 considers alternative choices. In Section 6, we reconsider the factor model proposed by Yogo (2006), which has been tested using the GMM procedure analyzed in our paper. It features the market return, the growth rate of nondurables and services consumption as well as the growth rate of durable consumption as factors. For 25 size- and book-to-market sorted portfolios he finds a cross-sectional R^2 of 0.935 and a highly significant market price of risk of around 170 for durable consumption growth. Starting from an exact replication of this result, we modify the GMM weighting matrix until we end up with a cross-sectional R^2 of 0.01 and a market price of risk of -138. Section 7 concludes and discusses approaches for applied researchers to diagnose and avoid the described bias.

2 Related literature

Our paper is linked to several strands of the literature. First of all, the use of GMM for cross-sectional asset pricing has become standard practice and is covered extensively in textbooks like Cochrane (2005), Campbell (2017), or Ferson (2019). However, although providing guidelines how to use GMM in practice, to our knowledge, none of these textbooks explicitly cover the peculiarities of the moment conditions discussed in our paper.

Besides, there are papers dealing with the econometric details of cross-sectional asset pricing with weak factors, like, e.g., Kan and Zhang (1999a,b), Kleibergen (2009), Gospodinov, Kan, and Robotti (2014), Kleibergen and Zhan (2015, 2020), Bryzgalova (2016), or Burnside (2016). Stock and Wright (2000) develop the general asymptotic theory for GMM estimators with weak identification. However, the issue that we raise in our paper applies to both weak and strong factors, irrespective of the sample size or the choice of a particular set of test assets.

Moreover, the concern raised in our paper is very different from the issue discussed in Lewellen et al. (2010). First, these authors argue that using the standard 25 Fama-French portfolios is problematic because the returns on these portfolios have a strong factor structure which is easy to pick up by chance when testing a factor. Our argument against the use of a specific form of GMM is relevant for *all* choices of test assets. The theory in our paper holds regardless of the choice of the test portfolios and of the structure of the true data-generating process. Second, Lewellen et al. (2010) conclude that the usual test statistics like cross-sectional R^2 or t-statistics are *uninformative* when using the Fama-French portfolios. We argue instead that these statistics can even be "set" (almost) arbitrarily via the GMM weighting matrix.

In the late 1990s a debate has emerged whether the classical Fama-MacBeth approach or the more advanced GMM-based method provides more reliable estimates of market prices of risk (see Kan and Zhou, 1999; Cochrane, 2001; Jagannathan and Wang, 2002). We do not take a stand in this debate. In particular, we do not argue that GMM should not be used for the estimation of cross-sectional models, but we rather point out a potentially severe problem with this method that researchers should be aware of.

The small sample properties of alternative GMM estimators are analyzed by Hansen, Heaton, and Yaron (1996). Their analysis suggests that *continuously updating GMM* (CUGMM), i.e., the joint estimation of parameters and the weighting matrix, performs better than the iterative approach. We find that CUGMM does not fix the problem described in our paper. Ronchetti and Trojani (2001) stick to the iterative approach, but suggest an alternative estimator which has a bounded influence function and is thus robust to outliers in the data. However, the concern raised in our paper is not resolved by this approach either as it does not originate from issues related to data quality or the exact distribution of error terms.

Section 6 of our paper is linked to the large literature investigating the performance of consumption-based asset pricing models for the pricing of the cross-section of expected returns. Major advances in this literature have been made recently by, e.g., Parker and Julliard (2005), Savov (2011), Ferson, Nallareddy, and Xie (2013), Boguth and Kuehn (2013), and Kroencke (2017). Some authors (rather indirectly) point towards the issue discussed in our paper, e.g., Parker and Julliard (2005) and Savov (2011). Instead of providing an in-depth analysis of the fundamental problem that we focus on, they rather suggest ad hoc procedures to robustify their empirical findings, such as fixing betas to their OLS counterparts.

Finally, besides the linear factor model which we criticize, the paper of Yogo (2006) also contains a non-linear model from which the linear model is derived by loglinearization. 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.

3 Cross-Sectional Regressions with GMM

This section consists of two parts. In the first, we discuss the GMM estimation approach under the assumption that the estimated model is correctly specified. In this case, the estimator is of course consistent. Here, we also introduce most of the notation and vocabulary used in the rest of the paper. In the second part, we discuss various situations of misspecified models. In the context of linear asset pricing models for the cross-section of expected returns, misspecification means that the cross-sectional R^2 is below 1 even asymptotically, i.e. that there is cross-sectional variation in expected returns that cannot be explained by exposures to the proposed factors.

3.1 Correctly specified models

Assume that there are n test assets with excess returns $R^e_{i,t}$ $(i = 1, \dots, n; t = 1, \dots, T)$ and k candidate pricing factors $F_{j,t}$ $(j = 1, \dots, k)$. An unconditional linear factor model implies that the expected excess returns of all assets are proportional to the assets' factor exposures:

$$\begin{aligned} E[R^e_i] &= \sum_{j=1}^{k} Cov(R^e_i, F_j)\lambda^*_j \\ \Leftrightarrow \quad E[R^e_i] &= \sum_{j=1}^{k} \left(E[R^e_i F_j]\lambda^*_j - E[R^e_i]E[F_j]\lambda^*_j\right) \qquad (1) \\ \Leftrightarrow \quad 0 &= E\left[R^e_i - \sum_{j=1}^{k} R^e_i (F_j - E[F_j])\lambda^*_j\right]. \end{aligned}$$

Here λ^*_j denotes the true market price of F_j-risk, scaled by the variance of F_j (i.e. $\lambda^*_j = \frac{MPR_j}{Var[F_j]}$), and is the parameter to be estimated.

The expected values of the factors, $E[F] = (E[F_1], \dots, E[F_k])'$, are typically unknown to the econometrician. Of course, $E_T[F] = \frac{1}{T}\sum_{t=1}^{T}(F_{1,t}, \dots, F_{k,t})'$ is an unbiased estimator, but with non-zero variance.[3] To set the uncertainty regarding the estimate for the factor means in relation to the uncertainty about $\lambda^* = (\lambda^*_1, \dots, \lambda^*_k)'$, the vector $E[F]$ is often replaced by a further parameter vector $\mu = (\mu_1, \dots, \mu_k)'$, which is estimated jointly with λ^*. Replacing $E[F]$ by μ and λ^* by λ, Equation (1) can be transformed as follows:

$$\begin{aligned} 0 &= E\left[R^e_i - \sum_{j=1}^{k} R^e_i (F_j - \mu_j)\lambda_j\right] \\ \Leftrightarrow E[R^e_i] &= \sum_{j=1}^{k} \left(E[R^e_i F_j]\lambda_j - E[R^e_i]\mu_j\lambda_j\right) \\ \Leftrightarrow E[R^e_i] &= E[R^e_i]\sum_{j=1}^{k}\Big(E[F_j] - \mu_j\Big)\lambda_j + \sum_{j=1}^{k} Cov(R^e_i, F_j)\lambda_j \end{aligned}$$

For the ease of exposition, consider the case of a model featuring a single factor (i.e., $k = 1$):

$$E[R^e_i] = \underbrace{E[R^e_i]\Big(E[F] - \mu\Big)\lambda}_{\text{constant term}} + \underbrace{Cov(R^e_i, F)\lambda}_{\text{covariance term}} \qquad (2)$$

If the model is correctly specified, the vector $(Cov(R^e_i, F))_i$ is non-zero and proportional to the vector of expected returns $(E[R^e_i])_i$, i.e. the true cross-sectional R^2 is equal to 1. In this case, the true parameter λ^* is simply given by $\lambda^* = \frac{E[R^e_i]}{Cov(R^e_i, F)}$ for any $i = 1, \dots, n$. However, Equation (2) can be solved in several ways. Indeed, substituting $Cov(R^e_i, F) = \frac{E[R^e_i]}{\lambda^*}$ in Equation (2) and

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

solving for μ gives

$$\mu = E[F] + \frac{1}{\lambda^*} - \frac{1}{\lambda}. \tag{3}$$

In other words, when estimated jointly with GMM, the parameters λ and μ cannot be identified separately using the pricing errors, i.e., the sample equivalent of Equation (2), as the only moment condition. Importantly, Equation (3) does not depend on i, so considering more test assets does not solve the problem.

There are two important special cases of the general set of solutions described by Equation (3). First, one can set $\mu = E[F]$ to eliminate the constant term in Equation (2). With this choice, λ is equal to λ^*, the true market price of risk divided by the factor variance. We refer to this solution as the *desired solution*.[4]

Second, consider cases where λ lies in a close neighborhood of zero, which implies that μ is very large in absolute terms. In the limit $\lambda \to 0$, the covariance term in Equation (2) vanishes and the constant term goes to one.[5] We refer to such cases as *trivial solutions*. They are trivial in the sense that the true statistical relation between factor exposures and expected returns is irrelevant, because the *covariance term* in Equation (2) can be made arbitrarily small. As opposed to the desired solution, a trivial solution yields parameter estimates which are far away from the true parameters.

In order to restore identification, it is a standard procedure to add a second set of moment conditions of the form $0 = E[F_j - \mu_j]$ for $j = 1, \ldots, k$, which identify the factor means.

In applications, the GMM point estimates are given by $(\hat{\lambda}, \hat{\mu})' = \text{argmin} f(\lambda, \mu)$, where

$$f(\lambda, \mu) = g_T'(\lambda, \mu)\, W\, g_T(\lambda, \mu). \tag{4}$$

Here, W denotes the GMM weighting matrix and g_T is the $(n+k) \times 1$-vector of sample moment conditions, given by

$$g_T(\lambda, \mu) = E_T \begin{bmatrix} R_i^e - \sum_{j=1}^k R_i^e (F_j - \mu_j)\lambda_j, & i = 1, \ldots, n \\ F_j - \mu_j, & j = 1, \ldots, k \end{bmatrix}. \tag{5}$$

We denote the second set of moment conditions as *penalty terms*, since they penalize deviations of $\hat{\mu}$ from the factor means.

The additional moment conditions influence the set of minima of the GMM objective function (4). Asymptotically, only the desired solution, i.e. $\hat{\lambda} = \lambda^*$ and $\hat{\mu} = E[F]$, will set all moment conditions to zero. It is, thus, the global minimum. A trivial solution, i.e. bringing the covariance term in Equation (2) close to zero by letting λ go to zero is now costly, since $\hat{\mu}$ will be far away from $E[F]$. All other "intermediate" solutions will also be penalized, since $E_T[F - \mu] = \frac{1}{\lambda} - \frac{1}{\lambda^*}$ is asymptotically nonzero unless $\lambda = \lambda^*$.

3.2 Misspecified models

We have shown above that, in case the model is correctly specified, the GMM estimator that minimizes the objective function (4) will asymptotically yield the true parameters, i.e., the desired solution. Obviously, this finding is just a special case of the general result of Hansen

[4] 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 is also used in the literature, see, e.g., the detailed discussion in Ferson (2019), p. 180, p. 220, or pp. 224ff.

[5] To see the latter, reformulate Equation (3) as $1 = \frac{\lambda}{\lambda^*} + (E[F] - \mu)\lambda$. The first term vanishes as λ goes to zero, so the second must go to one.

(1982). In applications, samples are small and asset pricing models are misspecified (see Kan and Zhang, 1999b; Fama and French, 2015). As a consequence, the sample covariances do not perfectly line up with the sample average returns, i.e., the cross-sectional R^2 is smaller than 1 and at least one pricing error is larger than 0. In this situation, there is no parameter vector $(\lambda, \mu)'$, for which $g_T(\lambda, \mu) = 0$. The minimum of the GMM objective function then depends on the choice of the weighting matrix W.

In the following, we show that certain standard choices of W lead to estimators suggesting that factors with poor pricing abilities appear important for pricing assets and so, heavily misspecified models appear close to correctly specified. For the ease of exposition, we study only weighting matrices of the form

$$W_x = \begin{pmatrix} I_n & 0 \\ 0 & 10^x\, I_k \end{pmatrix} = \mathrm{diag}(1, \ldots, 1, 10^x, \ldots, 10^x) \tag{6}$$

in this section.[6] With $W = W_x$, the GMM objective function is given by

$$f_x(\lambda, \mu) = SSE(\lambda, \mu) + 10^x \cdot SSP(\mu),$$

where SSE denotes the sum of squared pricing errors, i.e. the sum of the first n squared entries of g_T, and SSP denotes the sum of squared penalty terms. We have added the subscript x to the GMM objective function f to emphasize that the parameter x impacts the objective function, and, thus, potentially also the values that minimize it.

