Elephants and the Cross-Section of Expected Returns

This version: September 3, 2019

Abstract

Using GMM for cross-sectional asset pricing tests can generate spuriously high explanatory power for factor models by allowing 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 weak as well as strong factors, and to all sample sizes and test assets. To quantify the severity of the problem, we run tests based on simulated and empirical data.

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

JEL: G00, G12, C21, C13
1 Introduction

The amount of factor models that have been proposed to explain the cross-section of expected stock returns is enormous, and the literature keeps on growing. In light of the recent discussion about “taming the factor zoo”, we show theoretically, empirically, and via Monte Carlo simulations that using GMM for cross-sectional asset pricing tests 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 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 various researchers, e.g., Yogo (2006), Dhume (2010), Darrat et al. (2011), Maio and Santa-Clara (2012), Maio (2013), Lioui and Maio (2014), Da et al. (2016), and Chen and Lu (2017). The estimator that we consider uses the moment condition $0 = E[R^e_t - R^e_i (F - \mu_F) \lambda]$, which states that expected excess returns must be linear in the covariance between excess returns $R^e$ and a candidate factor $F$. Key to the estimator studied in this paper is a further moment condition of the form $0 = E[F - \mu_F]$ which identifies the factor mean $\mu_F$.

We show that the GMM estimator produces spurious results for particular choices of the initial GMM weighting matrix. Too small a weight on the latter moment condition can lead to an imprecise estimate of $\mu_F$ in favor of an improved cross-sectional fit, i.e., smaller pricing errors. Importantly, even with standard choices for that matrix, such as the identity matrix, the estimation results will be spurious in many cases.

We identify three major problems: First, weak factors can erroneously look “priced”, i.e., have large and significant market price of risk (MPR) estimates and high explanatory power for the cross-sectional variation in expected returns. Second, the point estimates of the MPR of

---

1 For a selection of the most recent approaches see, e.g., the papers by Harvey et al. (2016), Kogan and Tian (2014), Freyberger et al. (2017), or Harvey and Liu (2017).

2 Ideally, one would like to use the optimal weighting matrix. However, since this matrix is unknown, it has to be estimated from the data itself. As is well known from the literature (see, e.g., Cochrane (2005), p. 226), iterative procedures do not necessarily converge towards such an optimal weighting matrix. We find this pattern in our empirical examples as well.
strong factors, which truly have explanatory power for the cross-section, can be close to zero, i.e., much smaller than the true value, and thus be falsely considered irrelevant. In contrast to the first scenario, which represents a type I error, the second scenario corresponds to a type II error. Third, when weak and strong factors are considered jointly, weak factors can “drive out” strong factors, in the sense that MPR estimates of strong factors are small and insignificant, while those of weak factors are large and significant.

We exemplify these issues using an obviously economically meaningless factor, namely the log growth rate of the number of captive Asian elephants living in zoos around the world.\(^3\) Usually, there is a strong economic intuition why a candidate factor should be “priced” and if it should have a positive or negative MPR. Still, in empirical tests, we have to start from the null hypothesis that the factor is irrelevant. Our example shows that even in situations where this is undoubtedly true, GMM estimates can make the candidate factor look bright.

**Figure 1:**
Cross-sectional fit of the elephant model

The figure depicts realized average excess returns of the 25 Fama-French size- and book-to-market-sorted portfolios, plotted against the expected excess returns implied by a univariate model featuring the elephant factor. Model-implied expected excess returns are determined via a one-stage GMM procedure as outlined in Section 2, with three different weighting matrices for the moment conditions.

\(^3\)The data are available at [http://www.asianelephant.net/database.htm](http://www.asianelephant.net/database.htm). We thank the creators of this website, Jonas Livet and Torsten Jahn, for making these data publicly available.
Table 1: Market prices of risks for the elephant model

<table>
<thead>
<tr>
<th></th>
<th>A: Unit weight on $\mu_F$</th>
<th>B: High weight on $\mu_F$</th>
<th>C: Low weight on $\mu_F$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>101.25 (2.93)</td>
<td>558.25 (0.98)</td>
<td>10.47 (3.23)</td>
</tr>
<tr>
<td>Elephants</td>
<td>128.50 (5.81)</td>
<td>0.04 (0.00)</td>
<td>13.59 (5.90)</td>
</tr>
<tr>
<td>MKT</td>
<td>3.37 (2.94)</td>
<td>3.43 (2.91)</td>
<td>0.83 (3.49)</td>
</tr>
<tr>
<td></td>
<td>0.38 (0.46)</td>
<td>3.43 (2.88)</td>
<td>-0.01 (-0.06)</td>
</tr>
<tr>
<td>SMB</td>
<td>0.31 (0.19)</td>
<td>0.32 (0.20)</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td></td>
<td>-1.72 (-1.19)</td>
<td>0.32 (0.18)</td>
<td>-0.19 (-1.25)</td>
</tr>
<tr>
<td>HML</td>
<td>5.69 (3.49)</td>
<td>5.76 (3.42)</td>
<td>1.60 (6.04)</td>
</tr>
<tr>
<td></td>
<td>-0.47 (-0.39)</td>
<td>5.76 (3.52)</td>
<td>-0.13 (-1.02)</td>
</tr>
</tbody>
</table>

RMSE (in %) 1.26 2.00 1.04 5.64 2.02 2.02 0.09 0.43 0.07

$R^2$ (in %) 91.05 58.12 94.00 -174.15 56.64 56.64 99.90 97.47 99.94

The table reports estimates of $\lambda$ from a GMM estimation using the moment conditions $E[R_i - R_e (F - \mu_F)\lambda] = 0$ and $E[F - \mu_F] = 0$ along with three different weighting matrices. Heteroskedasticity- and autocorrelation-consistent (HAC) $t$-statistics (in parentheses) are calculated using a Bartlett kernel.

Graph A of Figure 1 shows the cross-sectional fit of a model featuring elephant population growth as the only factor. Using a quarterly postwar sample and the usual 25 size and book-to-market sorted portfolios as test assets, the cross-sectional $R^2$ is 91%, and the MPR estimate is statistically significantly different from zero with a $t$-statistic of 2.93. Table 1 contains point estimates, $t$-statistics, root mean squared pricing errors (RMSE), and cross-sectional $R^2$s, for this case and when elephant population growth is combined with the three factors from the Fama and French (1996) model.

Strikingly, the statistics reported in Panel A indicate that 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%), implying that just the elephant factor is enough to explain the cross-section of expected returns.

