Online Appendix
GMM Weighting Matrices in Cross-Sectional Asset Pricing Tests

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

This version: February 16, 2024

Abstract

This online appendix complements the paper "GMM Weighting Matrices in Cross-Sectional Asset Pricing Tests". The first section discusses and illustrates the bias documented in the original paper theoretically for several configurations of factors and test assets. The second section provides details about the simulation study referred to in the paper. The third section discusses papers in the literature that use the estimation approach we criticize. The fourth section provides further information regarding the estimation results of Yogo (2006).

Keywords: Asset pricing, cross-section of expected returns

JEL: G00, G12, C21, C13

[§] Lithuanian Green Finance Institute at INVEGA and Vilnius University, Sauletekio avenue 9, 10222 Vilnius, Lithuania. E-mail: nora.laurinaityte@gmail.com

[†] 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, Kaiserstrasse 12, 76131 Karlsruhe, Germany. E-mail: thimme@kit.edu, phone +49 721 681 41380

A Detailed analysis of the problem

This section is supposed to explain the rationale of the bias described in the paper more thoroughly. It 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. 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 there is cross-sectional variation in expected returns that cannot be explained by exposures to the proposed factors, i.e. that the cross-sectional R^2 is below 1 even in population.

A.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:

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

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. To set the uncertainty regarding the estimate of 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 (A.1) can be transformed as follows:

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

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

$$E[R_i^e] = E[R_i^e] \underbrace{\Big(E[F] - \mu\Big)\lambda}_{\text{“constant term”}} + \underbrace{Cov(R_i^e, F)\lambda}_{\text{“covariance term”}} \tag{A.2}$$

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 (A.2), as the only moment condition. Equation (A.2) can be solved in several ways. If the model is correctly specified, the vector $(Cov(R_i^e, F))_i$ is non-zero and proportional to the vector of expected returns $(E[R_i^e])_i$, i.e. the true cross-sectional R^2 is equal to 1. In this case, the true parameter λ^* is simply given by $\lambda^* = \frac{E[R_i^e]}{Cov(R_i^e, F)}$ for any $i = 1, \ldots, n$. Substituting $Cov(R_i^e, F) = \frac{E[R_i^e]}{\lambda^*}$ in Equation (A.2) and solving for μ gives

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

Equation (A.3) shows that there is more than one solution. Importantly, Equation (A.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 (A.3). First, one can set $\mu = E[F]$ to eliminate the “constant term” in Equation (A.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* in the paper.[1]

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 (A.2) vanishes and the "constant term" goes to one.[2] We refer to such cases as *trivial solutions* in the paper. 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 (A.2) can be made arbitrarily small. As opposed to the desired solution, a trivial solution yields parameter estimates that 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 $0 = E\left[F_j - \mu_j\right]$ for $j = 1, \dots, k$, which identify the factor means.

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

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

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

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

We denote the second set of moment conditions as *penalty terms* in this appendix, 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 (A.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

[1] 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.

[2] To see the latter, reformulate Equation (A.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.

the covariance term in Equation (A.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 $\lambda = \lambda^*$.

A.2 Misspecified models

We have shown above that, in case the model is correctly specified, the GMM estimator that minimizes the objective function (A.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 (1982). However, in applications, asset pricing models are always misspecified (see Kan and Zhang, 1999; Fama and French, 2015). As a consequence, the minimum of the GMM objective function then depends on the choice of the weighting matrix W. This is because the covariances of factors and returns do not perfectly line up with the average returns, i.e., at least one pricing error is larger than 0. In this situation, there is no parameter vector $(\lambda, \mu)'$, for which $g_T(\lambda, \mu) = 0$ and the weighting matrix determines the relative priority the estimator assigns to the moment conditions.

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 weighting matrices of the form

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

in this section. 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:

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

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.

A.2.1 Single factors

We discuss the case of single factor models first. The interaction of multiple factors is analyzed in Section A.2.2. In case of a single factor, the model is misspecified if $Cov^{\perp}$ is (asymptotically) 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 (A.2) now reads as

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

Case 1: Perfectly unpriced factors

We denote a factor for which $Cov^{ER} = 0$ a *perfectly unpriced* factor. 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, 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*. Details for the special cases of strong, priced factors and of perfectly weak factors are discussed below.[3]

We are looking for the parameters $(\lambda, \mu)'$ that minimize the GMM objective function f_x in Equation (A.4). To facilitate the understanding of the general solution, it is instructive to look at the minimization problem under one of the 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: 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. This follows because the constraint $\mu = E_T[F]$ implies a zero penalty term, but a nonzero pricing error. The GMM objective function f_x is equal to SSE and minimizing SSE also minimizes f_x for all x. The first n sample moment conditions are

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

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.

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

Importantly, this solution does not depend on the log weight x on the penalty term.

[3]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 (1999) and others.