In the remainder of the section, we discuss the solutions of the optimization problem $\min f_x(\lambda, \mu)$ for a given x, making different assumptions about the candidate pricing factors. To characterize factor j, we consider the following decomposition of the vector of sample covariances of the test asset returns with factor j:

$$(Cov_T(R_i^e, F_j))_i = (Cov_{ij}^{ER})_i + (Cov_{ij}^{\perp})_i, \tag{7}$$

where $(Cov_{ij}^{ER})_i$ is a multiple of $(E_T[R_i^e])_i$ and $(Cov_{ij}^{\perp})_i$ is orthogonal to $(Cov_{ij}^{ER})_i$ in sample, i.e., $\sum_i Cov_{ij}^{ER} Cov_{ij}^{\perp} = 0$ for all j.

We discuss different cases with single factors first. The interaction of multiple factors is analyzed in Section 3.3. In case of a single factor, the model is misspecified if $Cov^{\perp}$ is different from zero. In this situation, our goal still is to estimate the parameter λ^* which solves $E[R_i^e] = Cov^{ER}\lambda^*$. The sample equivalent of the ith moment condition (2) now reads as

$$g_{T,i}(\lambda, \mu) = E_T[R_i^e] - \Bigg(\underbrace{E_T[R_i^e]\Big(E_T[F] - \mu\Big)\lambda}_{\text{constant term}} + \underbrace{Cov_i^{ER}\lambda}_{\text{priced covariance term}} + \underbrace{Cov_i^{\perp}\lambda}_{\text{unpriced covariance term}} \Bigg) \tag{8}$$

3.2.1 Case 1: Strong perfectly unpriced factors

We call a factor *perfectly unpriced* if $Cov^{ER} = 0$. In this case, the vector of average test asset returns is orthogonal to the vector of factor exposures, which implies a true cross-sectional R^2 of zero for the model featuring the perfectly unpriced factor as the only factor. In the following section, we focus on the case where $Cov^{\perp} \neq 0$. This implies $Cov(R_i^e, F)_i \neq 0$, and we therefore label the factor as *strong*. If, instead, $Cov(R_i^e, F)_i$ is equal to zero, we call the factor *perfectly weak*. We will treat the special case of perfectly weak factors in Section 3.2.3.[7]

[6] Alternative, more general weighting matrices are analyzed in Section 5.

[7] There is no consistent nomenclature in the literature. For instance, factors that are labeled "perfectly weak" in our paper are called "useless" by Kan and Zhang (1999b) and others.

We are looking for the parameters $(\lambda, \mu)'$ that minimize the GMM objective function f_x in Equation (4). To facilitate the understanding of the general solution, it is instructive to look at the minimization problem under the two constraints $\mu = E_T[F]$ and $\mu = E_T[F] - \frac{1}{\lambda}$ first. In accordance with the previous subsection, we label these two constrained optima as *desired constrained optimum* and *trivial constrained optimum*. As will become clear below, the solutions of the two constrained minimization problems represent corner solutions of the unconstrained minimization problem, as x goes to infinity and to minus infinity, respectively.

Desired constrained optimum: The constraint $\mu = E_T[F]$ implies a zero penalty term, but a nonzero pricing error. The GMM objective function f_x is thus equal to SSE and minimizing SSE also minimizes f_x for all x. The first n sample moment conditions are given by

$$g_{T,i}(\lambda, \mu) = E_T[R_i^e] - Cov_T(R_i^e, F)\lambda = E_T[R_i^e] - Cov^{\perp}\lambda.$$

The GMM objective function thus reads as

$$f_x(\lambda, E_T[F])) = SSE(\lambda, E_T[F]) = \sum_{i=1}^{n}(E_T[R_i^e])^2 + \lambda^2(Cov^{\perp})'Cov^{\perp}.$$

The minimum is obviously given by $\lambda = 0$, i.e.

$$f_x(0, E_T[F]) = SSE(0, E_T[F]) = \sum_{i=1}^{n}(E_T[R_i^e])^2.$$

Thus, under the constraint $\mu = E_T[F]$, we end up in the desired solution, which, in this case, implies that the sum of squared pricing errors is equal to the sum of squared returns. This means that the estimated cross-sectional R^2 is equal to zero, in line with the true R^2. Importantly, this solution does not depend on the log weight x on the penalty term.

Trivial constrained optimum: Under the constraint $\mu = E_T[F] - \frac{1}{\lambda}$, the constant term in Equation (8) is equal to one, such that the pricing errors are equal to the covariance term. The GMM objective function is given by

$$f_x((E_T[F] - \mu)^{-1}, \mu) = (E_T[F] - \mu)^{-2}(Cov^{\perp})'Cov^{\perp} + 10^x(E_T[F] - \mu)^2$$

and obviously depends on x. The first term, i.e., the sum of squared pricing errors, has a supremum at $\mu = E_T[F]$ and converges to zero as μ goes to minus or plus infinity. The second term, i.e. the sum of squared penalty terms, shows the opposite behavior, it is minimized at $\mu = E_T[F]$.

When x is very negative, minimizing f_x essentially boils down to minimizing the sum of squared pricing errors SSE. Similar to the desired constrained minimum discussed above, λ ends up being close to zero. The difference, however, is that the estimated cross-sectional R^2 is close to 1 here, since the pricing errors are close to zero, while the estimated R^2 is equal to 0 in the desired constrained case. Thus, for very negative values of x, the trivial constrained optimization returns a trivial solution with spuriously high R^2 and tiny pricing errors.

For large positive values of x, the penalty term requires μ to be close to $E_T[F]$, which implies high pricing errors. In fact, $f_x((E_T[F] - \mu)^{-1}, \mu)$ becomes arbitrarily large if x is chosen large enough. In particular, for x large enough, the minimum of $f_x((E_T[F] - \mu)^{-1}, \mu)$ with respect to μ exceeds the desired constrained minimum $f(0, E_T[F])$, which is finite. The minimum under the constraint $\mu = E_T[F] - \frac{1}{\lambda}$ thus cannot be the global minimum when x is sufficiently high.

Global optimum: We have considered the minima of the GMM objective function under the constraints $\mu = E_T[F]$, corresponding to an estimated R^2 of 0 (which is equal to the true R^2), and $\mu = E_T[F] - \frac{1}{\lambda}$, corresponding to an estimated R^2 of close to 1 as λ goes to zero. These two cases coincide with the corner solutions of the unconstrained minimization in the cases where x goes to plus or minus infinity.

For large positive x, this results from the fact that the penalty term requires μ to be close to $E_T[F]$, and this is, by definition, the condition for the desired constrained optimization.

On the other hand, if x is chosen negative enough, we have seen that the value of the GMM objective function in the trivial constrained optimization can be brought arbitrarily close to zero. We thus conclude that the trivial constrained optimum represents the corner solution of the global optimization problem as x goes to minus infinity.

For intermediate values of x, the unconstrained minimum will typically be between these two cases. As we will see in the simulation study in Section 4, the parameter estimates go from the $R^2 = 1$ case for low values of x to the $R^2 = 0$ case for high values of x and the point estimates $(\hat{\lambda}, \hat{\mu})$ go from $(0, \infty)$ to $(\lambda^*, E_T[F])$. Importantly, for seemingly natural choices of x, such as $x = 0$, we typically find estimates that are far away from the desired solution.

3.2.2 Case 2: Strong priced factors

We next consider a single strong factor which has some explanatory power for the cross-section of test asset returns, i.e., $Cov^{ER} \neq 0$, but we assume that the model is not fully correctly specified, i.e., $Cov^{\perp} \neq 0$. This is the case, we typically have to deal with in empirical applications. It turns out that the rationale in this case is very similar to the case of perfectly unpriced factors.

We are interested in the parameter λ^* that solves $E_T[R_i^e] = \lambda^* Cov^{ER}$. Moment condition (8) is then given as

$$\begin{aligned} g_{T,i}(\lambda, \mu) &= E_T[R_i^e] - \Big(E_T[R_i^e]\Big(E_T[F] - \mu\Big)\lambda + \lambda Cov^{ER} + \lambda Cov^{\perp}\Big) \\ &= E_T[R_i^e] - \Big(E_T[R_i^e]\Big(E_T[F] - \mu\Big)\lambda + \frac{\lambda}{\lambda^*} E_T[R_i^e] + \lambda Cov^{\perp}\Big). \end{aligned} \tag{9}$$

We consider the same constrained optima as in the previous subsection.

Desired constrained optimum: Under the constraint $\mu = E_T[F]$, the sum of squared pricing errors SSE and thus f_x has a minimum in $\lambda = \lambda^*$, and we have

$$f_x(\lambda^*, E_T[F]) = (\lambda^*)^2 \, (Cov^{\perp})' Cov^{\perp}. \tag{10}$$

These parameters correspond to the *desired solution* and yield an estimated cross-sectional R^2 which is equal to the true R^2, i.e., $1 - (Cov^{\perp})'Cov^{\perp}/((Cov^{ER})'Cov^{ER} + (Cov^{\perp})'Cov^{\perp})$.

Trivial constrained optimum: Setting the constant term in Equation (8) to one by constraining $\lambda = (E_T[F] - \mu)^{-1}$ gives

$$\begin{aligned} f_x((E_T[F] - \mu)^{-1}, \mu) = (E_T[F] - \mu)^{-2} \Big((Cov^{ER})'Cov^{ER} + (Cov^{\perp})'Cov^{\perp}\Big) \\ + 10^x (E_T[F] - \mu)^2 \end{aligned} \tag{11}$$

Taking the first order condition of Equation (11) with respect to μ gives

$$\mu^{opt} = E_T[F] \pm \sqrt[4]{\frac{(Cov^{ER})'Cov^{ER} + (Cov^{\perp})'Cov^{\perp}}{10^x}}. \tag{12}$$

Substituting this solution back into f_x gives the minimum value

$$f_x((E_T[F] - \mu^{opt})^{-1}, \mu^{opt}) = 2\sqrt{10^x \Big((Cov^{ER})'Cov^{ER} + (Cov^{\perp})'Cov^{\perp}\Big)} \tag{13}$$

Global optimum: From Equations (10), (12) and (13) we can again conclude that the two constrained optima represent the corner solution of the unconstrained optimization problem for large positive or negative x.

For large positive x, Equation (12) shows that the estimate of μ is close to $E_T[F]$ also under the constraint $\lambda = (E_T[F] - \mu)^{-1}$. However, from Equation (13), we see that the value $f_x((E_T[F] - \mu^{opt})^{-1}, \mu^{opt})$ exceeds $f_x(\lambda^*, E_T[F])$ if x is large enough, so the trivial constrained optimum cannot be the global optimum. Instead, the corner solution, when x goes to plus infinity, is again given by the desired constrained optimum, exactly as in the previous subsection.

For large negative x, on the other hand, $f_x((E_T[F] - \mu^{opt})^{-1}, \mu^{opt})$ can get arbitrarily close to zero so that, again, the trivial constrained optimum represents the corner solution for the unconstrained minimization problem when x goes to minus infinity.

For finite but low enough weights, $f_x((E_T[F]-\mu^{opt})^{-1}, \mu^{opt})$ is smaller than $f_x(\lambda^*, E_T[F])$, which implies that the parameter estimates systematically move away from the true parameters as x gets smaller and smaller. We will analyze the relation between the true cross-sectional R^2, the weight x, and the parameter estimates more thoroughly in our simulation study in Section 4.

3.2.3 Case 3: (Perfectly) weak factors

Finally, we call a factor *perfectly weak* if $Cov^{ER} = 0$ and $Cov^{\perp} = 0$, i.e., if the time-series covariance $Cov_T(R_i^e, F)$ is equal to zero for all test assets $i = 1, \ldots, n$. Perfectly weak factors are necessarily perfectly unpriced. The results from Section 3.2.1 largely carry over to this case, except for the desired constrained optimization problem.