---

4Although 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 et al. (2016)) and factor structure in the Fama-French portfolios (see Lewellen et al. (2010)).
Note that in Panel A we use an identity matrix to weight the moment conditions. Panel B in turn shows that, with a weighting matrix putting a higher weight on the factor mean(s), the MPR estimate of the elephant factor is insignificant, and the cross-sectional $R^2$ is negative. Moreover, the MPR estimates of the strong factors MKT and HML are significant and stay unchanged when the three factor model is augmented by the elephant factor. The estimates in Panel B correspond to the results of standard Fama and MacBeth (1973) regressions.

While Panels A and B in Table 1 exemplify the problems in the cases of just a weak factor and when weak and strong factors are combined, Panel C shows that the choice of the weighting matrix can also impact MPR estimates of strong factors. There we use a low weight on the moment condition identifying the factor mean(s). While the MPR point estimates and standard errors of MKT and HML in the context of the three-factor model do not change much between Panels A and B (columns 2 and 5), the point estimates are much smaller in Panel C (column 8), such that a $t$-test clearly rejects the null that the MPR estimates are equal to the factor means (not tabulated). At the same time, all $R^2$s are very high in Panel C. We show in more detail in Section 3 that the $R^2$s vary between 1 and the (true) $R^2$ from a simple two-stage Fama-MacBeth regression as the weight on the moment condition identifying the factor mean varies from low to high.

To see the intuition behind these results, it is instructive to first rewrite the moment condition for the pricing errors:

\[
E[R_i^e - R_i^e(F - \mu_F)\lambda] = 0
\]

\[
\Leftrightarrow E[R_i^e] - E[R_i^e]E[F - \mu_F]\lambda - Cov(R_i^e, F)\lambda = 0.
\]

First consider the case of a weak factor $F$, i.e., a factor for which $Cov(R_i^e, F)$ is close to zero for all test assets $i$. The above moment condition will then be satisfied when $\lambda$ and $\mu_F$ are chosen such that $E[F - \mu_F]\lambda$ is (roughly) equal to one. In actual applications, the GMM algorithm will yield an estimate $\mu_F$ close to (but different from) the sample average of $F$, due to the second moment condition $E[F - \mu_F] = 0$. As a consequence, $\lambda \approx (E[F - \mu_F])^{-1}$ is large and appears
economically (and typically also statistically) significant.

In case of a strong factor $F$, i.e., when $\text{Cov}(R^c_i, F)$ is large in absolute terms and varies substantially across the test assets, the optimal $\lambda$ and $\mu_F$ are chosen differently. When the weight on the condition for the factor means is small enough, $\lambda$ will be close to zero in order to reduce the impact of the covariance term on the pricing errors. Again, $\mu_F$ is then set such that $E[F - \mu_F] \lambda$ is close to one. With $\lambda$ close to zero, this causes $\mu_F$ to be far away from the sample average of $F$.

In the presence of both weak and strong factors, weak factors have large $\lambda$ estimates and $\mu_F$ estimates such that $E[F - \mu_F] \lambda$ is close to one. Since the pricing errors are already close to zero due to the effect of the misestimation of the MPR for the weak factor, the MPRs of all other factors will be close to zero. As a consequence, weak factors seem to have a significant MPR, while strong factors may be considered irrelevant.

In all cases considered, small pricing errors come at the cost of not matching factor means. For instance, in the example presented in Graph A of Figure 1, the estimated factor mean is $-0.24$ percentage points per quarter, while the sample average is $0.47$ percentage points per quarter. We argue that researchers who estimate models for the cross-section of expected returns via GMM should not only report MPR estimates and cross-sectional $R^2$’s, but also estimated factor means, since deviations from the observed sample average can be informative about the problems we document in this paper.

Our paper comprises a theoretical and an empirical part. In Section 2, we discuss the theory for both weak and strong factors and perform simulations to analyze the impact of the weighting matrix in a controlled environment. In Section 3, we shed new light on the factor model proposed by Yogo (2006), which has been tested using the GMM procedure described above. It features the market return, the growth rate of nondurables and services consumption as well as the growth rate of durable consumption as factors. Results for 25 size- and book-to-market sorted portfolios are presented in Table 3 of his paper, most prominently 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 $-178$ for durable consumption growth. These results coincide with those from a standard Fama-MacBeth regression. Our empirical exercise also documents that a multi-stage GMM does not solve or even mitigate the concerns we raise. As is well known from the literature on cross-sectional asset pricing (see Cochrane (2005)), a multi-stage GMM does not necessarily converge towards an “optimal” weighting matrix, and we find this pattern here as well.

**Related literature.** Our paper is linked to several strands of the literature. First of all, 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 et al. (2014), Kleibergen and Zhan (2015), Bryzgalova (2016), or Burnside (2016). Stock and Wright (2000) develop the general asymptotic theory for GMM estimators with weak identification. However, we would like to emphasize that 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. In fact, there does not even have to be a hidden factor structure in returns that we want to uncover, as can be seen, e.g., from our simulation exercises. 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 choice of a certain structure of the GMM weighting matrix. Third, Lewellen et al. (2010) propose to compare the market price of risk estimates with the time series average returns of the factors in case the factors are traded. We argue much more generally
that researchers should compare the estimated factor mean with the time series average of the
factor, and these two quantities should be close to each other for all factors, no matter if traded
or not. A large divergence between them points towards a too small weight on the factor mean
in the GMM estimation and can lead to the problems investigated here in detail.

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.

In particular, the empirical parts of our paper are linked to the large literature inves-
tigating 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

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. Our paper shows that
the trade-off between estimating “correct” betas and fitting the cross-section translates into a
trade-off between cross-sectional fit and fit to macro fundamentals.

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 is, however, completely independent of the failure of the linear factor model that
we document. Besides Yogo (2006), other papers that estimate factor models with the GMM
technique described above are Dhume (2010), Darrat et al. (2011), Maio and Santa-Clara (2012),
Maio (2013), Lioui and Maio (2014), Da et al. (2016), and Chen and Lu (2017). None of these papers report estimated factor means or mention the trade-off that we discuss in our paper.

2 Cross-Sectional Regressions with GMM

We start our exposition by providing the theory behind our main argument that the results from estimating pricing models for the cross-section of expected stock returns via GMM are sensitive to the choice of the GMM weighting matrix. Although the argument is valid for both strong and weak factors, the rationale behind it depends on the nature of the candidate factors. We therefore split the exposition in three parts: We start with models that feature a single weak factor, move to a single strong factor in Section 2.2, and discuss the case of several factors in Section 2.3. In each subsection, we exemplify the theoretical argument through a simulation exercise in a controlled environment.