Trivial constrained optimum: Under the constraint $\mu = E_T[F] - \frac{1}{\lambda}$, the trivial constrained optimization returns a trivial solution with spuriously high R^2 and tiny pricing errors. This follows from the constant term in Equation (A.8) being 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 low, minimizing f_x essentially boils down to minimizing the sum of squared pricing errors SSE. Considering Equation (A.8), this is done by bringing the covariance terms to zero and the constant term to one. As a consequence, 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.

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 see in the simulation study in Section B.1 of this appendix and in the paper, 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 $(0, \pm\infty)$ 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.

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 (A.8) is then given as

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

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

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

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 (A.8) to one by constraining $\lambda = (E_T[F] - \mu)^{-1}$ gives

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

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

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

Substituting this solution back into f_x gives the minimum value

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

Global optimum: From Equations (A.10), (A.12) and (A.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 (A.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 (A.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 case of a perfectly unpriced factor.

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 B of this appendix.

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 A.2.1 largely carry over to this case, except for the desired constrained optimization problem.

Desired constrained optimum: Under the constraint $\mu = E_T[F]$, the moment condition (A.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:

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

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 (A.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 controlled environment in Section B of this appendix.

A.2.2 Multiple factors

With two or more factors, there is one "constant term" and one "covariance term" (see Equation (A.2)) 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 discussed above, it is relatively cheaper to do so in terms of penalty terms when a factor is relatively weaker. As a consequence, in empirical applications like the ones presented in Section 3 in the paper, weaker factors have the potential to "drive out" stronger factors, when the GMM weighting matrix is chosen inappropriately.

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

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

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 perfectly weak factor is enough to produce 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 large, either positive or negative.

As we will see in our controlled environment in Section B.3, factors which are rather 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 weaker 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 next section.

B Appendix to the simulation exercise

This section provides details about the data-generating process we use to simulate data in our controlled environment (Section B.1). In the other subsections, we provide additional information about types of factors we have left out in the paper. Section B.2 considers strong and priced factors, Section B.3 analyses the impact of factor strength, Section B.4 studies how the bias varies with changing sample sizes, and Section B.5 takes a closer look at multifactor models.

B.1 Data-generating process for simulation exercise

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

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

where the ε's are independent from one another and from the factors and are also i.i.d. normally distributed with means of 0 and standard deviations 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 B.4. 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, orthogonalize 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

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

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 and have a cross-sectional mean of zero.

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

$$\beta_{i,j} = \frac{\sigma_r}{\sigma_f} \cdot s_j \cdot \frac{b_{i,j}}{\max(|b_{\cdot,j}|)}, \tag{B.3}$$

$$\sigma_i = \frac{\sigma_r}{\sigma_f} \sqrt{1 - \sum_{j=1}^{k} \left(s_j \cdot \frac{b_{i,j}}{\max(|b_{\cdot,j}|)} \right)^2}, \tag{B.4}$$

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 (B.3) and (B.4) then reduce to $\beta_i = \frac{\sigma_r}{\sigma_f} \cdot s \cdot b_i / \max(|b|)$ and $\sigma_i = \frac{\sigma_r}{\sigma_f} \cdot \sqrt{1 - (s \cdot b_i / \max(|b|))^2}$, respectively. The time series R^2 of the factor model for return i is given by

$$\frac{(\beta_i \sigma_f)^2}{(\beta_i \sigma_f)^2 + (\sigma_i \sigma_\varepsilon)^2} = \frac{\beta_i^2}{\beta_i^2 + \sigma_i^2} = s^2 \frac{b_i}{\max(|b|)}, \tag{B.5}$$

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.[4] 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 (B.3) into (B.2), 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.

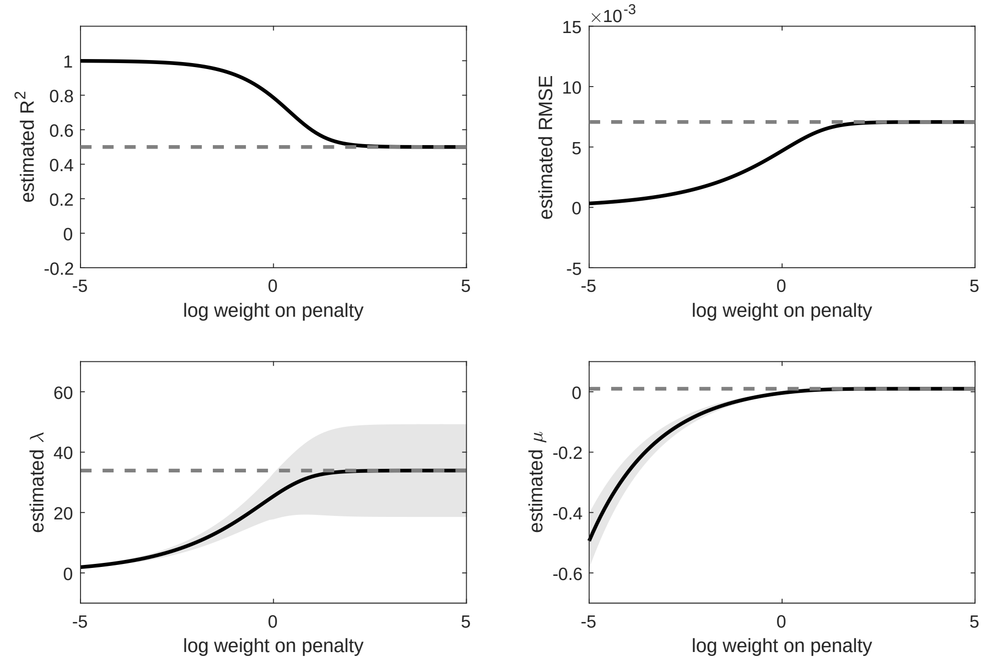
B.2 The impact of true cross-sectional explanatory power