Desired contrained optimum: Under the constraint $\mu = E_T[F]$, the moment condition (8) becomes

$$g_{T,i}(\lambda, \mu) = E[R_i^e] - 0 \cdot \lambda,$$

which implies that λ cannot be identified, independent of the choice of x.

Trivial constraint optimum: Setting the constant term to one by constraining $\lambda = (E_T[F] - \mu)^{-1}$ automatically eliminates all pricing errors. The GMM objective function then comprises only the sum of squared penalty terms:

$$f_x((E_T[F] - \mu)^{-1}, \mu) = 10^x (E_T[F] - \mu)^2. \tag{14}$$

Independent of x, it has an infimum at $\mu = E_T[F]$, which is, however, not a minimum, since the initial constraint $\lambda = (E_T[F] - \mu)^{-1}$ is not defined at that point. Nevertheless, the GMM objective function can be brought arbitrarily close to zero by choosing μ arbitrarily close to $E_T[F]$. Hence, in applications, the GMM estimator always ends up in a trivial solution if a factor is perfectly weak. In terms of estimates, this means that $\hat{\mu} \approx E_T[F]$, $\hat{\lambda}$ is extremely large in absolute terms and the estimated cross-sectional R^2 is close to 1.

A perfectly weak factor is of course a theoretical knife-edge case. In applications, however, factors are often weak, in the sense that the sample covariances of all test asset returns with the candidate factor are very close to zero. In this case, the rationale outlined above can still serve as an intuition. If the covariance term in Equation (2) is close to zero, even highly positive or negative values for λ do not make the covariance term particularly costly when minimizing SSE. We will analyze the relation between weakness of a factor and estimation results more thoroughly in our simulation study in Section 4.

3.3 Multiple factors

With two or more factors, there is one constant and one covariance term per factor. To reduce the pricing errors to zero, it is enough to set the constant term of one factor to one, i.e., choose parameters that correspond to a *trivial solution* for only one factor. As we have seen in the one factor cases, it is particularly cheap to do so in terms of penalty terms when a factor is weak. As a consequence, in empirical applications like the one presented in the introduction, weak factors have the potential to "drive out" strong factors, when the GMM weighting matrix is chosen inappropriately.

To exemplify this rationale, consider the case of two factors F_1 and F_2, where F_2 is perfectly weak. Equation (8) then reads

$$g_{T,i}(\lambda,\mu) = E_T[R_i^e] - \Big(E_T[R_i^e]\Big(E_T[F_1]-\mu_1\Big)\lambda_1 + Cov_T(R_i^e,F_1)\lambda_1 + E_T[R_i^e]\Big(E_T[F_2]-\mu_2\Big)\lambda_2\Big). \quad (15)$$

One solution of this equation is given by $(\lambda_1,\lambda_2,\mu_1,\mu_2)' = (0,(E_T[F_2]-\mu_2)^{-1},E_T[F_1],\mu_2)'$. In this case, the sum of squared penalty terms is equal to $(E_T[F_2]-\mu_2)^2$ and can be made arbitrarily small if μ_2 is set close to $E_T[F_2]$. The pricing errors also vanish for these parameters, so they must minimize the GMM objective function globally.

Hence, when estimating a linear model with multiple factors, one weak factor is enough to end up with an estimated cross-sectional R^2 of one and zero pricing errors, independent of the weighting matrix. Importantly, the estimated parameters are far away from the true parameters in a particularly misleading fashion. The market price of risk λ_1 of the strong and (imperfectly) priced factor is estimated at zero and, thus, appears irrelevant. Moreover, the market price of risk λ_2 of the perfectly weak factor is seemingly extremely large, either positive or negative.

As we will see in our application with simulated data in Section 4, factors which are "almost" perfectly weak (i.e., for which the time series covariances with the test asset returns are non-zero, but small) lead to similar results when combined with strong factors. Importantly, just as stated above, λ_1 is estimated to be equal to zero in these cases, falsely implying that the first factor was unpriced, even if it is actually (imperfectly) priced. Moreover, since the factor mean of the weak factor is typically close to the factor mean in sample, λ_2 is estimated to be large and significant.

We discuss the situation of two factors with different strengths and different explanatory powers for the cross-section of test asset returns thoroughly in the simulation study in Section 4.

4 Simulation Study

Having discussed the behavior of the GMM objective function in the limits $x \to \pm\infty$ and for instructive special cases, we now analyze the full spectrum of possible factors and GMM weighting matrices in a simulation study. The general idea is as follows. We make various choices for the data-generating processes of factors and of test portfolio excess returns. For each choice, we draw one representative monthly sample of factor realizations and excess returns. Finally, we run the GMM estimation on each of these samples for various choices of the weighting matrix W.[8]

[8] This approach is different from other studies that simulate many samples to analyze the finite sample properties of estimators. As shown in Section 4.5, the properties we describe do not depend on the sample size. As discussed below, we make sure that all properties of factors and returns even hold in sample.

4.1 Data-generating processes

The factors $F_1, \ldots, F_k$ are drawn independently from an i.i.d. normal distribution with means and standard deviations which are all set to 1 percent per quarter. The data generating process for excess returns is given as

$$R^e_{i,t} = E[R^e_{i,t}] + \sum_{j=1}^{k} \beta_{i,j}(F_{j,t} - E[F_{j,t}]) + \sigma_i \, \varepsilon_{i,t}, \tag{16}$$

where the ε's are independent from one another and from the factors and are also i.i.d. normally distributed with means and standard deviations of 1 percent. In particular, the standard deviation of the factors, σ_f, is equal to the standard deviation of the error terms, σ_ε, which will facilitate the interpretation of the coefficients $\beta_{i,j}$ and σ_i as explained in detail below. They are chosen such that the volatility of excess returns, $\sigma_r = \sqrt{\beta^2_{i,j}\sigma^2_f + \sigma^2_i\sigma^2_\varepsilon}$, is 0.06 for all assets.

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 4.5. The factor time series as well as the ε's are sampled from an i.i.d. normal distribution and orthogonalized subsequently, to make sure that they are perfectly orthogonal even in our small sample.

The cross-sectional variation in expected returns $E[R^e_{i,t}]$ is modeled as follows. We draw a true factor exposure $b_{i,j}$ of the return of asset i to factor j from a normal distribution with 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 $(b_{i,j})_{i=1,\ldots,n}$ for different factors j are supposed to be orthogonal. In our simulation study, we draw vectors of factor exposures, orthogonolize them, and scale them subsequently, such that they all have a mean and standard deviation of 1 percent in sample. To allow for model misspecification and varying degrees of explanatory power of the factor exposures in our simulation framework, we set

$$E[R^e_{i,t}] = \sum_{j=1}^{k} (r_j b_{i,j} + \sqrt{1 - r_j^2}\, e_{i,j}), \tag{17}$$

where the coefficients r_j are chosen between 0 and 1 and the $e_{i,j}$'s have exactly the same properties as the $b_{i,j}$'s, but are orthogonal to them.

To exemplify the intuition, assume that there is only one factor, such that the expected excess return of asset i is given as $rb_i + \sqrt{1 - r^2}e_i$. Then, the cross-sectional correlation between the vector of factor exposures and the vector of expected returns is r. In other words, the true cross-sectional R^2 is equal to r^2. With multiple factors, the factor model's cross-sectional R^2 is given by $\frac{1}{k}\sum_{j=1}^{k} r_j^2$, i.e. factor j contributes r_j^2/k to the overall cross-sectional R^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 set

$$\beta_{i,j} = \frac{\sigma_r}{\sigma_f} \cdot s_j \cdot \frac{b_{i,j}}{\max(|b_{\cdot,j}|)}, \tag{18}$$

$$\sigma_i = \frac{\sigma_r}{\sigma_f} \sqrt{1 - \sum_{j=1}^{k} \left(s_j \cdot \frac{b_{i,j}}{\max(|b_{\cdot,j}|)} \right)^2}, \tag{19}$$

so that the time series correlation between factors and returns is controlled by the choice variables $s_j \in [0,1]$.

Again, the intuition is best understood in the case of a single factor. Equations (18) and (19) then reduce to $\beta_i = \frac{\sigma_r}{\sigma_f} \cdot s \cdot b_i / \max(|b|)$ and $\sigma_i = \frac{\sigma_r}{\sigma_f} \cdot \sqrt{1 - s \cdot b_i/\max(|b|)}$, respectively. The time series R^2 of the factor model for return i is given by

$$\frac{(\beta_i \sigma_f)^2}{(\beta_i \sigma_f)^2 + (\sigma_i \sigma_\varepsilon)^2} = \frac{\beta_i^2}{\beta_i^2 + \sigma_i^2} = s^2 \frac{b_i}{\max(|b|)}, \tag{20}$$

where we make use of the fact that $\sigma_f = \sigma_\varepsilon$. Normalizing the β_i and σ_i by $\max(|b|)$ ensures that s can easily be interpreted as the time series R^2 of the asset with the highest absolute factor exposure. Assets with lower factor exposures have lower time series R^2 accordingly.[9] The scaling factor $\frac{\sigma_r}{\sigma_f}$ ensures that the volatility of excess returns is equal to 0.06 for all assets, as explained above. This implies an annual return volatility of 20.78% for all test assets.

To sum up, the true cross-sectional R^2 is always equal to $\frac{1}{k}\sum_{j=1}^{k} r_j^2$, the true μ_j is always equal to 0.01 by assumption, and, from plugging (18) into (17), we obtain the true λ_j as

$$\lambda_j = \begin{cases} \frac{r_j \max(|b_{\cdot,j}|)}{s_j \sigma_r \sigma_f} & \text{if } s_j > 0, \\ \text{not identified} & \text{if } s_j = 0. \end{cases}$$

In the following, we analyze the behavior of the GMM estimator using data from the time series and cross-sectional model introduced above. We always hold all parameters fixed, with the exception of s, r and the GMM weighting matrix W.

4.2 The impact of true R^2

We start by analyzing the impact of the GMM weighting matrix for factors with varying cross-sectional explanatory power, i.e., for different values for r. Throughout this section, we analyze strong factors only, i.e. we set $s = 1$, and we stick to one factor models and weighting matrices W_x of the form studied in Section 3.

We first consider a single strong but perfectly unpriced factor, i.e. we set $s = 1$ and $r = 0$. We vary the log weight x on the penalty term and study the effects on the estimated R^2, RMSE, point estimates, and 95% confidence bands. Figure 1 shows these quantities for values of x between -5 and 5.

We find exactly the pattern described in Section 3.2.1. If x exceeds a critical value, in this case around 0.4, the estimates are in line with the true values, i.e. an R^2 of zero, a root mean squared pricing error that is equal to the cross-sectional standard deviation of expected returns, and point estimates $\hat{\lambda} = 0$ and $\hat{\mu} = 0.01$.[10] For such high values of x, the weight on the penalty term is so high that selecting a $\hat{\mu}$ different from the true factor mean in order to reduce the pricing error is too costly.

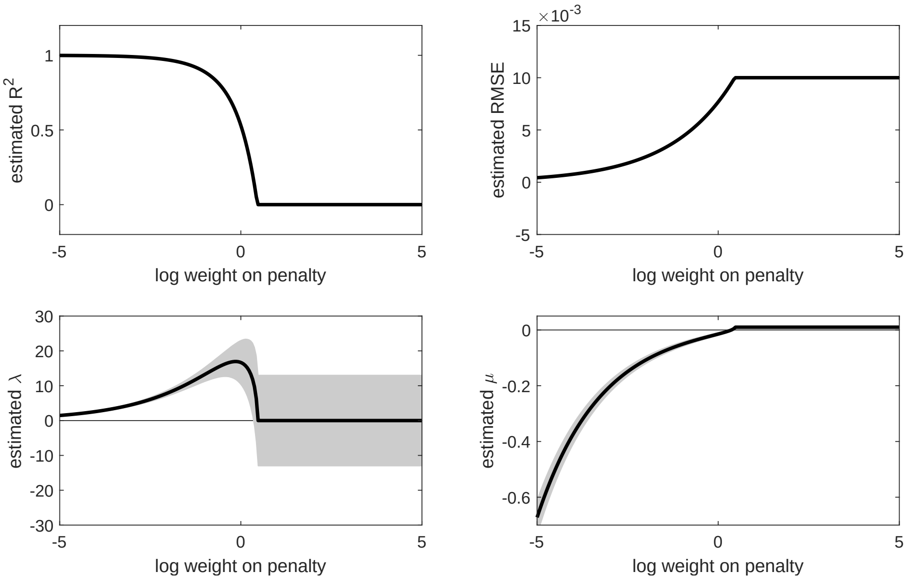
For $x \to -\infty$, the figure suggests that the estimated R^2 goes to 1, the RMSE goes to zero, and $\hat{\lambda}$ and $\hat{\mu}$ go to 0 and $-\infty$, respectively. This reflects what we labeled the trivial solution in the previous section.