2.1 A single weak factor

Assume that there are \( n \) test assets with excess returns \( R_{i,t}^e \) (\( i = 1, \ldots, n \)) and a single candidate pricing factor \( F \). The standard moment conditions for a cross-sectional GMM estimation of this one-factor model are the following (for \( i = 1, \ldots, n \)):

\[
\begin{align*}
E[R_{i}^e] &= Cov[R_{i}^e, F] \lambda \\
&\Leftrightarrow E[R_{i}^e] = E[R_{i}^e F] \lambda - E[R_{i}^e] E[F] \lambda \\
&\Leftrightarrow 0 = E[R_{i}^e - R_{i}^e (F - E[F]) \lambda].
\end{align*}
\]

(1)

Here \( \lambda \) denotes the market price of \( F \)-risk, scaled by the variance of \( F \) (i.e. \( \lambda = \frac{MPR_F}{Var[F]} \)) and is the parameter to be estimated. If the pricing factor \( F \) is uncorrelated with all test asset excess returns, i.e. \( Cov(R_{i}^e, F) = 0 \) for all \( i \), then \( \lambda \) cannot be identified from these moment conditions alone since either Equation (1) holds for any value of \( \lambda \) (if all expected excess returns are zero).
or it does not hold for any value of λ (if at least one of the expected excess returns is nonzero).

The expectation of the pricing factor, \( E[F] \), is generally unknown. Of course, \( \bar{F} \equiv E_T[F] \) is an unbiased estimator, but with non-zero variance. To set this uncertainty in relation to the uncertainty about \( \lambda \), \( E[F] \) is often treated as a further parameter that is estimated jointly with \( \lambda \). Before discussing this case in detail, we replace \( E[F] \) by a parameter \( \mu_F \) and study the consequences of different choices of this parameter. The sample moment condition becomes

\[
0 = E_T[R_t^e - R_t^e (F - \mu_F)\lambda].
\]

If we set \( \mu_F \) to a value that is not exactly equal to the sample mean \( E_T[F] \), this condition can be rewritten as

\[
E_T[R_t^e] = E_T[R_t^e F]\lambda - E_T[R_t^e]\mu_F\lambda \\
\Leftrightarrow E_T[R_t^e] = E_T[R_t^e] \left( E_T[F] - \mu_F \right) \lambda + Cov_T(R_t^e, F)\lambda
\]

If the sample covariance \( Cov_T(R_t^e, F) \) is equal to zero for all test assets \( i \) and \( E_T[R_t^e] \neq 0 \) for at least one \( i \), there is a unique solution for \( \lambda \):

\[
\lambda = (E_T[F] - \mu_F)^{-1}
\]

Note that this relation holds irrespective of the sample size or of how the moment conditions are weighted. It implies that all test assets are priced perfectly, so that the cross-sectional \( R^2 \) is equal to 1. Summing up, in the theoretical case that \( Cov_T(R_t^e, F) = 0 \) for all \( i \), any choice of \( \mu_F \neq \bar{F} \) allows to explain the cross-section of expected returns perfectly, i.e. there exists a \( \hat{\lambda} \) for which pricing errors are equal to zero for all test assets.

The analogue of this result in the language of standard two-stage regressions a la Fama and MacBeth (1973) is explained in detail in Appendix A. We show that running the first stage

\footnote{Throughout the paper, we use the notation of Hansen (1982) in which a subscript \( T \) denotes the sample equivalent of a given moment.}
regression (the time-series regression that estimates the factor exposures) without an intercept, or setting the intercept to an arbitrary value that is different from the sample equivalent, leads to a spuriously large explanatory power of average returns by factor exposures in the second stage regression.

In applications, $\mu_F$ is typically not set to a particular value but estimated using a further moment condition of the form $E[F - \mu_F]$. Throughout the paper we assume that it is our goal to minimize the GMM objective function

$$f(\lambda, \mu_F) = g_T'(\lambda, \mu_F) W g_T(\lambda, \mu_F),$$

(4)

where $W$ denotes the GMM weighting matrix, and $g_T$ is the vector of sample moment conditions given by

$$g_T(\lambda, \mu_F) = E_T \left[ \begin{array}{c} R_1^e - R_1^e(F - \mu_F)\lambda \\ \vdots \\ R_n^e - R_n^e(F - \mu_F)\lambda \\ F - \mu_F \end{array} \right].$$

(5)

Intuitively, if the sample covariances between the test asset returns and the factor are all close to zero, the objective function is minimized by setting $\mu_F$ very close to $\bar{F}$ (due to the last moment condition) and choosing $\lambda \approx (\bar{F} - \mu_F)^{-1}$. The estimate for $\lambda$ is thus very large in absolute terms, and its sign depends on the sign of $\bar{F} - \mu_F$.

To substantiate this intuition, we perform a simulation exercise to study the impact of the weighting matrix in a clean environment. An unconstrained GMM estimation based on the moment conditions (5) yields a $\mu_F$ estimate which is arbitrarily close to $\bar{F}$ and lets $\hat{\lambda}$ explode in order to match asset pricing moments. We therefore constrain the estimate $\hat{\lambda}$ to be between $-B$ and $B$. We set $B = 50$, but any other number would deliver the same qualitative results.

We assume the data generating process $R_{i,t}^e = \alpha_i + \sigma \varepsilon_{i,t}$ for 25 test assets, where the
alphas are randomly assigned and vary between 0.3 and 2.7 percentage points per quarter. The return volatility is set to 8 percentage points quarterly for all assets and the $\varepsilon_{i,t}$ are i.i.d. standard normally distributed. We simulate only one sample with a sample size of 240, which corresponds to a post-war dataset with quarterly data. The realizations of the useless factor $F_t$ are also drawn from a normal distribution with a mean of 2 percentage points and a standard deviation of 4 percentage points. To make sure that the factor is uncorrelated with all excess returns even in the simulated sample, we then make the factor orthogonal to all 25 test assets. Afterwards we scale and shift the factor so that its mean and standard deviation are exactly equal to 2 and 4 percentage points in the finite sample, in order to ease the interpretation of the following numerical results.

To weight the sample moment conditions $g_T$ we use a matrix of the form

$$W = \begin{pmatrix}
1 & \ldots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots \\
0 & \ldots & 1 & 0 \\
0 & \ldots & 0 & 10^x
\end{pmatrix} = \text{diag}(1, \ldots, 1, 10^x).$$

$x$ denotes the log weight of the last moment condition, which is used to estimate the mean of the useless factor. In the following, we vary $x$ between -5 and 5 and study the result of the cross-sectional regression in terms of $R^2$'s and point estimates.\(^6\)

The upper graph in Figure 2 shows the cross-sectional $R^2$ as a function of $x$. We find that for a weight of $10^{-5}$, i.e., when we assign basically zero weight to the condition for the factor mean, the cross-sectional $R^2$ is equal to 1. Using the identity matrix (corresponding to $x = 0$) delivers qualitatively the same result. Increasing the relative weight further then ultimately leads to a decrease in $R^2$. The true $R^2$ of 0 is obtained when the weight approaches the upper limit of $10^5$.