To understand the impact of the true cross-sectional R^2 on the estimated statistics, we 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 B.1 shows the estimation results. They suggest that the case of (imperfectly) priced factors is very similar to the case of perfectly unpriced factors discussed in the paper. 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.

[4]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 (B.4) is positive for all i. The most conservative way of doing so is to assume that $\sum_{j=1}^{k} s_j \leq 1$.

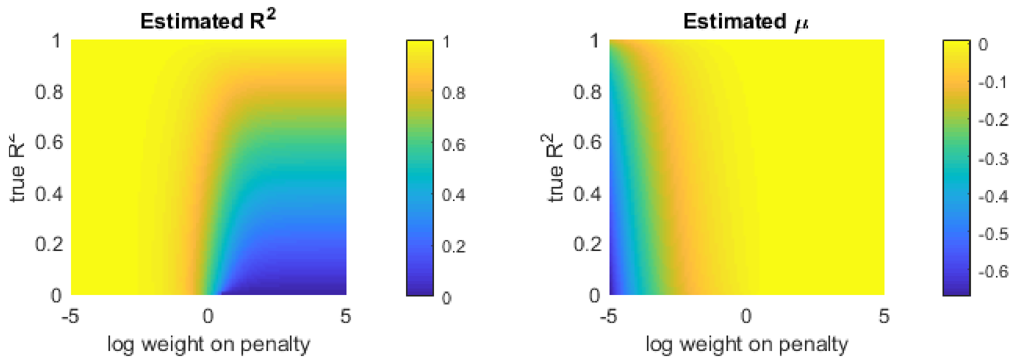
Figure B.1:
A single strong factor with a cross-sectional R^2 of 0.5



We apply a GMM estimation with the moment conditions in Equation (2) and the weighting matrix in Equation (5) in the paper. 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 (B.1)-(B.4) with $r = \sqrt{0.5}$ and $s = 1$.

Figure B.2 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 B.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 (2) and the weighting matrix in Equation (5) in the paper. 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 (B.1)-(B.4) with $s = 1$.

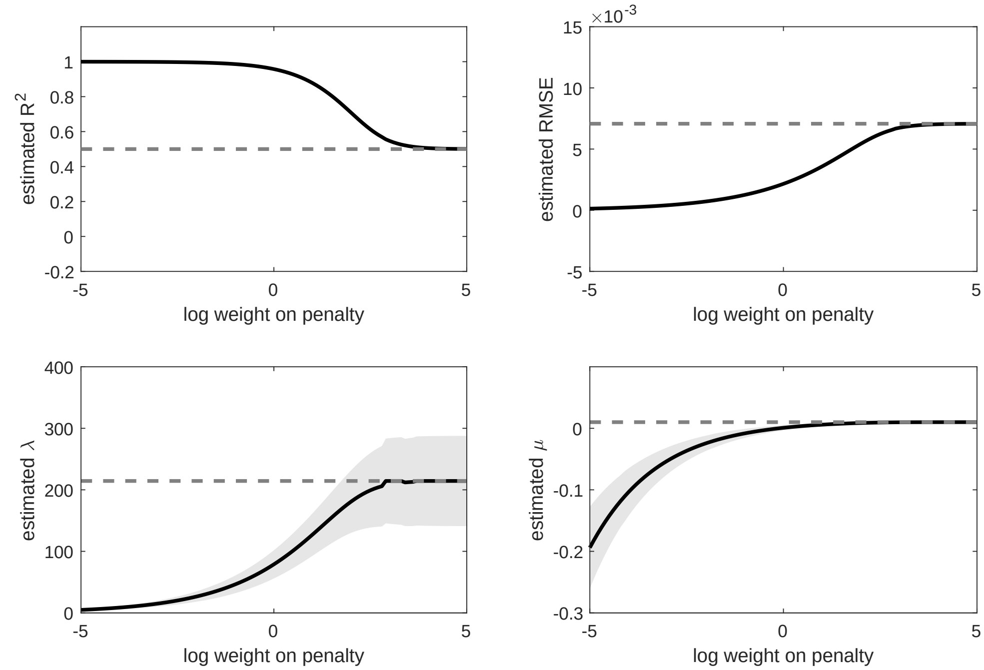
B.3 The impact of factor strength

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 (B.3) and (B.4), while keeping r fixed at $\sqrt{0.5}$. As a starting point, Figure B.3 shows the estimated quantities as functions of x for a very low s of $\sqrt{0.025}$.[5] 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 B.1 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 B.3 compared to Figure B.1.