[9] Note that an asset with a zero factor exposure necessarily has a time series R^2 of zero, so we decide to couple the individual asset's times series R^2 to its factor exposure in general. In settings with multiple factors, the values $s_1, \ldots, s_k$ have to be chosen such that the term under the root in Equation (19) is positive for all i. The most conservative way of doing so is to assume that $\sum_{j=1}^{k} s_j \leq 1$.

[10] Although the four functions shown in Figure 1 seem to have a kink at the critical value, they are actually differentiable. Making the grid finer in a close neighborhood of that point, we find that the first derivatives change rapidly but smoothly.

Figure 1:

A single strong and perfectly unpriced factor



We apply a GMM estimation with the moment conditions in Equation (5) and the weighting matrix in Equation (6). 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. Returns are simulated according to Equations (16)-(19) with $r = 0$ and $s = 1$.

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.

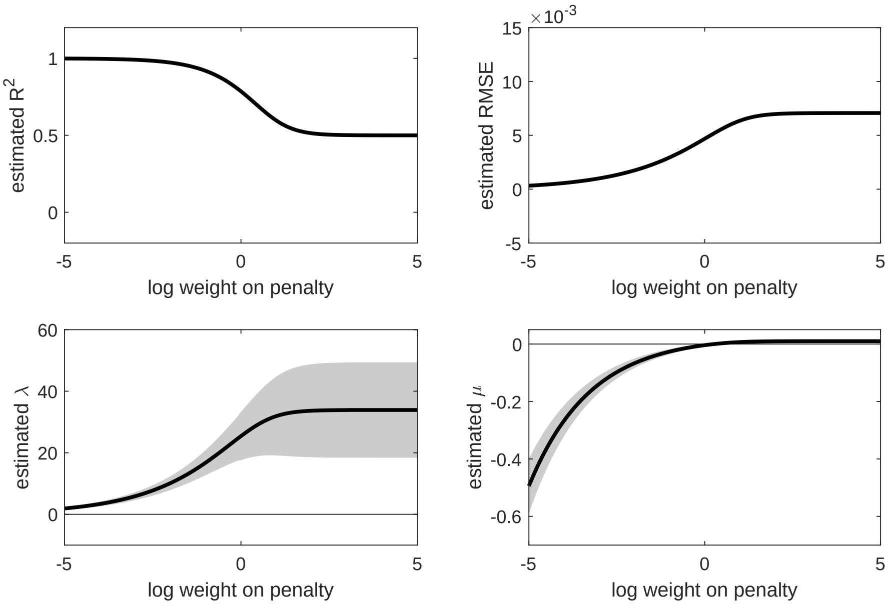
To understand the impact of the true cross-sectional R^2 on the estimated statistics, we next simulate data under the assumptions that the factor is strong and the cross-sectional R^2 is equal to 0.5, i.e. $r = \sqrt{0.5}$ and $s = 1$, which implies a true λ of 33.90.

Figure 2 shows the estimation results. They suggest that the case of (imperfectly) priced factors is very similar to the case of perfectly unpriced factors. For sufficiently high weights on the penalty term, the estimated R^2 and RMSE, as well as the point estimates of λ and μ are close to the true values. Low values of x, on the other hand, lead to inflated R^2's and biased parameter estimates. The figure again shows that even seemingly moderate choices of the weighting matrix, such as the identity matrix, can result in such a pattern.

Figure 3 shows the results for all possible choices of r (including the cases $r^2 = 0$ and $r^2 = 0.5$ from above). It depicts heatmaps of the estimated R^2 and the point estimate of μ as functions of x (on the horizontal axis) and the true R^2 (on the vertical axis). It is apparent that the estimated R^2 is far too high for low weights x when the true R^2 is small. However, we also observe that the parameter μ (and, thus, also λ which is close to $(E_T[F] - \mu)^{-1}$ in this region) is seriously biased even for high values of the true R^2 if x is low. This shows that, unless a factor is perfectly priced, there is potential for biased parameter estimates, if the weight on the factor mean is chosen inappropriately.

Figure 2:

A single strong factor with a cross-sectional R^2 of 0.5



We apply a GMM estimation with the moment conditions in Equation (5) and the weighting matrix in Equation (6). 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. Returns are simulated according to Equations (16)-(19) with $r = \sqrt{0.5}$ and $s = 1$.

4.3 The impact of factor strength

Next, we analyze the impact of the strength of a factor on the estimated R^2, RMSE, λ, and μ, by varying the coefficient s in Equations (18) and (19), while keeping r fixed at $\sqrt{0.5}$. As a starting point, Figure 4 shows the estimated quantities as functions of x for a very low s of $\sqrt{0.025}$.[11] With such a low value of s, most of the time series variation in the test asset returns come from the unsystematic part and the factor is only weakly correlated with them. Figure 2 can serve as a benchmark, since the true cross-sectional R^2 is the same in both figures and only the time series R^2, i.e., the strength of the factor, is reduced dramatically in Figure 4 compared to Figure 2.

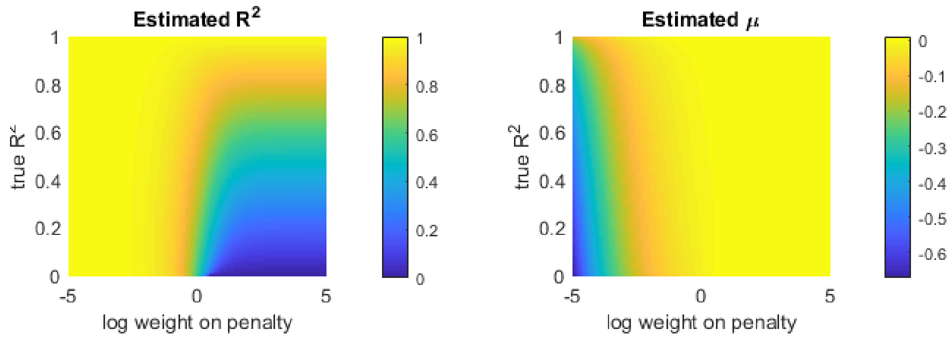
Qualitatively, we observe a pattern that is very similar to Figure 2. Quantitatively, one difference is that the true λ is equal to 214 with these parameters. More interestingly, the range in which the estimated R^2 drops from 1 to 0 has moved to higher values of x, relative to Figure 2. Thus, when fixing a particular weighting matrix, for example the identity matrix, biased parameter estimates and inflated R^2's are the more likely the weaker the factor being tested. As discussed in Section 3.2.3, if a factor is very weak, the covariance terms in Equation (8) are close to zero, so even a very high λ does not lead to high pricing errors under the constraint $\lambda = (E_T[F] - \mu)^{-1}$.

The heatmaps in Figure 5 show the estimated R^2 and μ as functions of the log weight x on the penalty and of s^2, the strength of the factor. The true cross-sectional R^2 of the factor is again set to 0.5. In line with the discussion above, we observe that for very weak factors (s close to zero), the estimated R^2 is close to 1 even for $x = 1$. Apart from such extreme choices of s, the

[11] Analyzing the case $s = 0$ is not feasible, because the estimation algorithm does not converge to a finite λ.

Figure 3:

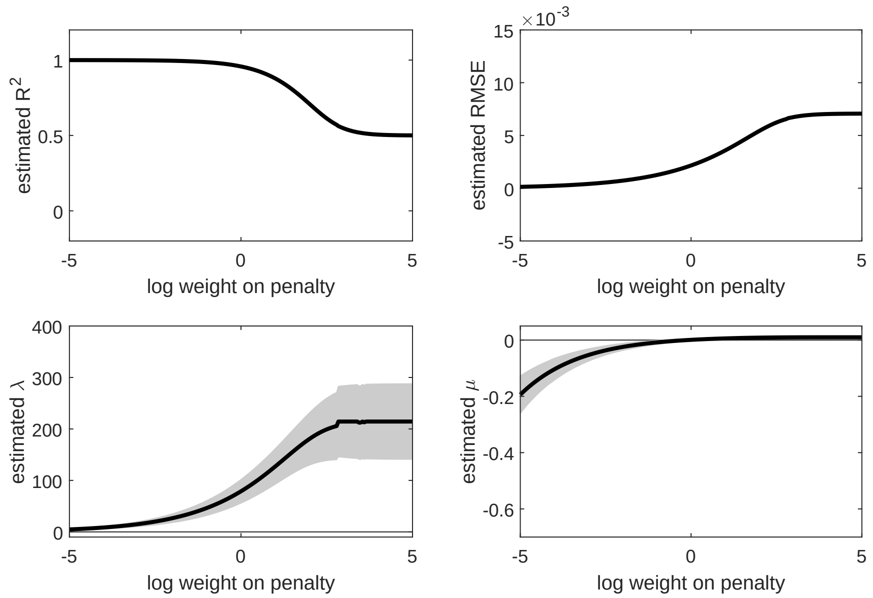
A single strong factor with a cross-sectional R^2 of 0.5



We apply a GMM estimation with the moment conditions in Equation (5) and the weighting matrix in Equation (6). The figure shows estimated R^2 and the point estimates of μ as functions of x, the log weight on the moment condition that identifies the factor mean, and the true R^2. Returns are simulated according to Equations (16)-(19) with $s = 1$.

Figure 4:

A single weak factor with a cross-sectional R^2 of 0.5

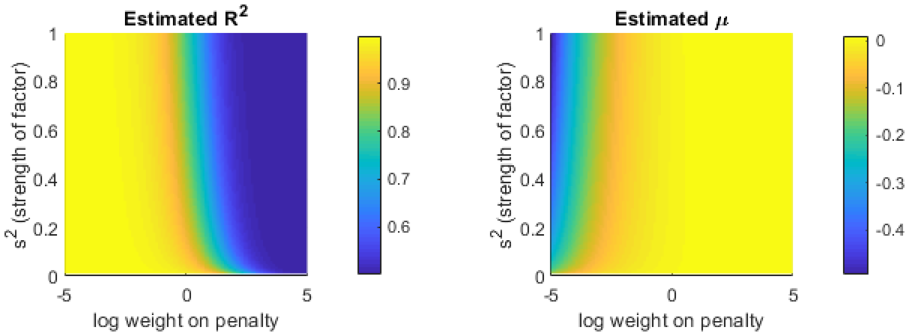


We apply a GMM estimation with the moment conditions in Equation (5) and the weighting matrix in Equation (6). The figure shows estimated R^2 and RMSE and the point estimates of λ and μ, together with 95% confidence bounds as functions of x, the log weight on the moment condition that identifies the factor mean. Returns are simulated according to Equations (16)-(19) with $r = \sqrt{0.5}$ and $s = \sqrt{0.025}$.

pattern in estimated R^2's and also the parameter estimates as functions of x are rather stable across levels of s. This shows that the strength of a factor has only a minor impact on the bias we describe, relative to the true explanatory power of a factor, as analyzed in Section 3.2.2. Stated differently and as also pointed out in the introduction, the issue we raise does not apply to weak factors only.

Figure 5:

A single strong factor with a cross-sectional R^2 of 0.5



We apply a GMM estimation with the moment conditions in Equation (5) and the weighting matrix in Equation (6). The figure shows estimated R^2 and the point estimates of μ as functions of x, the log weight on the moment condition that identifies the factor mean, and the true R^2. Returns are simulated according to Equations (16)-(19) with $s = 1$.

4.4 Models with two factors

We continue with two factors which differ in terms of both cross-sectional and time series explanatory power. We start 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 $s_2 = \sqrt{0.01}$). This resembles the setup studied in Section 3.3. The true cross-sectional R^2 is equal to $0.5 = \frac{1}{2}(r_1^2 + r_2^2)$. Figure 6 shows the usual statistics.