\(^6\)To keep things simple, we perform only single-stage GMM estimations in this section. The case of multi-stage GMM is considered in Section 3.
The figure depicts $R^2$ and point estimates of coefficients as functions of the logarithm of the weight on the moment condition that identifies $\mu_F$. Test asset returns and the factor are simulated as outlined in Section 2.1.

The lower two graphs show point estimates of $\lambda$ in the left and $\mu_F$ in the right plot. The point estimate of $\lambda$ is equal to the upper bound $B$ for all values of the relative weight by construction as explained above. Economically more interesting are the shaded areas, which indicate confidence bounds produced by the estimation algorithm. These are obtained by adding and subtracting 1.96 HAC standard errors to the point estimates. It is standard in the asset pricing literature to consider such confidence bounds in order to evaluate whether an estimate is statistically significantly different from zero. In our case, we would conclude that the useless factor is useful in explaining cross-sectional variation in expected returns as long as the log weight on the moment condition for $\mu_F$ is less than roughly 2.14, i.e., the weight itself is below 140.
The lower right plot shows point estimates of $\mu_F$. With a small weight on the last moment condition, the estimated factor mean is $\bar{F} - \frac{1}{B}$, the solution of $B = (\bar{F} - \mu_F)^{-1}$, which is equal to zero in our case. The confidence band does not include the sample mean of 0.02. Putting more weight on the last moment condition brings the point estimate of $\mu_F$ closer to the sample mean $\bar{F} = 0.02$, which is marked by a line in the graph. With a high weight on the associated moment condition, the estimate of $\mu_F$ is equal to $\bar{F} = 0.02$ and, at the same time, the confidence band around the estimate of $\lambda$ is huge, which is in line with the true structure of expected returns. $\lambda$ should not be identified given that all sample covariances of returns with the factor are zero.

For a given choice of the weighting matrix, the objective function resulting from the moment condition (5) has two local minima, one where $\mu_F > \bar{F}$ and one where $\mu_F < \bar{F}$. Depending on which local minimum the numerical minimization will run into, the estimated $\hat{\lambda}$ will either be positive or negative. In the clean environment with zero correlation discussed in this subsection, these two local minima will both be global minima with identical function values. In the empirical cases discussed below, only one of these two minima is the global minimum. However, numerically, this implies that small changes in the weighting matrix can make the estimated $\hat{\lambda}$ switch sign and turn from a large positive number to a large negative number (or vice versa). In our implementation, we try to avoid this numerical problem by hand, for instance by trying out different starting values for the minimization algorithm. In particular, in the figures reported in this paper, we make sure that the $\hat{\lambda}$’s never switch sign (in the simulation exercises) or that we indeed find the global minimum (in the empirical cases).

### 2.2 A single strong factor

So far, our analysis has focused on weak factors, i.e., on the case when the covariance between the factor and the test portfolio excess returns is (close to) zero. Next, we document similar results for the case of strong factors. Remember that the moment condition targeting the pricing
errors can be rewritten as

\[
E_T[R^e_i] = E_T[R^e_i]\left(E_T[F] - \mu_F\right)\lambda + \text{Cov}_T(R^e_i, F)\lambda
\]  

(6)

A factor is called “strong” when the covariance term \(\text{Cov}_T(R^e_i, F)\) is different from zero and large, which allows to investigate the cross-sectional relation between factor exposures and expected returns. When \(F\) is a strong factor, there are conceptually two ways to minimize the GMM objective function. The first (and economically preferred) one would be to choose \(\mu_F\) close to the sample average \(\bar{F}\), i.e., to make the first term in Equation (6) close to zero. As a consequence, \(\lambda\) would pick up the cross-sectional relation between average returns and the covariance of returns with the factor, as desired.

The second way would be to choose \(\lambda\) close to zero in order to make the second term in Equation (6) small. Then we would be back in the situation outlined in the previous subsection and the result from the GMM estimation would be a \(\mu_F\) estimate that is far away from \(\bar{F}\) and a \(\lambda \approx (\bar{F} - \mu_F)^{-1}\) close to zero.

Depending on the GMM weighting matrix, the estimation result will be a mixture of these two extreme solutions. To show this, we run basically the same simulation exercise as in the previous subsection, but now we choose a different data generating process. We set \(R^e_{i,t} = \alpha_i + \beta_i F_t + \sigma \varepsilon_{i,t}\) for \(i = 1, \ldots, 25\) with different randomly assigned \(\alpha_i\)'s ranging from 0.3 to 2.7 percentage points, as in Section 2.1. \(F\) has a mean of 2 and a standard deviation of 4 percentage points in quarterly terms. As before, we make sure that these conditions exactly hold in sample. By construction, the true beta of portfolio \(i\) \((i = 1, \ldots, 25)\) with respect to the useful factor is \(0.5 + \frac{i-1}{24}\), i.e., the betas range between 0.5 and 1.5. As before, the noise terms \(\varepsilon\) are i.i.d. standard normals, the standard deviation \(\sigma\) is 8 percentage points, and we simulate one sample with 60 years of quarterly data.

Figure 3 shows the results from estimations analogous to the ones in the previous subsection. Most importantly, the results regarding the cross-sectional \(R^2\) found in the previous
subsection carry over to the case of a strong factor, since the cross-sectional goodness of fit ranges from the true value (implied by the data-generating process) of 0.3 to 1.

The figure depicts $R^2$ and point estimates of coefficients as functions of the logarithm of the weight on the moment conditions for $\mu_F$. Test asset returns and factors are simulated as outlined in Section 2.2.

The results regarding the parameter estimates support the intuition outlined above. The higher the weight on the moment condition targeting the factor mean, the closer the estimate of $\lambda$ is to the theoretically correct value of 20. With a small weight on this moment condition, $\lambda$ is set close to zero to downplay the impact of the sample covariance term in Equation (6). At the same time, if this weight approaches zero, the estimate of $\mu_F$ becomes arbitrarily large and negative.

Comparing these findings to the ones from Section 2.1, we see that too low a weight on the moment condition identifying the factor mean affects weak and strong factors differently.
It leads to large and significant estimates of the market price of risk for weak factors, while for strong factors it leads to estimates close to zero although the true parameter is actually much larger. Stated differently, the GMM test for the significance of a pricing factor is prone to both Type I and Type II errors. In both cases, however, we find that a small weight on the last moment condition yields point estimates of the factor mean which are far away from the sample average. Thus, researchers who use the cross-sectional estimation technique discussed here should always report the estimate of the factor mean and relate it to the sample average.

### 2.3 Weak and strong factors together

Our analysis so far has focused on one-factor models, but in practice, of course, models often feature multiple factors. As a final step, we document that our results carry over to multi-factor models and are also valid in the presence of both weak and strong factors. The theoretical mechanism is a mixture of the two mechanisms outlined in the previous subsections, so we turn directly to the results from the simulation exercise.