Qualitatively, we observe a pattern that is very similar to Figure B.1. 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 B.1. 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

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

Figure B.3:
A single weak factor with a cross-sectional R^2 of 0.5

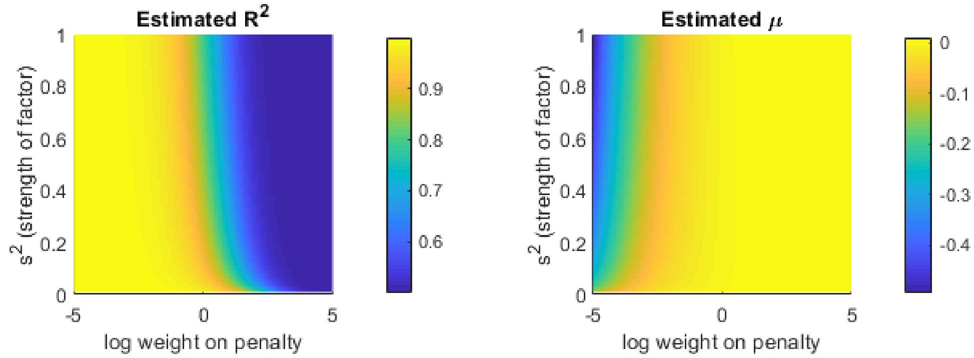


We apply a GMM estimation with the moment conditions in Equation (2) and the weighting matrix in Equation (5) in the paper. The figure shows estimated R^2 and RMSE and the point estimates of λ and μ, together with 95% confidence bounds as functions of x, the log weight on the moment condition that identifies the factor mean. Returns are simulated according to Equations (B.1)-(B.4) with $r = \sqrt{0.5}$ and $s = \sqrt{0.025}$.

tested. As discussed in Section A.2 of this appendix, if a factor is very weak, the covariance terms in Equation (A.2) 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 B.4 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 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.

Figure B.4:
A single strong factor with a cross-sectional R^2 of 0.5



We apply a GMM estimation with the moment conditions in Equation (2) and the weighting matrix in Equation (5) in the paper. 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 (B.1)-(B.4) with $s = 1$.

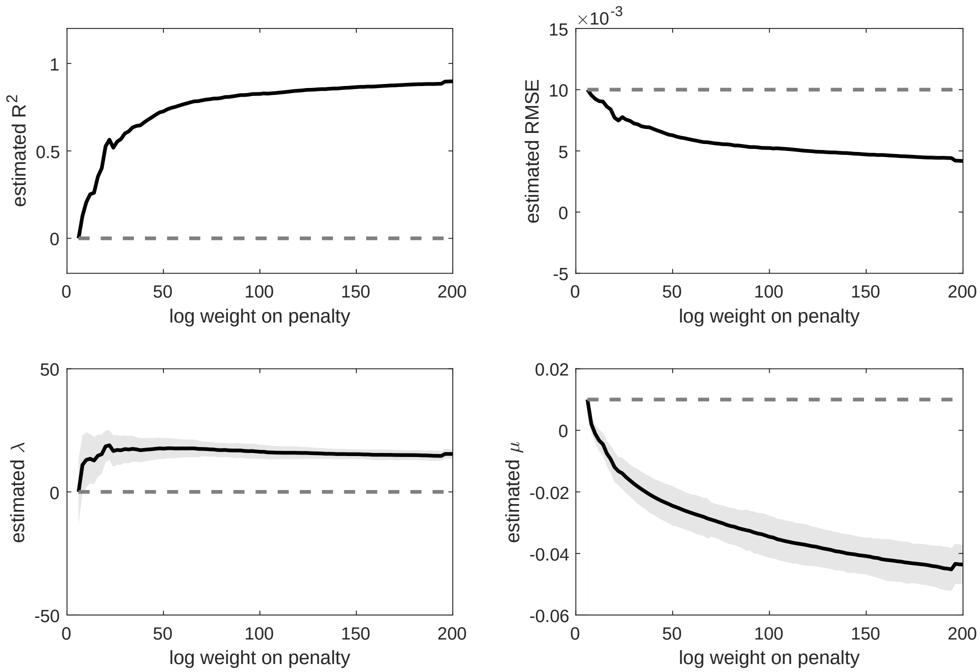
B.4 The impact of sample size and number of test assets

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 population counterparts 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 A.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 B.5 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 B.5:
The impact of the number of test assets



We apply a GMM estimation with the moment conditions in Equation(2) in the paper 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 (B.1)-(B.4) 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.