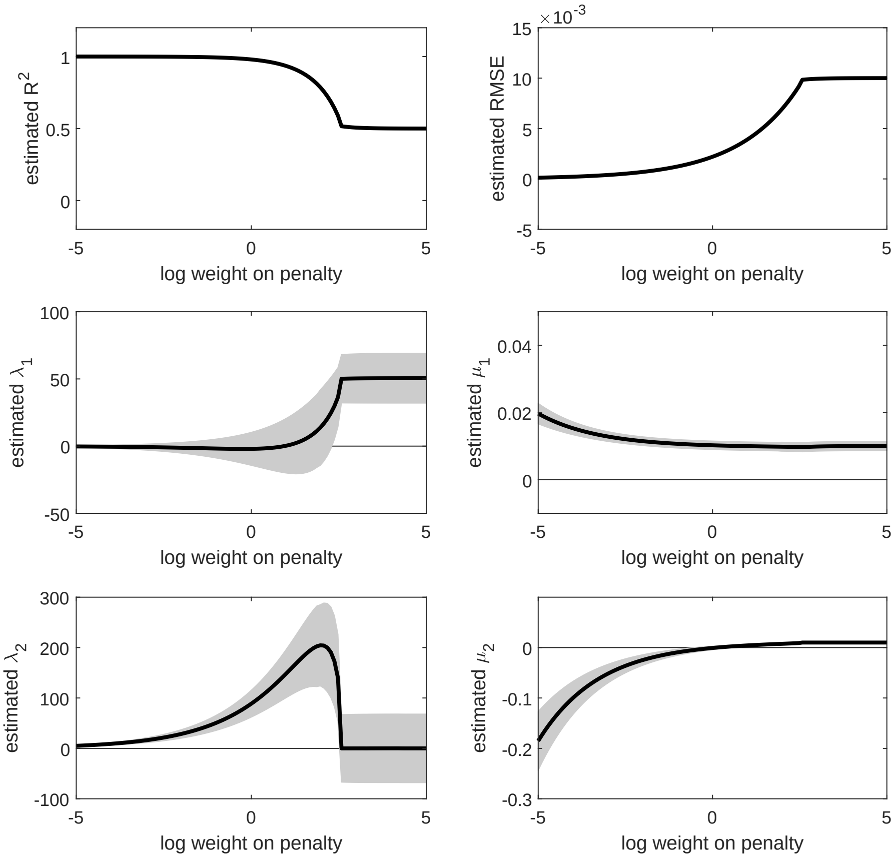
As before, the estimated R^2 and RMSE go to the true values as x increases. For log weights x above a certain threshold, in this case $x = 2.6$, deviations of $\hat{\mu}$ from the sample means of the factors are so costly in terms of the penalty terms that the minimum of the GMM objective function is equal to the true parameters. For values of x below that threshold, it is cheaper to reduce the pricing errors at the cost of a penalty. This can be achieved by setting λ close to $(E_T[F] - \mu)^{-1}$ for one of the factors, i.e., bringing the constant term in Equation (8) close to one. As discussed in Section 3.3, the pricing error for the strong and imperfectly priced factor is equal to $\lambda Cov^{\perp}$ and can only be reduced by setting λ close to zero since $Cov^{\perp}$ is large. A small λ, however, results in a high penalty. For the weak factor, $Cov^{\perp}$ is already close to zero, so that there is no need to choose a small λ. We find exactly this pattern in the estimated λ_2 in Figure 6. For weights below the critical value, λ_2 estimates are positive and significant and μ_2 estimates are close to $E_T[F_2] - \frac{1}{\lambda_2}$. The pattern in $(\hat{\lambda}_2, \hat{\mu}_2)'$ as a function of x is qualitatively similar to the case of a single strong and perfectly unpriced factor, shown in Figure 1. The difference is that the critical log weight x is larger in Figure 6, since the weak and unpriced factor F_2 here is much weaker than the unpriced factor in Figure 1.

For low values of x, the parameters $(\hat{\lambda}_2, \hat{\mu}_2)'$ correspond to a trivial solution, applied to the weak factor, which brings the pricing errors close to zero. The strong and (imperfectly) priced factor F_1 hampers the good pricing performance of the model. To shut down its impact on the pricing errors, λ_1 is estimated close to zero for low values of x. As a consequence, the weak and unpriced factor F_2 appears priced while the strong and unpriced factor appears irrelevant.

In numbers, when weighting the moment conditions with the identity matrix, the market price of risk 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 market price of risk λ_2 of the weak 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

Figure 6:

One strong and priced and one weak and unpriced factor



We apply a GMM estimation with the moment conditions in Equation (5) and the weighting matrix in Equation (6). 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 (16)-(19) with $r_1 = 1$, $r_2 = 0$, $s_1 = \sqrt{0.9}$, and $s_2 = \sqrt{0.01}$.

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%).

To complete the analysis, we finally turn to the case of two equally strong factors, in the sense that

$$(Cov_T(R^e, F_1))' Cov_T(R^e, F_1) = (Cov_T(R^e, F_2))' Cov_T(R^e, F_2) \neq 0,$$

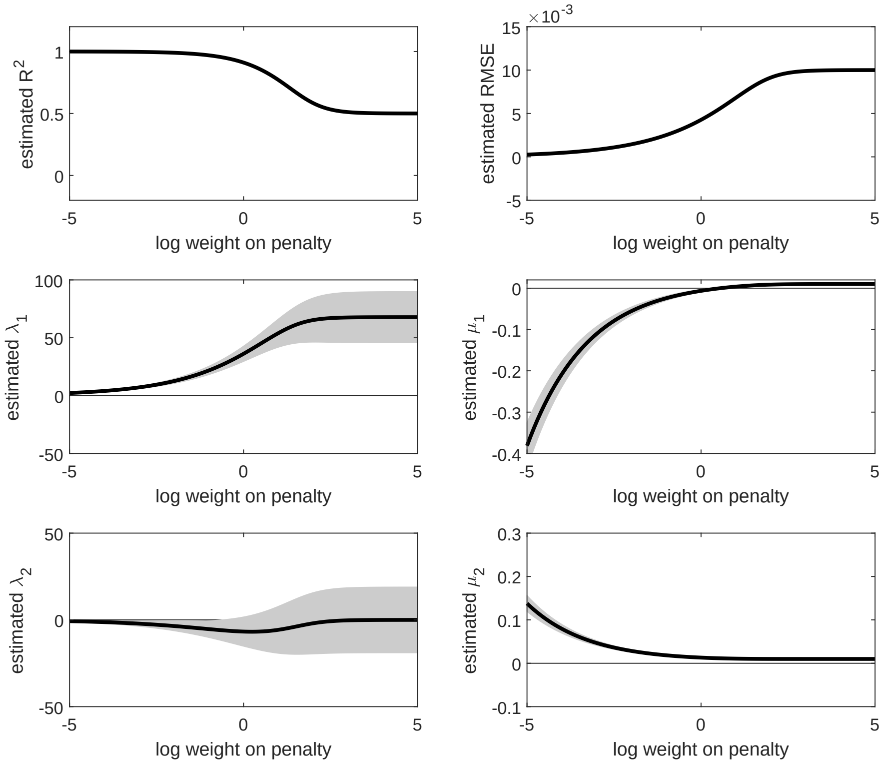
where factor F_1 is priced and factor F_2 is perfectly unpriced. In terms of the parameters in Equations (16) to (19), we set $s_1 = s_2 = \sqrt{0.5}$, $r_1 = 1$, and $r_2 = 0$, again implying a true cross-sectional R^2 of 0.5 for the two factor model. Figure 7 shows estimated R^2 and RMSE, together with the point estimates and 95% confidence intervals of the parameters.

We again find a critical value of x (now at 1.9), which separates the desired estimates (for x above that value) from biased estimates (for x below that value). Compared to Figure 6, the critical value is now slightly smaller, since the two factors are much stronger (compare the analysis in Section 3.2.3).

Interestingly, in terms of the point estimates of the parameters, the two factors “share the work”

Figure 7:

One strong and priced and one strong and perfectly unpriced factor



We apply a GMM estimation with the moment conditions in Equation (5) and the weighting matrix in Equation (6). 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 (16)-(19) with $r_1 = 1$, $r_2 = 0$, $s_1 = \sqrt{0.5}$, and $s_2 = \sqrt{0.5}$.

for low log weights x. Since the two factors are equally strong, it is equally costly in terms of the penalty term to let the μ estimates deviate from the sample means of the factors to decrease the pricing errors. Compared to the analysis of Figure 6, which showed that weak factors drive out strong factors, we thus cannot conclude that unpriced factors drive out priced factors if they are comparable in terms of strength. Still, however, parameter estimates are biased and R^2's are inflated when the weight on the penalty term is too low.

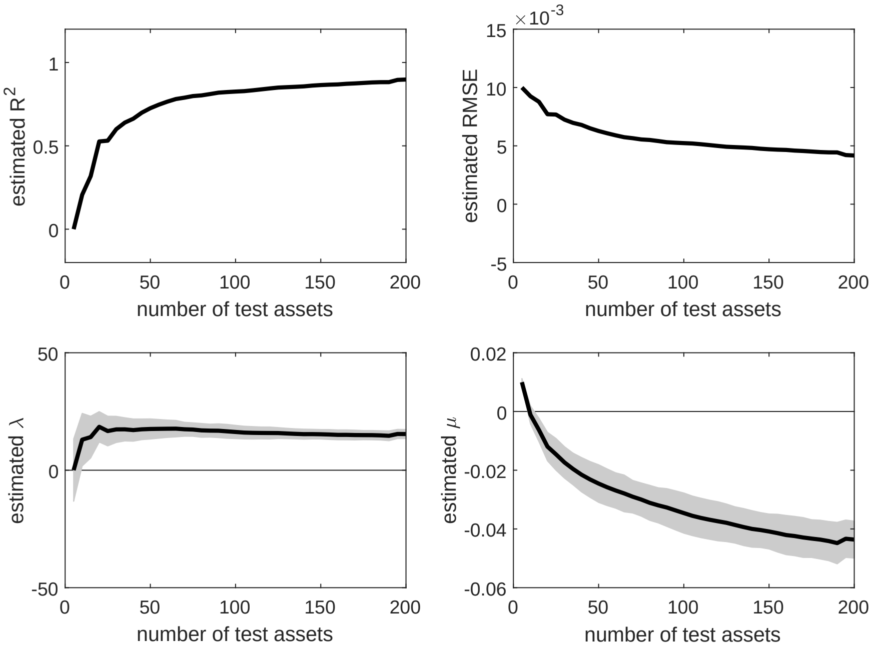
4.5 The impact of sample size and number of test assets

None of our arguments involve the sample size. Indeed, shrinking the pricing error at the cost of a higher penalty is not more costly when the sample size is particularly small or large. To corroborate this intuition, we simulate data with different sample sizes between 100 and 1.000.000 and repeat the estimation. We find that the parameter estimates and pricing statistics are exactly equal across sample sizes. This stability is due to the fact that we orthogonalize and standardize all time series in order to set all sample moments equal to their true moments even in small samples.

We also analyze if and how the number of test assets has an impact on the estimated parameters and the model performance statistics. We simulate factors and returns with $s = 1$ and $r = 0$, corresponding to a strong and perfectly unpriced factor, as in Section 3.2.1, but the following intuition carries over to cases of weaker and priced factors. We fix $x = 0$, which means that we use the identity matrix for weighting the moment conditions, and we simulate between 5 and 200 test asset return time series. Figure 8 shows estimated cross-sectional R^2's, root mean squared pricing errors, and the point estimates of λ and μ, together with the 95% confidence intervals, as functions of the number of test assets.

Figure 8:

The impact of the number of test assets



We apply a GMM estimation with the moment conditions in Equation (5) and the identity weighting matrix ($x = 0$). The figure shows estimated R^2 and RMSE and the point estimates of λ and μ, together with 95% confidence bounds as functions of the number of test assets. Returns are simulated according to Equations (16)-(19) with $r = 0$ and $s = 1$.

Interestingly, we find that the bias in parameter estimates is more severe if the number of test assets is large. With only five test assets, the estimated R^2 coincides with the true R^2 of zero, and the estimates of λ and μ are equal to the true values of 0 and 0.01. With an increasing number of test assets the estimated R^2 goes to 1 and the estimated μ goes to pronounced negative values.