We assume the same data-generating processes as in Section 2.2 and add a weak factor that is constructed as explained in Section 2.1. The weak factor $F_2$ is orthogonal to all the test portfolio excess returns and to the strong factor $F_1$ in the finite sample. Both $F_1$ and $F_2$ have a mean of 2 and a standard deviation of 4 percentage points in quarterly terms. As before, we make sure that these conditions exactly hold in the sampled time series which contains 60 years of quarterly data.

We then perform a one-step GMM estimation of $\lambda_1$, $\lambda_2$, $\mu_{F,1}$, and $\mu_{F,2}$ using the moment
conditions

\[ g_t(\lambda_1, \lambda_2, \mu_{F,1}, \mu_{F,2}) = \begin{bmatrix}
R_{1,t} - R_{1,t}^e (F_{1,t} - \mu_{F,1}) \lambda_1 - R_{1,t}^e (F_{2,t} - \mu_{F,2}) \lambda_2 \\
\vdots \\
R_{n,t} - R_{n,t}^e (F_{1,t} - \mu_{F,1}) \lambda_1 - R_{n,t}^e (F_{2,t} - \mu_{F,2}) \lambda_2 \\
F_{1,t} - \mu_{F,1} \\
F_{2,t} - \mu_{F,2}
\end{bmatrix}. \]

We again assign varying weights to the latter two moment conditions using the weighting matrix

\[ W = \begin{pmatrix}
1 & \ldots & 0 & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots \\
0 & \ldots & 1 & 0 & 0 \\
0 & \ldots & 0 & 10^x & 0 \\
0 & \ldots & 0 & 0 & 10^x
\end{pmatrix} = \text{diag}(1, \ldots, 1, 10^x, 10^x). \]

We produce graphs analogous to those presented above, but now we plot \( \hat{\mu}_{F,1}, \hat{\mu}_{F,2} \) and the two market prices of risk estimates \( \hat{\lambda}_1 \) and \( \hat{\lambda}_2 \) with confidence bands. Like before, we constrain \( \hat{\lambda}_2 \) to be between -50 and 50.

Figure 4 depicts the results. The upper picture again shows the cross-sectional \( R^2 \) as a function of the log weight assigned to the moment condition for the factor means. For a log weight below zero, the cross-sectional \( R^2 \) is again equal to 1, which (falsely) indicates that the two-factor model delivers a perfect cross-sectional fit. As the weight increases, the \( R^2 \) converges to the true \( R^2 \) which is again equal to 0.3.

The lower graphs again show point estimates of the market prices of risk and factor means. For small values of \( x \), the market price of risk of the strong factor is estimated to be zero while that of the weak factor is large and significant. We have seen the same patterns for low values of \( x \) in Sections 2.1 and 2.2 and in the introductory elephant example. To find out which factor is actually causing the pricing errors to be low and the \( R^2 \) to be close to 1 in this case, we have
The figure depicts $R^2$ and point estimates of coefficients as functions of the logarithm of the weight on the moment conditions for $\mu_{F,1}$ and $\mu_{F,2}$. Test asset returns and factors are simulated as outlined in Section 2.3. $F_1$ denotes the useful factor and $F_2$ denotes the useless factor.

to consider the point estimates of the factor means. Here, we see that the mean of the strong factor is correctly estimated for low values of $x$ which means that all the explanatory power for expected returns in the cross-section is coming from the weak factor. Indeed, the estimate of $\lambda_1$
is not only close to zero, as in Section 2.2, but virtually equal to zero, such that the covariance term in Equation (6) does not matter. In this case, the estimate of $\mu_1$ is irrelevant for the pricing errors, so the algorithm estimates this factor mean correctly to avoid any penalty from the associated moment condition, even if the weight on this condition is low.

We find a similar pattern when the identity matrix is used for weighting. The returns and factors we simulate have moments and covariance structures similar to standard test asset returns and factors that are used in the empirical asset pricing literature. In the light of these results, the finding documented in Table 1 in the introduction, in particular that the market prices of risk of the three Fama-French factors become insignificant when the elephant factor is included and the identity matrix is used for weighting, is not surprising.

Increasing $x$ leads to larger confidence bands for the market price of risk of the weak factor (which should not be identified by the moment conditions), a market price of risk estimate of the strong factor that is significantly positive, and point estimates of the factor means that are close to the sample means.

3 Empirical evaluation of the durable consumption model

3.1 Model and moment conditions

We now turn to one example from the literature in which the estimation design outlined above plays a major role, and show step by step how changes in the weighting matrix can severely affect the pricing performance of the model. 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), the log growth rate of consumption of durable goods ($F_2$), and the log return on the aggregate stock market ($F_3$). We denote the vector including these three factors by $F$. 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.\footnote{See \url{http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html}.} We denote the excess returns by \( R_{1,t}, \ldots, R_{25,t} \). The sample moment conditions used to estimate the model parameters and evaluate the goodness of fit via GMM are

\[
g_T(\lambda, \mu_F) = E_T \begin{bmatrix} R_{1,t}(1 - (F_{1,t} - \mu_{F,1})\lambda_1 - (F_{2,t} - \mu_{F,2})\lambda_2 - (F_{3,t} - \mu_{F,3})\lambda_3) \\ \vdots \\ R_{25,t}(1 - (F_{1,t} - \mu_{F,1})\lambda_1 - (F_{2,t} - \mu_{F,2})\lambda_2 - (F_{3,t} - \mu_{F,3})\lambda_3) \\ F_{1,t} - \mu_{F,1} \\ F_{2,t} - \mu_{F,2} \\ F_{3,t} - \mu_{F,3} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{7}
\]

The parameters to be estimated are the three market prices of risk (scaled by the factor variance) \( \lambda = (\lambda_1, \lambda_2, \lambda_3)' \), and the factor means \( \mu_F = (\mu_{F,1}, \mu_{F,2}, \mu_{F,3})' \). In total, there are 28 moment conditions and 6 parameters to be estimated.

## 3.2 Estimating the model

### 3.2.1 Algorithm

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 3.2.2 and 3.2.3). Then, in Section 3.3, we perform an analysis similar to the one in Section 2, 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 \textsc{Gauss} and available on Motohiro Yogo’s website.\footnote{See \url{https://sites.google.com/site/motohiroyogo/}.} Our replication code is a line-by-line translation to \textsc{Matlab}. 

\[
R_{25,t}(1 - (F_{1,t} - \mu_{F,1})\lambda_1 - (F_{2,t} - \mu_{F,2})\lambda_2 - (F_{3,t} - \mu_{F,3})\lambda_3) \\
F_{1,t} - \mu_{F,1} \\
F_{2,t} - \mu_{F,2} \\
F_{3,t} - \mu_{F,3} 
\]
In the following, we describe the exact GMM algorithm used by Yogo (2006) to estimate the six parameters $\lambda$ and $\mu_F$.