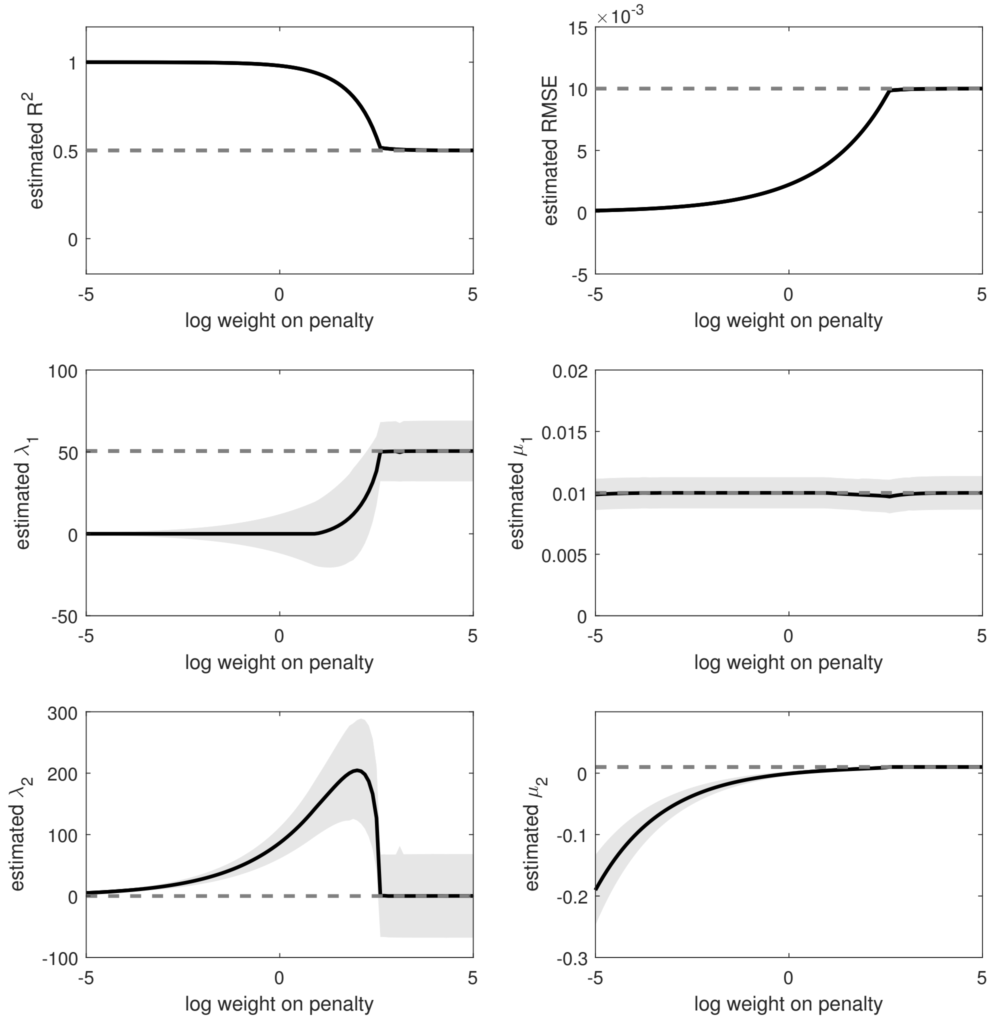
We conclude that the bias in parameter estimates does not depend on the sample size, but the number of test assets matters. Ceteris paribus, the bias is more severe if the number of test assets is large. This is in contrast to, e.g. Lewellen et al. (2010), who suggest increasing the number of test assets 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 worse.

B.5 Models with two factors

In this subsection, we discuss cases 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 Appendix A.2.2. The true cross-sectional R^2 is equal to $0.5 = \frac{1}{2}(r_1^2 + r_2^2)$. Figure B.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 (A.8) close to one. As discussed in Section A.2.2, 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 weaker 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 B.6. For weights below the critical value, λ_2 estimates are positive and seem 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

Figure B.6:
One strong and priced and one weak and unpriced factor



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

in Figure 1 in the paper. The difference is that the critical log weight x is larger in Figure B.6, since the weaker unpriced factor F_2 here is much weaker than the unpriced factor in Figure 1 in the paper.

For low values of x, the parameters $(\hat{\lambda}_2, \hat{\mu}_2)'$ correspond to a trivial solution, applied to the weaker 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 weaker 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 weaker and perfectly unpriced factor F_2 is estimated at 88.69 (true value is 0) with a t-statistic of 6.51. The estimated cross-sectional R^2 is equal to 97.96% (true value is 50%) and the estimated root mean squared pricing error is 0.22% (true value is 1%).