Intuitively, when increasing the number of test assets, we also increase the number of moment conditions. However, the number of moment conditions that identify the factor mean stays unaltered and, thus, they become less relevant relative to the pricing errors. In that sense, increasing the number of test assets indirectly increases the weight on the pricing errors or, in other words, decreases the relative weight on the penalty terms.

We conclude that the bias in parameter estimates does not depend on the sample size, but the number of test assets matters. Ceteris paribus, the bias is more severe if the number of test assets is large. This is in contrast to, e.g. Lewellen et al. (2010), who suggest increasing the number of test assets as an easy way to address the challenges laid out in their paper. However, for the particular problem discussed in our paper, it makes the situation even worse.

5 Alternative Weighting Matrices

The analysis in Sections 3 and 4 emphasizes the key role of the GMM weighting matrix for the bias in the estimation. For ease of exposition, we focused on prespecified weighting matrices of the form

$$W = \begin{pmatrix} I_n & 0 \\ 0 & 10^x I_k \end{pmatrix} = \operatorname{diag}(1, \ldots, 1, 10^x, \ldots, 10^x). \tag{21}$$

In applications, however, weighting matrices are often chosen endogenously, for instance by inverting a prior estimate of the covariance matrix of error terms. Therefore we now analyze the behavior of the GMM estimator for alternative endogenous specifications of W. We start with the theoretically optimal weighting matrix, namely the inverse of the true covariance matrix of the errors. Of course, we typically do not know the true variances and covariances of the moment conditions in applications. Therefore, we also discuss two popular alternatives: A diagonal matrix with the inverses of the variances of the moment conditions on the main diagonal, and the inverse of an estimate of the covariance matrix.

5.1 Inverse of the true covariance matrix

In general, the idea of using the inverse of the covariance matrix goes back to Hansen (1982), who showed that the corresponding GMM estimator is efficient in the class of asymptotically normal estimators. In applications, this approach requires a prior estimation of the variances of the moment conditions, since the true parameters, and, thus, the true variances are unknown. In our controlled simulation environment, however, we know the true parameters, in particular the variances of the moment conditions. The statistics for pricing performance, such as R^2 and RMSE, are often not interpretable with this specification because the GMM objective function no longer minimizes the pricing errors of the test assets, but the pricing errors of portfolios of them. However, we can still investigate to what extent the bias in parameter estimates is mitigated by this choice of the weighting matrix.

We start with the case of a strong and perfectly unpriced factor. We decompose the inverse of the true covariance matrix of the moment conditions as

$$W = \begin{pmatrix} W_{pe} & * \\ * & * \end{pmatrix},$$

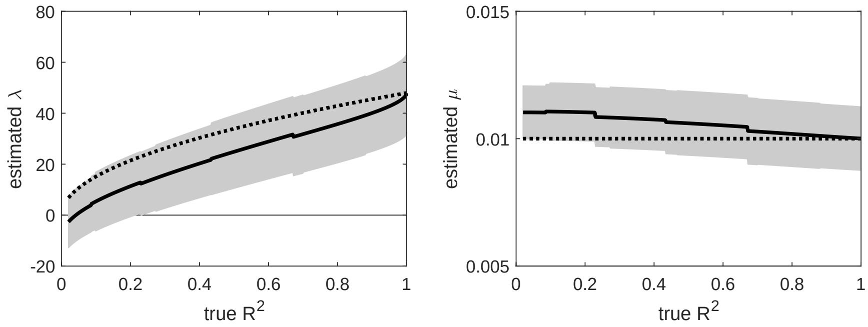
where W_{pe} is the weighting matrix for the pricing errors. Given the structure of our data-generating process, all off-diagonal elements of W_{pe} are equal to zero. The weights on the main diagonal are positively related to the respective test assets' factor exposures in the clean environment of our simulation study. The remaining parts of W, which are marked by asterisks, are populated by non-zero elements. Intuitively, the first n moment conditions are correlated through their dependence on the factor F. Efficient GMM gets rid of this joint variation by considering differences between these moment conditions and $F - \mu$.

To investigate whether the inverse of the true covariance matrix leads to correct parameter estimates, we simulate a strong factor ($s = 1$) with varying true cross-sectional R^2's. Figure 9 shows the true parameters λ and μ (dotted lines), the parameter estimates (solid lines), and the 95% confidence intervals, as functions of the true cross-sectional R^2 of the simulated factor. All other parameters in the simulation are as described in Section 4.

Overall, the figure shows that the point estimates are close to their true values. In additional simulations not reported here, we vary the variances of returns and of the factor, the number of test assets, and the strength of the factor, and we find this result to be very robust across

Figure 9:

Using the inverse of the true covariance matrix for weighting



We apply a GMM estimation with the moment conditions in Equation (5) and the use the inverse of the true covariance matrix of the moment conditions as weighting matrix. The figure shows the true λ's and μ's and point estimates of λ and μ, together with 95% confidence bounds as functions of the true R^2 of the factor model. Returns are simulated according to Equations (16)-(19) with $s = 1$.

specifications. Our findings suggest that using the inverse of the true covariance matrix mitigates the bias we describe in this paper. Unfortunately, in applications, the true covariance matrix is typically unknown and needs to be estimated. In the following, we show that two popular alternatives lead to biases that are comparable to those described in Section 4.

5.2 Diagonal weighting matrices

In asset pricing applications, a weighting matrix with off-diagonal elements implies that the estimated parameters only minimize the pricing errors of portfolios of the test assets, but not of the test assets themselves. Such portfolios often have extreme weights if there is a common component in test asset returns that is not captured by the factors (a *time effect*, see Petersen, 2009). Cochrane (2005) argues that a prespecified weighting matrix with zero off-diagonal elements avoids this problem and allows for a clear interpretation of the estimated parameters and a simple evaluation of the model's pricing performance. Instead of assigning the different moment conditions a unit or some other arbitrary weight, Yogo (2006) suggests to use the inverses of the variances of the moment conditions as weights to account for the fact that less volatile moments are more informative about the parameters to be estimated (see Section 6).

In applications, this reduced-form approach also requires a prior estimation of the variances of the moment conditions, so we again study it in our controlled simulation environment where we know the true parameters. Again, we consider the situation of a strong and perfectly unpriced factor first. In this case, the true λ is equal to zero, such that the variances of the individual moment conditions in

$$g(\lambda, \mu) = \begin{bmatrix} R_i^e - R_i^e(F-\mu)\lambda, \quad i = 1, \ldots, n \\ F - \mu \end{bmatrix} \tag{22}$$

are simply given by the variances of the test asset returns and of the factor. The corresponding diagonal weighting matrix is then simply given by $\text{diag}(Var(R_1^e)^{-1}, \ldots, Var(R_n^e)^{-1}, Var(F)^{-1})$.

In asset pricing, it is common to use factor mimicking portfolios which are typically about as volatile as the test asset returns themselves. But if the pricing factor F is as volatile as the test asset returns, we are back in the case of an identity matrix as weighting matrix with all the consequences described in Section 4.

If the pricing factor and the test asset returns have very different volatilities, we can simply scale the pricing factor up or down. The true pricing performance of linear factor models is invariant to affine linear transformations of the factors. Our theoretical analysis, however, reveals that the estimated pricing performance is not invariant to scaling. When using a diagonal matrix with inverse variances of the moment conditions on the main diagonale, multiplying a pricing factor by a scalar m is tantamount to multiplying the weight on the penalty term by $1/m^2$.

For priced factors, this rationale does not change much. When the true λ is different from zero, the variances of the pricing errors typically increase moderately, relative to the case where λ is equal to zero. However, the intuition from above still applies. The weight on the penalty term essentially reflects the magnitude of the factor variance relative to the return variances of the test assets.

5.3 Iterated and continuously updating GMM

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

We again simulate a strong ($s = 1$) and perfectly unpriced ($r = 0$) factor 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.

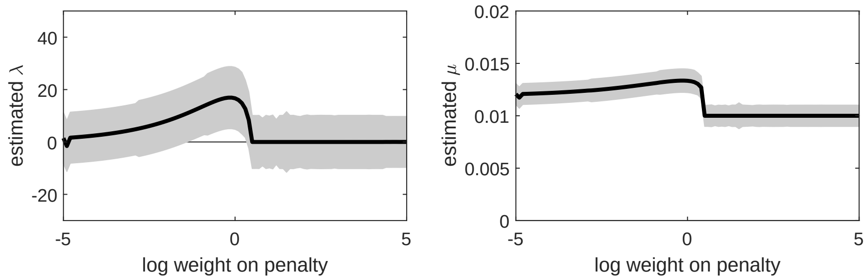
Figure 10 shows the point estimates of λ and μ, together with 95% confidence intervals, as functions of the log weight x on the penalty term in the first stage GMM estimation. If the iteration always converged to the true parameters, the point estimates should be independent of the weight x in the first stage.

The figure clearly shows that this is not the case. In fact, the pattern in point estimates is remarkably similar to the one observed in Figure 1. Note that the results in Figure 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.

Our results suggest that biased first stage estimates carry over to higher stages in the iteration. This is not surprising. Suppose the log weight x in the first stage is very low. As we have argued in Sections 3 and 4, we then end up with parameter estimates that are close to a trivial 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 in Equation (22), i.e., the pricing errors, approximately simplify to $R - R$. This means that not only the sample average of these moment conditions, but in fact each single observation is virtually equal to zero. Thus, the variances of the first n moment conditions and their covariances with the penalty terms are close to zero. The covariance matrix

Figure 10:

Using the inverse of the estimated covariance matrix for weighting



We apply an iterated GMM estimation with the moment conditions in Equation (5) and use the inverse of the covariance matrix of the moment conditions, estimated using the point estimates from the previous stage, as 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 penalty term in the first stage. Returns are simulated according to Equations (16)-(19) with $s = 1$ and $r = 0$.

of the moment conditions after the first stage approximately has the form

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

and is (close to) singular. Intuitively, small but finite log weights x in the first stage lead to parameters that we labeled as a trivial solution, so that the weights on the pricing errors in the second stage will be very high. The estimated variance of the pricing errors is close to zero which makes these moments seem much more informative about the parameters to be estimated than the penalty term. This again leads to biased parameter estimates, and, consequently, biased estimates of the covariance matrix, in the subsequent stages.[12]

As an alternative to iterated GMM, we also investigate continuously updating GMM (CUGMM), suggested 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.

Summing up, popular numerical strategies, such as iterated or CUGMM, do not mitigate the biases in the parameter estimates that we describe in Section 3.

[12] We find the same pattern for the point estimates also with much larger sample sizes, suggesting that the bias does not vanish as $T \to \infty$. The estimate of $Cov(g)$ does not depend on T, implying that the weights do not change as T increases. Thus, the intuition from Section 4.5 carries over to the case of iterated GMM.

6 Example: The Durable Consumption Model

We finally turn to an example from the literature in which the estimation design outlined above plays a major role, and which is therefore very susceptible to the bias outlined in our paper. Yogo (2006) proposes a factor model in which expected excess returns of stocks or portfolios are linear in the assets' exposures (betas) with respect to three factors: the log growth rate of consumption of nondurable goods and services (denoted by F_1 subsequently), with a sample mean of 0.513%, the log growth rate of consumption of durable goods (F_2), with a mean of 0.915%, and the log return on the aggregate stock market (F_3), with a mean of 1.88%. The model is tested using quarterly excess returns of the standard 25 size- and book-to-market-sorted portfolios from 1951:Q1 to 2001:Q4, which are available on Kenneth French's webpage.[13] The sample moment conditions used to estimate the model parameters and evaluate the goodness of fit via GMM are the ones given in Equation (5). In total, there are 28 moment conditions and 6 parameters to be estimated.