1. **Initial parameter values.** The initial value for $\mu_F$ is set to the sample average of $F$ and the initial value for $\lambda$ is the solution of the system of linear equations $E_T[R^e] = [(R^e)'(F - \mu_F)/T] \lambda$. 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 $\lambda$ is calculated.

2. **Covariance matrix of moment conditions.** The initial values calculated in Step 1 are plugged into the moment condition function (Equation 7). This allows the estimation of the covariance matrix $\hat{\Omega}^1$ 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

$$W_1 = \begin{pmatrix} \det(\hat{\Omega}^1_{1,\ldots,25})^{-\frac{1}{2}} I_{25} & 0 \\ 0 & (\hat{\Omega}^1_{26,\ldots,28})^{-1} \end{pmatrix}$$

where $\hat{\Omega}^1_{i,\ldots,j}$ denotes the submatrix of $\hat{\Omega}^1$ 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:

$$\lambda_3 < 1$$

$$\lambda_1 + \lambda_2 + \lambda_3 > 0.$$ (8)

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

4. **Cross-sectional $R^2$ 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 (7) to obtain a second estimate $\hat{\Omega}^2$ of the covariance matrix of the moment conditions.

6. **GMM: second stage.** $(\hat{\Omega}^2)^{-1}$ 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 $\lambda$ 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 $\mu_F$ 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). 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 2 and 3 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”).

---

9We 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](https://personalpages.manchester.ac.uk/staff/Alastair.Hall/GMMGUI.html).
Table 2: GMM — First stage

<table>
<thead>
<tr>
<th></th>
<th>Constrained estimation</th>
<th>Unconstrained estimation</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yogo</td>
<td>Replica</td>
<td>25 portf &amp; nonpara</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>15.535</td>
<td>15.535</td>
<td>69.109</td>
<td>256.877</td>
<td>264.701</td>
<td>70.902</td>
<td>125.284</td>
<td>189.463</td>
</tr>
<tr>
<td></td>
<td>[113.722]</td>
<td>[113.722]</td>
<td>[323.112]</td>
<td>[124.084]</td>
<td>[129.167]</td>
<td>[84.769]</td>
<td>[146.072]</td>
<td>[139.604]</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>152.410</td>
<td>152.410</td>
<td>227.504</td>
<td>-251.322</td>
<td>-258.972</td>
<td>-170.725</td>
<td>-291.337</td>
<td>-320.007</td>
</tr>
<tr>
<td></td>
<td>[39.410]</td>
<td>[39.410]</td>
<td>[120.099]</td>
<td>[131.489]</td>
<td>[139.646]</td>
<td>[47.550]</td>
<td>[100.835]</td>
<td>[135.343]</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>-5.548</td>
<td>-5.685</td>
<td>-2.375</td>
<td>-3.954</td>
<td>-5.053</td>
</tr>
<tr>
<td></td>
<td>[2.702]</td>
<td>[2.702]</td>
<td>[6.876]</td>
<td>[3.570]</td>
<td>[3.753]</td>
<td>[1.694]</td>
<td>[3.151]</td>
<td>[3.726]</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>0.5</td>
<td>0.480</td>
<td>0.421</td>
<td>0.453</td>
<td>0.458</td>
<td>0.536</td>
<td>0.575</td>
<td>0.497</td>
</tr>
<tr>
<td></td>
<td>[0.1]</td>
<td>[0.052]</td>
<td>[0.070]</td>
<td>[0.048]</td>
<td>[0.048]</td>
<td>[0.051]</td>
<td>[0.054]</td>
<td>[0.046]</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>0.3</td>
<td>0.299</td>
<td>0.600</td>
<td>1.065</td>
<td>1.053</td>
<td>1.441</td>
<td>1.195</td>
<td>1.105</td>
</tr>
<tr>
<td></td>
<td>[0.3]</td>
<td>[0.255]</td>
<td>[0.216]</td>
<td>[0.078]</td>
<td>[0.077]</td>
<td>[0.151]</td>
<td>[0.153]</td>
<td>[0.086]</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>1.6</td>
<td>1.595</td>
<td>1.569</td>
<td>1.406</td>
<td>1.441</td>
<td>2.121</td>
<td>2.157</td>
<td>1.495</td>
</tr>
<tr>
<td></td>
<td>[0.6]</td>
<td>[0.555]</td>
<td>[0.575]</td>
<td>[0.570]</td>
<td>[0.569]</td>
<td>[0.559]</td>
<td>[0.562]</td>
<td>[0.568]</td>
</tr>
<tr>
<td>MAE</td>
<td>0.122</td>
<td>0.122</td>
<td>0.198</td>
<td>0.249</td>
<td>0.260</td>
<td>0.120</td>
<td>0.212</td>
<td>0.247</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.935</td>
<td>0.935</td>
<td>0.784</td>
<td>0.683</td>
<td>0.657</td>
<td>0.941</td>
<td>0.811</td>
<td>0.724</td>
</tr>
</tbody>
</table>

The table presents the results from Table 3 of Yogo (2006), from our replication and from several modifications of the replication. $\mu_1$, $\mu_2$, $\mu_3$, and mean absolute pricing errors (MAE) are expressed in percentage points. HAC standard errors are in brackets.
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.

3.2.2 GMM: First stage

We are going to discuss the results from the first stage of the GMM (Table 2) 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 $\mu$ and $\lambda$ 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 $\lambda_2$ remains statistically significant throughout the cases, it switches sign, which challenges the economic interpretation of the estimated coefficient. As discussed at the end of Section 2.1, such a sign switch upon a slight change of the weighting matrix can occur when the explanatory power of the factor stems from misestimation of the factor mean. 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 $\hat{\Omega}_{1,\ldots,25}$. 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 3.2.1, 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.

### 3.2.3 GMM: Second stage

Next, we analyze the robustness of the results from the second stage of the GMM estimation (Table 3). 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 $\lambda_3$ is always greater than 1, which implies a negative intertemporal elasticity of substitution in the theoretical 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.
The table presents the results from Table 3 of Yogo (2006), from our replication and from several modifications of the replication. $\mu_1$, $\mu_2$, and $\mu_3$ are expressed in percentage points. HAC standard errors are in brackets.