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

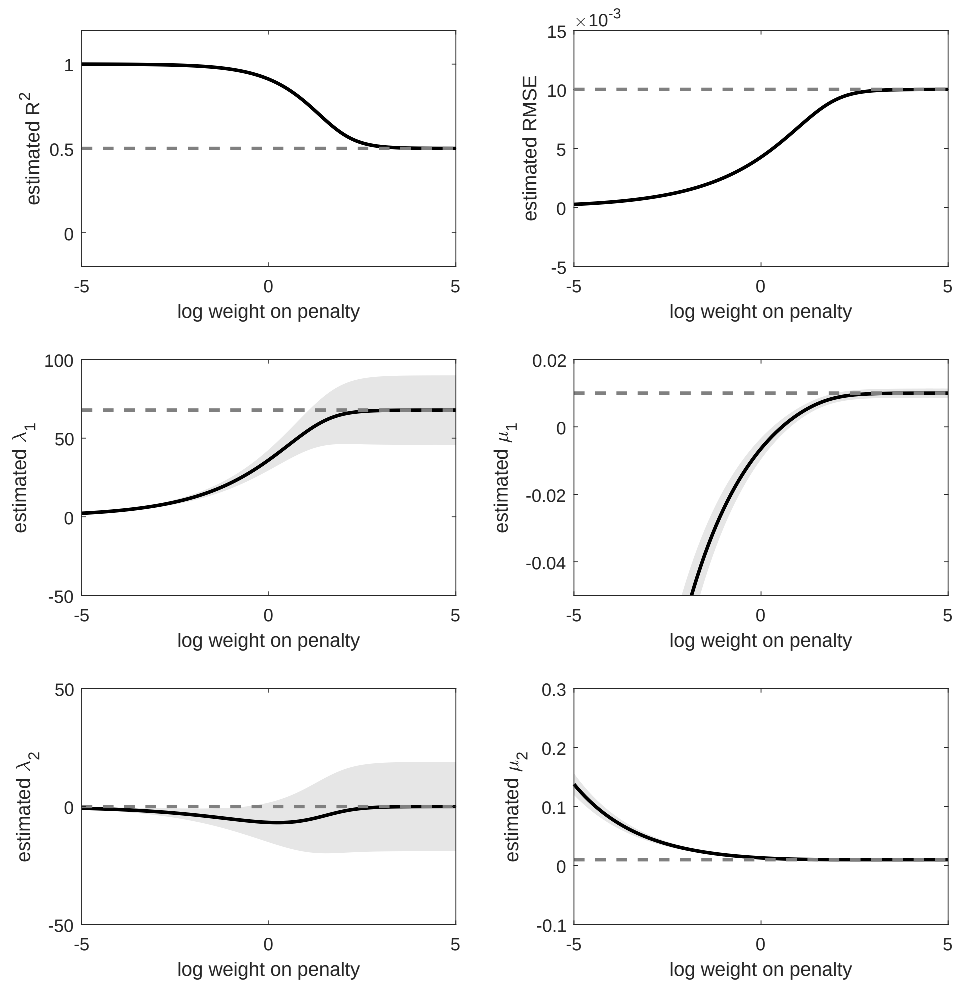
$$(Cov(R^e, F_1))'Cov(R^e, F_1) = (Cov(R^e, F_2))'Cov(R^e, F_2) \neq 0,$$

where factor F_1 is priced and factor F_2 is perfectly unpriced. In terms of the parameters in Equations (B.1) to (B.4), 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 B.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 B.6, the critical value is now slightly smaller, since the two factors are much stronger (compare the analysis in Section B.3).

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

Figure B.7:
One strong and priced and one strong and perfectly unpriced factor



We apply a GMM estimation with the moment conditions in Equation (2) and the weighting matrix in Equation (5) in the paper. 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 (B.1)-(B.4) with $r_1 = 1$, $r_2 = 0$, $s_1 = \sqrt{0.5}$, and $s_2 = \sqrt{0.5}$.

to decrease the pricing errors. Compared to the analysis of Figure B.6, showing that weaker factors drive out stronger factors, we thus cannot conclude that unpriced factors drive out priced factors if they are comparable in terms of strength. Still, parameter estimates are biased and R^2's are inflated when the weight on the penalty term is too low.

C Detailed review of papers using the GMM design discussed in our paper

In this part of the appendix, we list a few empirical asset pricing papers that rely on GMM with moment conditions as discussed in our paper.[6] Some papers report estimates from a Fama and MacBeth (1973) regression in addition to the results of the GMM estimation, but do not discuss differences from the GMM estimates. We also briefly highlight further extensions of the GMM procedures if there are any. Importantly, none of these extensions is reportedly addressing the issue discussed in our paper.