The exact estimation procedure for the model is somewhat involved, so that a detailed description is delegated to Appendix A. In short, a two-stage GMM estimation is performed, where the weighting matrix in the second stage is the inverse of the covariance matrix estimated in the first stage. Most important in the context of our paper is the weighting matrix $W^{(1)}$ in the first stage. This matrix is obtained using initial values for the parameters λ and μ (see Appendix A). More specifically, $W^{(1)}$ is chosen as

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

where $\widehat{\Omega}^{(1)}$ is the estimate of the 28×28-covariance matrix of the moment conditions, given these initial values. $\widehat{\Omega}^{(1)}_{i,\ldots,j}$ denotes its submatrix

$$\begin{pmatrix} \widehat{\Omega}^{(1)}_{i,i} & \cdots & \widehat{\Omega}^{(1)}_{i,j} \\ \vdots & \ddots & \vdots \\ \widehat{\Omega}^{(1)}_{j,i} & \cdots & \widehat{\Omega}^{(1)}_{j,j} \end{pmatrix}$$

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 $(\widehat{\Omega}^{(1)}_{26,\ldots,28})^{-1}$ in $W^{(1)}$ by a factor 10^x, varying x between -4 and 4.

The results are shown in Table 2. Confirming our theoretical and simulation-based results, 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 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 3, 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.

As a final step, we also present point estimates and the cross-sectional R^2 from a standard Fama-MacBeth two-pass regression in the last column of Table 2. The R^2 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

[13]See `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

Table 2: **Parameter estimates for different weighting matrices**

	$x=-4$	$x=0$	$x=1$	$x=2$	$x=4$	FMB
First stage results						
λ_1	-5.448	15.537	73.012	313.970	335.626	278.738
	[26.026]	[113.722]	[331.228]	[191.854]	[235.004]	[76.001]
λ_2	17.783	152.409	237.669	-307.570	-181.367	-103.759
	[43.913]	[39.411]	[86.587]	[240.471]	[272.894]	[64.467]
λ_3	0.229	1.000	1.000	-6.398	-5.002	3.129
	[1.354]	[2.702]	[6.651]	[5.966]	[6.626]	[0.964]
μ_1	0.254	0.480	0.495	0.516	0.513	0.513
	[0.004]	[0.001]	[0.000]	[0.001]	[0.001]	[0.051]
μ_2	-4.783	0.299	0.593	1.007	0.916	0.915
	[0.157]	[0.002]	[0.003]	[0.001]	[0.001]	[0.074]
μ_3	-0.741	1.595	1.730	1.921	1.880	1.880
	[0.038]	[0.006]	[0.006]	[0.006]	[0.006]	[0.549]
MAE	0.012	0.122	0.209	0.337	0.415	0.416
R^2	0.999	0.935	0.767	0.409	0.013	0.009
Second stage results						
λ_1	-5.292	17.901	90.914	-34.333	139.981	
	[3.266]	[31.280]	[58.334]	[39.597]	[43.216]	
λ_2	19.196	170.569	267.807	34.729	-137.893	
	[1.787]	[15.561]	[34.157]	[43.131]	[53.529]	
λ_3	0.198	0.659	1.000	0.998	-2.077	
	[0.078]	[0.849]	[1.642]	[1.061]	[1.313]	
μ_1	0.395	0.533	0.491	0.506	0.556	
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
μ_2	-4.291	0.278	0.575	0.768	0.937	
	[0.005]	[0.001]	[0.001]	[0.001]	[0.001]	
μ_3	0.861	1.637	1.848	1.454	1.592	
	[0.004]	[0.005]	[0.005]	[0.004]	[0.005]	
J	7.838	23.170	32.769	60.055	29.108	
p-val	0.998	0.392	0.065	0.000	0.142	

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. HAC standard errors are in brackets. Details regarding the estimation of the model are presented in Appendix A.

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.

The wide range of estimates in Table 2, in particular in comparison with the Fama-MacBeth results, suggests that the estimates with $x = 4$ may be regarded as close to the true parameters, and those with $x = 0$ should be deemed wrong. However, we wish to emphasize that this conclusion cannot be drawn with this generality. Contrary to the controlled simulation exercise in Section 4, 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. the factor is in fact much stronger 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 (also about the factor means) than the factor time series.

On the other hand, our extensive analysis in Sections 3, 4, and 5 has shown 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 theoretical analysis predicts exactly the pattern that we find in Table 2, and this pattern strongly suggests that the durable goods-model does a bad job in explaining the cross-section of expected returns.

7 Summary and Discussion

Standard GMM cross-sectional asset pricing tests can generate spuriously high explanatory power for linear factor models and biased parameter estimates for the market prices of factor risks. Tests based on simulated and empirical 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 the problem is the more severe the larger the number of test assets.

We identify two major problems. First, unpriced factors look priced. For instance, the market factor and even the population growth of captive Asian elephants seemingly explain the cross-section of expected stock returns close to perfectly, as long as a small weight is put on matching the mean of these factors. The market prices of factor risk estimates are highly significant, although their true value is close to zero, a typical type I error. Importantly, the problem applies to strong and weak factors.

Second, strong factors with some explanatory power for the cross-section of expected returns can be "crowded out" by weak and unpriced factors and become seemingly insignificant, a typical type II error. For instance, in the introductory example, the elephant factor renders the three Fama-French factors obsolete in a joint test.

In GMM estimations it is natural to assign more informative moments a higher weight. In the light of this argument, one may conclude that a serious misestimation of the factor means in favor of an improved cross-sectional fit is a natural result of the trade-off between the uncertainty about the true factor mean and the pricing errors. However, as we show in simulations, standard techniques to choose an "optimal" weighting matrix result in parameter estimates that show a bias that is similar to the one we thoroughly describe for prespecified weighting matrices.

How can asset pricing researchers diagnose if their results are biased and the pricing performance of their model is spuriously high? As mentioned by Parker and Julliard (2005), one can perform several GMM estimations with varying weights on the moment condition that identifies 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 in the desired solution (which means that the parameter estimates are unbiased) or in 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 $1/\lambda$.

Alternatively, we suggest a sanity check of the factor mean estimates. Typically, these estimates will be far away from the sample averages of the factors if the pricing performance of the factors is inflated. As described in detail in Section 3.2.3, weak factors are the exception. As a conclusion, researchers who apply the GMM estimation approach described here should run established tests to see if their factors are weak (see Kleibergen and Paap, 2006; Kroencke, 2020) and, in addition, always report the estimated factor means, together with estimated standard errors.

Finally, a simple 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 market prices of risk should be compared to Fama and MacBeth estimates as well. In case of a traded factor, the λ estimate should not only be tested for being significantly different from zero. Instead, researchers should also check if it is close to the sample average of the factor divided by the factor variance.

References

Boguth, O. and L. Kuehn (2013). Consumption volatility risk. *Journal of Finance 68*, 2589–2615.

Borri, N. and G. Ragusa (2017). Sensitivity, moment conditions and the risk-free rate in Yogo (2006). *Critical Finance Review 6*(2), 381–393.

Bryzgalova, S. (2016). Spurious factors in linear asset pricing models. *Working Paper*.

Burnside, C. (2016). Identification and inference in linear stochastic discount factor models with excess returns. *Journal of Financial Econometrics 14*(2), 295–330.

Campbell, J. (2017). *Financial Decisions and Markets: A Course in Asset Pricing.* Princeton University Press.

Chen, Z. and A. Lu (2017). Seeing the unobservable from the invisible: The role of co2 in measuring consumption risk. *Review of Finance 22*(3), 977–1009.

Cochrane, J. (2001). A rehabilitation of stochastic discount factor methodology. *Working Paper*.

Cochrane, J. (2005). *Asset Pricing.* New Jersey: Princeton University Press.

Da, Z., W. Yang, and H. Yun (2016). Household production and asset prices. *Management Science 62*(2), 387–409.

Darrat, A., B. Li, and J. Park (2011). Consumption-based CAPM models: International evidence. *Journal of Banking and Finance 35*, 2148–2157.

Den Haan, W. and A. Levin (2000). Robust covariance matrix estimation with data-dependent var prewhitening order. NBER Working Paper T0255.

Dhume, D. (2010). Using durable consumption risk to explain commodities returns. *Working Paper*.

Fama, E. and K. French (1992). The cross-section of expected stock returns. *Journal of Finance 47*, 427–465.

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

Fama, E. and J. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607–636.

Ferson, W. (2019). *Empirical Asset Pricing: Models and Methods*. MIT Press.

Ferson, W., S. Nallareddy, and B. Xie (2013). The "out-of-sample" performance of long-run risk models. *Journal of Financial Economics 107*(3), 537–556.

Gospodinov, P., R. Kan, and C. Robotti (2014). Misspecification-robust inference in linear asset pricing models with irrelevant risk factors. *Review of Financial Studies 27*(7), 2139–2170.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*, 1029–1084.

Hansen, L., J. Heaton, and A. Yaron (1996). Finite sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics 14*, 262–280.

Harvey, C., Y. Liu, and H. Zhou (2016). ... and the cross-section of expected returns. *Review of Financial Studies 29*(1), 5–68.

Jagannathan, R. and Z. Wang (2002). Empirical evaluation of asset pricing models: A comparison of SDF and beta methods. *Journal of Finance 57*(5), 2337–2367.

Kan, R. and C. Zhang (1999a). Gmm tests of stochastic discount factor models with useless factors. *Journal of Financial Economics 54*(1), 103–127.

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

Kan, R. and G. Zhou (1999). A critique of the stochastic discount factor methodology. *Journal of Finance 54*(4), 1221–1248.

Kleibergen, F. (2009). Tests of risk premia in linear factor models. *Journal of Econometrics 149*(1), 149–173.

Kleibergen, F. and R. Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics 133*(1), 97–126.

Kleibergen, F. and Z. Zhan (2015). Unexplained factors and their effects on second-pass r-squared's. *Journal of Econometrics 189*(1), 101–116.

Kleibergen, F. and Z. Zhan (2020). Robust inference for consumption-based asset pricing. *forthcoming in Journal of Finance*.

Kroencke, T. (2017). Asset pricing without garbage. *Journal of Finance 72*(1), 47–98.

Kroencke, T. (2020). On robust inference for consumption-based asset pricing. Working Paper.

Lewellen, J., S. Nagel, and J. Shanken (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics 96*, 175–194.

Lioui, A. and P. Maio (2014). Interest rate risk and the cross-section of stock returns. *Journal of Financial and Quantitative Analysis 49*(2), 483–511.

Maio, P. (2013). Intertemporal CAPM with conditioning variables. *Management Science 59*(1), 122–141.

Maio, P. and P. Santa-Clara (2012). Multifactor models and their consistency with the icapm. *Journal of Financial Economics 106*(3), 586–613.

Newey, W. and K. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Newey, W. and F. Windmeijer (1994). Generalized method of moments with many weak moment conditions. *Econometrica 77*, 687–719.

Parker, J. and C. Julliard (2005). Consumption risk and the cross section of expected returns. *Journal of Political Economy 113*, 185–222.

Petersen, M. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies 22*(1), 435–480.

Ronchetti, E. and F. Trojani (2001). Robust inference with gmm estimators. *Journal of Econometrics 101*(1), 37–69.

Savov, A. (2011). Asset pricing with garbage. *Journal of Finance 66*, 177–201.

Stock, J. and J. Wright (2000). GMM with weak identification. *Econometrica 68*, 1055–1096.

Yogo, M. (2006). A consumption-based explanation of expected stock returns. *Journal of Finance 61*, 539–580.

Appendix A The Durable Consumption Model: Details

A.1 Estimation algorithm

In this appendix we provide details on the analysis of the durable consumption model discussed above in Section 6. The moment conditions are given in Equation (5), and the parameters to be estimated are the three market prices of risk (scaled by the factor variance) $\lambda = (\lambda_1, \lambda_2, \lambda_3)'$ as well as the factor means $\mu = (\mu_1, \mu_2, \mu_3)'$.

Our analysis proceeds in two steps. We first present an exact replication of Yogo's results together with some modifications of the numerical procedure that already have a pronounced effect on the GMM weighting matrix (Sections A.2 and A.3). Then we perform an analysis similar to the one in Section 4, i.e., we explicitly manipulate the GMM weighting matrix, to study the impact of this variation on the estimation results.

We start by replicating the results presented in Table 3 in Yogo (2006). The original code is written in Gauss and available on Motohiro Yogo's website.[14] Our replication code is a line-by-line translation to Matlab.

In the following, we describe the exact GMM algorithm used by Yogo (2006) to estimate the six parameters λ and μ.