<table>
<thead>
<tr>
<th></th>
<th>Yogo</th>
<th>Replica</th>
<th>25 portf</th>
<th>nonpara</th>
<th>25 portf &amp; nonpara</th>
<th>Replica</th>
<th>25 portf</th>
<th>nonpara</th>
<th>25 portf &amp; nonpara</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>17.898</td>
<td>17.899</td>
<td>92.025</td>
<td>-44.590</td>
<td>-43.977</td>
<td>-60.167</td>
<td>-66.300</td>
<td>-60.905</td>
<td>-49.488</td>
</tr>
<tr>
<td></td>
<td>[31.280]</td>
<td>[31.280]</td>
<td>[56.000]</td>
<td>[38.454]</td>
<td>[38.714]</td>
<td>[24.788]</td>
<td>[30.996]</td>
<td>[41.054]</td>
<td>[40.081]</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>170.569</td>
<td>170.569</td>
<td>259.283</td>
<td>74.121</td>
<td>71.670</td>
<td>141.629</td>
<td>151.692</td>
<td>59.016</td>
<td>69.862</td>
</tr>
<tr>
<td></td>
<td>[15.561]</td>
<td>[15.561]</td>
<td>[35.139]</td>
<td>[30.003]</td>
<td>[50.566]</td>
<td>[19.751]</td>
<td>[22.148]</td>
<td>[32.477]</td>
<td>[31.786]</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>0.659</td>
<td>0.659</td>
<td>1.000</td>
<td>0.998</td>
<td>0.998</td>
<td>2.523</td>
<td>3.194</td>
<td>2.514</td>
<td>2.278</td>
</tr>
<tr>
<td></td>
<td>[0.849]</td>
<td>[0.849]</td>
<td>[1.624]</td>
<td>[1.087]</td>
<td>[1.093]</td>
<td>[0.605]</td>
<td>[0.804]</td>
<td>[1.216]</td>
<td>[1.195]</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>0.5</td>
<td>0.533</td>
<td>0.454</td>
<td>0.526</td>
<td>0.526</td>
<td>0.511</td>
<td>0.520</td>
<td>0.513</td>
<td>0.511</td>
</tr>
<tr>
<td></td>
<td>[0.0]</td>
<td>[0.040]</td>
<td>[0.041]</td>
<td>[0.043]</td>
<td>[0.043]</td>
<td>[0.034]</td>
<td>[0.034]</td>
<td>[0.043]</td>
<td>[0.043]</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>0.3</td>
<td>0.278</td>
<td>0.587</td>
<td>0.746</td>
<td>0.751</td>
<td>0.418</td>
<td>0.642</td>
<td>0.762</td>
<td>0.766</td>
</tr>
<tr>
<td></td>
<td>[0.14]</td>
<td>[0.073]</td>
<td>[0.056]</td>
<td>[0.056]</td>
<td>[0.056]</td>
<td>[0.094]</td>
<td>[0.079]</td>
<td>[0.056]</td>
<td>[0.056]</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>1.6</td>
<td>1.637</td>
<td>1.834</td>
<td>2.053</td>
<td>2.048</td>
<td>1.596</td>
<td>1.671</td>
<td>2.085</td>
<td>2.049</td>
</tr>
<tr>
<td></td>
<td>[0.5]</td>
<td>[0.473]</td>
<td>[0.506]</td>
<td>[0.516]</td>
<td>[0.516]</td>
<td>[0.400]</td>
<td>[0.438]</td>
<td>[0.522]</td>
<td>[0.520]</td>
</tr>
<tr>
<td>$J$</td>
<td>23.170</td>
<td>23.170</td>
<td>35.515</td>
<td>254.120</td>
<td>255.947</td>
<td>21.376</td>
<td>25.200</td>
<td>179.853</td>
<td>177.939</td>
</tr>
<tr>
<td>$p$-val</td>
<td>0.392</td>
<td>0.392</td>
<td>0.034</td>
<td>0.000</td>
<td>0.000</td>
<td>0.498</td>
<td>0.288</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
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. 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.

### 3.3 The impact of the weighting matrix

Our previous analysis suggests that the choice of the (initial) weighting matrix has a major impact on the inference and the parameter estimates. We now take a closer look at this relation. As a starting point we use the same estimation routine as described in Section 3.2.1, but use all 25 portfolios in Step 1 and use the nonparametric covariance matrix estimator, i.e., the case reported in the fifth column in Table 2. We then replace the square matrix $\hat{\Omega}_{26,\ldots,28}^{-1}$ in the lower right block of $W_1$ (Step 3 in Section 3.2.1) by the matrix $w \cdot (\hat{\Omega}_{26,\ldots,28})^{-1}$, with a scalar $w \in \mathbb{R}_+$. Note that we exclusively manipulate the weighting matrix for the first stage of the GMM estimation.

We vary $w$ from $10^{-4}$ to $10^4$ and report the results in Table 4. The pricing performance varies dramatically with $w$. With $w = 10^{-4}$, the pricing error is only 0.028% and the $R^2$ is

---

10This pattern has, for instance, been described in Cochrane (2005), p. 226.
Table 4: Parameter estimates for different weighting matrices

<table>
<thead>
<tr>
<th>Parameter Estimates for Different Weighting Matrices</th>
</tr>
</thead>
<tbody>
<tr>
<td>w = 10^{-4}</td>
</tr>
<tr>
<td>λ₁</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>λ₃</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>μ₁</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>μ₂</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>μ₃</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>MAE</td>
</tr>
<tr>
<td>R²</td>
</tr>
</tbody>
</table>

First stage results

Second stage results

The table presents the results from our replication when the weighting matrix is manipulated. We start from the case reported in the fifth column in Table 2 and then multiply the weights on the moment conditions identifying the factor means by a scalar $w$. The last column reports results from 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.
0.997. Moreover, the first stage estimates of the factor means differ dramatically from the sample averages because the weight on these moment conditions is very low. Increasing $w$ leads to an increase in mean absolute pricing errors and a decrease in $R^2$. The second column ($w = 1$) is identical to the fifth column in Table 2. For $w = 100$, the $R^2$ is already down to 0.034, and for $w = 10^4$ we have only 0.01 left. 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 4. The values for $\mu_1$, $\mu_2$, $\mu_3$ reported in this column are just the sample averages of the factors. The $R^2$ is very close to the one for $w = 10^4$. There are no degrees of freedom for the factor means in a Fama-MacBeth regression, which is similar to the case $w = 10^4$ in which the additional parameters $\mu_i$ are also basically fixed to the sample averages of the factors.

4 Conclusion

Standard GMM cross-sectional asset pricing tests can generate spuriously high explanatory power for factor models by allowing the estimated factor means to substantially deviate from the observed sample averages. 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. For instance, in terms of the motivating example presented in the introduction, even the population growth of captive Asian elephants can explain the cross-section of expected stock returns as long as a small weight is put on matching the mean of this “factor”. The seemingly high pricing performance comes along with a highly significant estimate of the market price of risk, a typical type I error.