Maio and Santa-Clara (2012): The moment conditions in this paper are exactly the ones discussed here:

$$\begin{bmatrix} E[R^e_{t+1} - R^e_{t+1}(f_{t+1} - \mu_f)'\lambda] \\ E[f_{t+1} - \mu_f] \end{bmatrix} = 0. \tag{C.1}$$

In their benchmark specification, the authors apply a one-stage GMM with an identity matrix for the weights. In robustness checks, they extend the procedure, e.g., by running a two-stage GMM, adding further test assets, or implementing bootstrap procedures to robustify the p-values for the estimated parameters. They also run two-pass time-series/cross-sectional regressions, but do not interpret the results of these regressions.

Maio (2013a,b): The benchmark empirical estimation relies on a one-stage GMM using the exact design studied here, with identity matrix for weighting. Again, the author runs the GMM estimation on various sets of test assets, and he also includes a detailed analysis of the individual pricing errors of each test portfolio, as well as a bootstrap analysis for the risk premia estimates. In Table 15 of the online appendix, the paper also reports results from two-pass

[6] We skip the paper of Yogo (2006) because it is discussed at length in Sections 3.2 and 3.3 of the paper and in Section D of this appendix.

time-series/cross-sectional regression.

Lioui and Maio (2014): The main empirical estimation relies on a GMM design similar to Maio and Santa-Clara (2012) with identity weighting matrix. The analysis also provides a two-stage GMM estimator, using the inverse of the covariance matrix estimated with the parameters of the first stage for weighting. The authors also provide a weighted least squares coefficient of determination R^2_{WLS}, which assigns less weight to noisier pricing errors. The authors analyze the factor loadings (betas) in more detail to understand the economic intuition behind their results. In Table 7 of the online appendix, they report results from two-pass time-series/cross-sectional regressions.

Tedongap (2015): This author roughly also follows the procedure of Yogo (2006). The linearized model features consumption growth, expected consumption growth and consumption volatility as factors. It is estimated via one-stage GMM with two sets of moment conditions: one targeting the pricing errors (similar to the moment condition discussed here, but including an intercept), the other one targeting time series moments of consumption growth. The weighting matrix is block-diagonal.

Darrat et al. (2011): These authors closely follow the procedure of Yogo (2006) (see Section 3.3 of the paper and Section D of this appendix), but include an instrument z_t (as it is also done by Yogo (2006) in an extension of the GMM procedure). The moment conditions are

$$\left[\begin{array}{c} E[R^e_{t+1} - R^e_{t+1}(f_{t+1} - \mu_f)'\lambda] \\ E[f_{t+1} - \mu_f] \end{array} \right] \otimes z_t = 0. \tag{C.2}$$

The weighting matrix in the first stage is given by

$$W = \begin{pmatrix} \nu I_N & 0 \\ 0 & \Sigma^{-1}_{ff} \end{pmatrix} \otimes z_t \tag{C.3}$$

The instruments include a constant and the lagged values of world consumption growth, the US wealth-consumption ratio, and the US term spread.

Chen and Lu (2017): The moment conditions are again the ones discussed in our paper (with and without instruments). The authors apply a two-stage GMM estimator, where the weighting matrix in the first stage is chosen following Yogo (2006):

$$W = \begin{pmatrix} kI_N & 0 \\ 0 & \Sigma_{ff}^{-1} \end{pmatrix} \tag{C.4}$$

where $\Sigma_{ff} = E_T[(f - \mu_f)(f - \mu_f)']$ and $\mu_f = E_T[f]$. For robustness, the authors also run Fama-MacBeth regressions.

Da et al. (2016): The paper analyzes models with log-linear pricing kernels as well as standard linear factor models. For these linear factor models, they apply the familiar moment conditions

$$E\begin{bmatrix} R_t^x - R_t^x(f_t - \mu_f)'\lambda \\ f_t - \mu_f \end{bmatrix} = 0. \tag{C.5}$$

They run a two-stage GMM estimation, where the initial values for the parameters are taken from OLS regressions. The weighting matrix in the first stage is a block-diagonal matrix as suggested by Cochrane (2005) and based on these OLS estimates.

Grammig et al. (2009): The moment conditions for the linearized model in this paper are similar to the ones discussed here, but include an intercept:

$$\begin{bmatrix} E[R_{t+1}^e - \alpha\iota_N - R_{t+1}^e(f_{t+1} - \mu_f)'\lambda] \\ E[f_{t+1} - \mu_f] \end{bmatrix} = 0. \tag{C.6}$$

The authors apply a one-stage GMM estimator with an identity weighting matrix.