1. **Initial parameter values.** The initial value for μ is set to the sample average of F and the initial value for λ is the solution of the system of linear equations $E_T[R^e] = [(R^e)'(F-\mu)/T]\ \lambda$. Here, T denotes the sample size, F denotes the $T \times 3$ matrix containing the time series of the factors, and R^e denotes a $T \times 24$ matrix that contains the time series of excess returns of portfolios 2 to 25. In particular, the small growth portfolio is dropped when the initial value for λ is calculated.

2. **Covariance matrix of moment conditions.** The initial values calculated in Step 1 are plugged into the moment condition function (Equation (5)). This allows the estimation of the covariance matrix $\widehat{\Omega}^1$ of the moment conditions. For this purpose, a parametric estimator along the lines of Den Haan and Levin (2000) is used.

3. **GMM: first stage.** The moment conditions are weighted by a sparse weighting matrix of the form

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

where $\widehat{\Omega}^{(1)}_{i,\ldots,j}$ denotes the submatrix $\begin{pmatrix} \widehat{\Omega}^{(1)}_{i,i} & \ldots & \widehat{\Omega}^{(1)}_{i,j} \\ \vdots & & \vdots \\ \widehat{\Omega}^{(1)}_{j,i} & \ldots & \widehat{\Omega}^{(1)}_{j,j} \end{pmatrix}$ of $\widehat{\Omega}^{(1)}$ and I_{25} denotes the 25-dimensional identity matrix. The initial values for the optimizer are the same as in Step 1. The point estimates are constrained in the following way:

$$\begin{aligned} \lambda_3 &< 1 \\ \lambda_1 + \lambda_2 + \lambda_3 &> 0. \end{aligned} \tag{23}$$

In the theoretical model presented in Yogo (2006), these inequalities mean that the implied relative risk aversion and the implied elasticity of intertemporal substitution parameters are constrained to be positive.

[14] See `https://sites.google.com/site/motohiroyogo/`.

4. **Cross-sectional $\mathbf{R}^2$ and pricing errors.** These are calculated based on the estimates from Step 3.

5. **Covariance matrix of moment conditions.** The parameter estimates from Step 3 are plugged into the moment condition function (5) to obtain a second estimate $\widehat{\Omega}^{(2)}$ of the covariance matrix of the moment conditions.

6. **GMM: second stage.** $(\widehat{\Omega}^{(2)})^{-1}$ is used to weight the moment conditions. The point estimates from Step 3 are used as initial values for the optimization routine.

Table 3 in Yogo (2006) shows the estimates of λ and the J-statistic from the second stage (Step 6) and the mean absolute pricing error and the R^2 from the first stage (Step 4). The estimates of μ_F are not reported.

Steps 2 and 5 involve the estimation of covariance matrices of autocorrelated time series. Yogo (2006) uses a parametric estimation approach of the spectral density matrix as described by Den Haan and Levin (2000). An alternative is to use a non-parametric estimator in the spirit of Newey and West (1987).[15] In our application, the first 25 moment conditions (the pricing errors) and the factor mean of the market portfolio return are barely affected by the choice of the covariance estimator. On the other hand, the parametric estimates for the variance of nondurable (durable) consumption growth are 1.8 (14.4) times higher than the estimates that we obtain using the nonparametric estimation procedure.

Tables 3 and 4 show the results from the estimation as reported in Table 3 in Yogo (2006), together with the results from eight variations of the estimation procedure. The first one is our one-to-one replication of Yogo (2006) (labeled "Replica"). Then, we keep the small growth portfolio when calculating the initial covariance matrix for the first stage in Step 1 ("25 portf"). For the next two replications, we use the nonparametric instead of the parametric covariance matrix estimator, once without and once with including the small growth portfolio in Step 1 ("nonpara" and "25 portf & nonpara", respectively). These four variations of the estimation are performed with and without imposing the constraints on the parameters in Steps 3 and 6.

A.2 GMM: First stage

We are going to discuss the results from the first stage of the GMM (Table 3) first. Numbers in italics are not reported in the paper but only in the text file `est_dur` available in the supplementary material provided on Motohiro Yogo's website.

A comparison of the first two columns shows that we perfectly replicate the results reported in the paper when we use the original setup. The picture changes drastically, however, when we consider the variations discussed above. The most important statistics from the first stage are the cross-sectional R^2 and the mean absolute pricing error. Both statistics reveal that the superior pricing performance of the durable consumption model is already severely weakened when we use all 25 portfolios for the estimation of the covariance matrix in the first stage.

The point estimates for μ and λ are reported for informational purposes only, as the estimates in Table 3 of Yogo (2006) are obtained from the second stage of the GMM. However, we can see from our replication that the point estimates from the first stage are not robust either. Most importantly, although λ_2 remains statistically significant throughout the cases, it switches sign, which challenges the economic interpretation of the estimated coefficient. Equation (12) shows that the estimated μ can easily change from above to below $E_T[F]$ or vice versa upon a slight change in the weighting matrix, depending on the sign of the root. The associated λ estimate

[15] We rely on the function `longvar` from the GMM package of Kostas Kyriakoulis. The full package can be downloaded at `https://personalpages.manchester.ac.uk/staff/Alastair.Hall/GMMGUI.html`.

Table 3: **GMM — First stage**

		Constrained estimation				Unconstrained estimation			
	Yogo	Replica	25 portf	nonpara	25 portf & nonpara	Replica	25 portf	nonpara	25 portf & nonpara
λ_1	*15.535*	15.535	69.109	256.877	264.701	70.902	125.284	189.463	201.017
	[113.722]	[113.722]	[323.112]	[124.084]	[129.167]	[84.769]	[146.072]	[139.604]	[143.598]
λ_2	*152.410*	152.410	227.504	-251.322	-258.972	-170.725	-291.337	-320.007	-327.990
	[39.410]	[39.410]	[120.099]	[131.489]	[139.646]	[47.550]	[100.835]	[135.343]	[144.721]
λ_3	*1.000*	1.000	1.000	-5.548	-5.685	-2.375	-3.954	-5.053	-5.254
	[2.702]	[2.702]	[6.876]	[3.570]	[3.753]	[1.694]	[3.151]	[3.726]	[3.911]
μ_1	*0.5*	0.480	0.421	0.453	0.458	0.536	0.575	0.497	0.496
	[0.1]	[0.052]	[0.070]	[0.048]	[0.048]	[0.051]	[0.054]	[0.046]	[0.046]
μ_2	*0.3*	0.299	0.600	1.065	1.053	1.441	1.195	1.105	1.090
	[0.3]	[0.255]	[0.216]	[0.078]	[0.077]	[0.151]	[0.153]	[0.086]	[0.084]
μ_3	*1.6*	1.595	1.569	1.406	1.441	2.121	2.157	1.495	1.516
	[0.6]	[0.555]	[0.575]	[0.570]	[0.569]	[0.559]	[0.562]	[0.568]	[0.568]
MAE	0.122	0.122	0.198	0.249	0.260	0.120	0.212	0.247	0.258
R^2	0.935	0.935	0.784	0.683	0.657	0.941	0.811	0.724	0.698

The table presents the results from Table 3 of Yogo (2006), from our replication and from several modifications of the replication. μ_1, μ_2, μ_3, and mean absolute pricing errors (MAE) are expressed in percentage points. HAC standard errors are in brackets.

will then switch sign. Comparing the unconstrained and the constrained estimation, one can also see that the constraints are always binding once they are imposed. The estimate of $\lambda_1 + \lambda_2 + \lambda_3$ is always negative, which implies a negative risk aversion coefficient in the theoretical model.

To understand why our small changes to the estimation procedure affect the results so heavily, it is instructive to look at the effect of the algorithm design on the initial GMM weighting matrix. The small growth portfolio is the one with the largest pricing error in a simple regression-based asset pricing test. Dropping this portfolio in Step 1 leads to less volatile pricing errors for the remaining 24 portfolios, and, in turn, to a lower determinant of $\widehat{\Omega}^{(1)}_{1,\ldots,25}$. As a consequence, when we reintroduce the small growth portfolio in Step 1, the weight on the first 25 moment conditions decreases from 299.69 to 222.81. Thus, our adjusted estimation puts a lot less emphasis on having low pricing errors at the benefit of better matching the factor means.

As mentioned at the end of Section A.1, the use of the parametric covariance estimator mainly affects the weight on the 27th moment condition that identifies the durable consumption growth factor mean. While omitting the small growth portfolio in Step 1 only affects the weighting matrix in the first stage of the GMM, the choice of the covariance estimator affects the weighting matrix in both GMM stages. It allows the algorithm to estimate the durable consumption growth factor mean very imprecisely in order to decrease the pricing errors.

A.3 GMM: Second stage

Next, we analyze the robustness of the results from the second stage of the GMM estimation (Table 4). Point estimates, J-statistic and p-value reported in Yogo (2006), Table 3, are based on these second stage estimates.

First of all we find that the three-factor model is rejected by the J-test in most cases. In particular after replacing the parametric covariance estimator by the nonparametric one, the specification test rejects the model relative to all conventional significance levels. Comparing the unconstrained and the constrained estimation, the estimate of λ_3 is always greater than 1, which implies a negative intertemporal elasticity of substitution in the equilibrium model.

Comparing the point estimates from the second stage to those from the first stage, we see that these estimates are relatively close to each other in the original Yogo (2006) paper. However, adjusting the design of the GMM estimator, we find pronounced differences across all our replications. One possible conclusion could be that, after our modifications, two-stage GMM is no longer enough and we need additional stages to have more reliable estimates. We perform multi-stage GMM estimations (results not reported here for brevity) and find that the algorithm does not converge towards an efficient point estimate. Instead, in all the cases considered, the multi-stage GMM oscillates between two different point estimates.[16] We conclude that the fact that the point estimates in Yogo (2006) seem to converge after two stages of GMM is an artefact and adjusting the code destroys this property.

Finally, it is also interesting to look at the estimated factor means. The sample averages of the factors are 0.513, 0.915, and 1.880. The estimated mean growth rate of durable consumption in our replication of Yogo (2006) is only 0.278 which is very low compared to the sample average of 0.915. However, as the 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. At the same time, the model is rejected by the J-test.

[16]This pattern is described, for instance, in Cochrane (2005), p. 226.

Table 4: **GMM — Second stage**

		Constrained estimation				Unconstrained estimation			
	Yogo	Replica	25 portf	nonpara	25 portf & nonpara	Replica	25 portf	nonpara	25 portf & nonpara
λ_1	17.898	17.899	92.025	-44.590	-43.977	-60.167	-66.300	-60.905	-49.488
	[31.280]	[31.280]	[56.000]	[38.454]	[38.714]	[24.788]	[30.996]	[41.054]	[40.081]
λ_2	170.569	170.569	259.283	74.121	71.670	141.629	151.692	59.016	69.862
	[15.561]	[15.561]	[35.139]	[30.003]	[30.566]	[19.751]	[22.148]	[32.477]	[31.786]
λ_3	0.659	0.659	1.000	0.998	0.998	2.523	3.194	2.514	2.278
	[0.849]	[0.849]	[1.624]	[1.087]	[1.093]	[0.605]	[0.804]	[1.216]	[1.195]
μ_1	*0.5*	0.533	0.454	0.526	0.526	0.511	0.520	0.513	0.511
	[0.0]	[0.040]	[0.041]	[0.043]	[0.043]	[0.034]	[0.034]	[0.043]	[0.043]
μ_2	*0.3*	0.278	0.587	0.746	0.751	0.418	0.642	0.762	0.766
	[0.1]	[0.104]	[0.073]	[0.056]	[0.056]	[0.094]	[0.079]	[0.056]	[0.056]
μ_3	*1.6*	1.637	1.834	2.053	2.048	1.596	1.671	2.085	2.049
	[0.5]	[0.473]	[0.506]	[0.516]	[0.516]	[0.400]	[0.438]	[0.522]	[0.520]
J	23.170	23.170	35.515	254.120	255.947	21.376	25.200	179.853	177.939
p-val	0.392	0.392	0.034	0.000	0.000	0.498	0.288	0.000	0.000

The table presents the results from Table 3 of Yogo (2006), from our replication and from several modifications of the replication. μ_1, μ_2, and μ_3 are expressed in percentage points. HAC standard errors are in brackets.