While there is a large literature that documents similar phenomena for weak factors, our mechanism also affects strong pricing factors, that is factors that are strongly related to returns in the time series, but whose factor exposures are not necessarily related to average returns
in the cross-section. We even show that strong factors that do have some explanatory power for the cross-section of expected returns can be “crowded out” by spurious 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 general, though, our findings apply to all sample sizes, all factors, and all sets of test assets.

The most prominent symptom of the described effect is that the estimate of the mean of the spurious factor is very different from the sample average. A technical conclusion from our paper is thus that researchers using the estimation strategy that we discuss should not only report the estimated market prices of risk and the cross-sectional $R^2$, but also the estimated factor means. Against the backdrop of the current scientific debate that is concerned with taming the “zoo of pricing factors”, our paper makes an important contribution to the empirical asset pricing literature in that it points out a fundamental flaw in cross-sectional asset pricing tests that may lead researchers to think that a model has explanatory power when it actually does not. Instead, statements claiming success at the cross-sectional front sometimes have to be taken with a certain dose of skepticism since they may be the result of sacrificing the fit to fundamentals.
A Fama-MacBeth Regressions

To better understand the theory behind our GMM result, it is instructive to formulate its implications for Fama-MacBeth two-stage regressions. The first stage involves time series regressions of excess returns on the factor to estimate the betas for all assets. In the second stage, average excess returns are regressed on these betas to estimate the market price of risk from the cross-section of average returns.

The beta of an asset with respect to the factor $F$ is given as

$$\beta_i^F = \frac{\text{Cov}[R_i^e, F]}{\text{Var}[F]} = \frac{E[R_i^e(F - E[F])]}{\text{Var}[F]} = \frac{E[R_i^e(F - E[F])]}{E[(F - E[F])^2]}$$

which is estimated via the sample analogues as

$$\hat{\beta}_i^F = \frac{E_T[R_i^e(F - E_T[F])]}{E_T[(F - E_T[F])^2]}.$$  

An econometrician who sets the mean of the factor to an arbitrary value $\mu_F$ instead of estimating it from the data will obtain the following beta:

$$\hat{\beta}_i^F = \frac{E_T[R_i^e(F - E_T[F])]}{E_T[(F - E_T[F])^2]} = \frac{\text{Cov}_T[R_i^e, F - \mu_F] + E_T[R_i^e]E_T[(F - \mu_F)]}{E_T[(F - \mu_F)^2]}$$

$$= \frac{\text{Cov}_T[R_i^e, F]}{E_T[(F - \mu_F)^2]} + \frac{E_T[(F - \mu_F)]}{E_T[(F - \mu_F)^2]}E_T[R_i^e]. \quad (9)$$

The first term in this representation is roughly equal to the true factor beta: it is the sample covariance between excess returns and the factor, scaled by a term that is roughly equal to the sample variance of the factor. However, this true beta is confounded by a second term, which is equal to the asset’s average excess return, multiplied by a constant which does not depend on $i$.

Again, consider the case of a weak factor first. Assuming that the sample covariance $\text{Cov}_T[R_i^e, F]$ is equal to zero for all $i \in \{1, \ldots, n\}$, only the second term remains, i.e., the
estimated betas are multiples of the average excess returns. Given this, the second-pass cross-sectional regression

$$E_T[R^e_i] = \ell_0 + \ell_1 \hat{\beta}^F_i + \varepsilon_i$$

can be rewritten as

$$E_T[R^e_i] = \ell_0 + \ell_1 \frac{E_T[F - \mu_F]}{E_T[(F - \mu_F)^2]} E_T[R^e_i] + \varepsilon_i.$$  

This regression obviously delivers $\ell_0 = 0$ and $\ell_1 = \frac{E_T[(F - \mu_F)^2]}{E_T[F - \mu_F]}$ with a cross-sectional $R^2$ of 1, i.e. a perfect explanatory power of the (misspecified) model. The coefficient $\ell_1$ is related to the GMM estimate via $\hat{\lambda} = \frac{\ell_1}{\text{Var}_T[F]}$. Since $E_T[(F - \mu_F)^2]$ is the (incorrect) estimate of the variance of $F$ in this context, we again obtain the result $\hat{\lambda} = (E_T[F] - \mu_F)^{-1}$.

A special case of this setup would be to set $\mu_F = 0$, while the sample mean $E_T[F]$ is nonzero. Then the estimate of beta would be

$$\hat{\beta}^F_i = \frac{E_T[R^e_i(F - \mu_F)]}{E_T[(F - \mu_F)^2]} = \frac{\text{Cov}_T[R^e_i, F - \mu_F] + E_T[R^e_i]E_T[(F - \mu_F)]}{E_T[(F - \mu_F)^2]} = \frac{E_T[F]}{E_T[F^2]} E_T[R^e_i],$$

and this is just the coefficient from a first-pass regression of $R^e_i$ on $F$ without intercept. In a similar vein, for $\mu_F \neq 0$, the estimated beta would equal the coefficient from a first-pass regression of $R^e_i$ on $F$ with an exogenously fixed nonzero intercept. On the other hand, if we choose the free parameter $\mu_F$ to be equal to the sample mean, the second term in Equation (9) vanishes and all betas are (correctly) estimated to be zero. To sum up, in case of a weak factor, the GMM estimator discussed above can be translated into a Fama-MacBeth two-stage regression with an intercept in the first stage that is different from the intercept in a standard OLS regression.

Now consider the case of a strong factor. Then the first term in Equation (9) is nonzero,
but the estimated beta is still distorted by the second term. The second-pass regression becomes

\[ E_T[R_t^c] = \ell_0 + \ell_1 \left( \frac{\text{Cov}_T[R_t^c, F]}{E_T[(F - \mu_F)^2]} + \frac{E_T[(F - \mu_F)]}{E_T[(F - \mu_F)^2]} E_T[R_t^e] \right) + \varepsilon_t. \]

If the free parameter \( \mu_F \) is allowed to take any arbitrary value, then, to minimize the pricing error from this second-pass regression, \( \mu_F \) must differ from \( E_T[F] \) as much as possible, so that the estimated beta is dominated by the second term. Note that \( E_T[F - \mu_F]/E_T[(F - \mu_F)^2] \) also goes to zero when \( F - \mu_F \) is getting large, but the convergence speed is much slower than for the first term, since \( \text{Cov}_T[R_t^c, F] \) is a fixed number that does not depend on \( F - \mu_F \). Accordingly, with \( \ell_1 = \frac{E_T[(F - \mu_F)^2]}{E_T[F - \mu_F]} \), the cross-sectional \( R^2 \) is close to 1, for the same reasoning as above. If on the other hand the estimate of \( \mu_F \) is exogenously set close to the sample mean, the second term is small and the estimated beta is close to the true beta. If \( \mu_F = E_T[F] \), again all betas are correctly estimated and the cross-sectional regression delivers the desired outcome.
References