D The durable consumption model: details

D.1 Estimation algorithm

In this appendix we provide details on the analysis of the durable consumption model discussed in Section 3.3 of the paper. The moment conditions are given in Equation (A.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 D.2 and D.3). Then we perform an analysis similar to the one in Sections 2.3, 2.4, and 3 of the paper, 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.[7] 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 (A.5)). This allows the estimation

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

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

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

where $\widehat{\Omega}^{(1)}_{i,\dots,j}$ denotes the submatrix $\begin{pmatrix} \widehat{\Omega}^{(1)}_{i,i} & \dots & \widehat{\Omega}^{(1)}_{i,j} \\ \vdots & & \vdots \\ \widehat{\Omega}^{(1)}_{j,i} & \dots & \widehat{\Omega}^{(1)}_{j,j} \end{pmatrix}$ 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:

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

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.

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 (A.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).[8] 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 D.1 and D.2 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.

D.2 GMM: First stage

We are going to discuss the results from the first stage of the GMM (Table D.1) 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 re-

[8]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 D.1: **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)
		[0.137]	[0.214]	[2.070]	[2.049]	[0.836]	[0.858]	[1.357]	[1.400]
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)
		[3.867]	[1.894]	[-1.911]	[-1.854]	[-3.590]	[-2.889]	[-2.364]	[-2.266]
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)
		[0.370]	[0.145]	[-1.554]	[-1.515]	[-1.402]	[-1.255]	[-1.356]	[-1.343]
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)
		[-0.577]	[-1.271]	[-1.188]	[-1.083]	[0.510]	[1.204]	[-0.283]	[-0.304]
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)
		[-2.435]	[-1.481]	[1.859]	[1.727]	[3.450]	[1.797]	[2.151]	[2.024]
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)
		[-0.514]	[-0.541]	[-0.832]	[-0.772]	[0.431]	[0.493]	[-0.678]	[-0.641]
MAE (in %)	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. Heteroskedasticity- and autocorrelation-consistent (HAC) standard errors (in parentheses) are calculated using a Bartlett kernel. Test statistics of t-tests of the hypotheses $\lambda = 0$ and $\mu = \bar{F}$ are shown in brackets.

ported 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. 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 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 unconstrained 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 D.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.

D.3 GMM: Second stage

Next, we analyze the robustness of the results from the second stage of the GMM estimation (Table D.2). 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.[9] We conclude that the fact that the point estimates in Yogo (2006) seem to converge after two stages of GMM is an artefact and slight changes in the estimation procedure destroys this property.

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

Table D.2: **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)
		[0.572]	[1.643]	[-1.160]	[-1.136]	[-2.427]	[-2.139]	[-1.484]	[-1.235]
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)
		[10.961]	[7.379]	[2.470]	[2.345]	[7.171]	[6.849]	[1.817]	[2.198]
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)
		[0.776]	[0.616]	[0.918]	[0.913]	[4.170]	[3.973]	[2.067]	[1.906]
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)
		[0.575]	[-1.365]	[0.372]	[0.372]	[0.029]	[0.294]	[0.070]	[0.023]
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)
		[-6.173]	[-4.562]	[-3.107]	[-3.018]	[-5.340]	[-3.519]	[-2.821]	[-2.750]
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)
		[-0.514]	[-0.091]	[0.335]	[0.326]	[-0.710]	[-0.477]	[0.393]	[0.325]
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. Heteroskedasticity- and autocorrelation-consistent (HAC) standard errors (in parentheses) are calculated using a Bartlett kernel. Test statistics of t-tests of the hypotheses $\lambda = 0$ and $\mu = \bar{F}$ are shown in brackets.

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.

References

Chen, Z. and A. Lu (2017): “Seeing the Unobservable from the Invisible: The Role of CO2 in Measuring Consumption Risk,” *Review of Finance*, 22, 977–1009.

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, 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.

Fama, E. and K. French (2015): “A Five-Factor Asset Pricing Model,” *Journal of Financial Economics*, 116, 1–22.

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

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

Grammig, J., A. Schrimpf, and M. Schuppli (2009): “Long-horizon consumption risk and the cross-section of returns: new tests and international evidence,” *European Journal of Finance*, 15, 511–532.

Hansen, L. (1982): “Large Sample Properties of Generalized Method of Moments Estimators,” *Econometrica*, 50, 1029–1084.

Kan, R. and C. Zhang (1999): “Two-Pass Tests of Asset Pricing Models with Useless Factors,” *Journal of Finance*, 54, 203–235.

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, 483–511.

Maio, P. (2013a): “Intertemporal CAPM with Conditioning Variables,” *Management Science*, 59, 122–141.

——— (2013b): “Return decomposition and the Intertemporal CAPM,” *Journal of Banking and Finance*, 37, 4958–4972.

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

Newey, W. and K. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Tedongap, R. (2015): "Consumption Volatility and the Cross-Section of Stock Returns," *Review of Finance*, 19, 367–405.

Yogo, M. (2006): "A Consumption-Based Explanation of Expected Stock Returns," *Journal of Finance*, 61, 539–580.