Empirical Evaluation of Overspecified Asset Pricing Models*

Elena Manresa
New York University, 19 West 4th St, New York, NY 10012, USA
<elena.manresa@nyu.edu>

Francisco Peñaranda
Queens College CUNY, 65-30 Kissena Blvd, Flushing, NY 11367, USA
<francisco.penaranda@qc.cuny.edu>

Enrique Sentana
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<sentana@cemfi.es>

June 2014, revised June 2020

Abstract

Empirical asset pricing models with possibly unnecessary risk factors are increasingly common. Unfortunately, they can yield misleading statistical inferences. Unlike previous studies, we estimate the identified set of SDFs and risk prices compatible with a given model's asset pricing restrictions. We also propose tests that detect problematic situations with economically meaningless SDFs uncorrelated to the test assets. Empirically, we estimate linear subspaces of SDFs compatible with popular extensions of the traditional and consumption versions of the CAPM, which are typically two-dimensional. Moreover, we often find that all the SDFs in those linear spaces are uncorrelated with the test assets' returns.

Keywords: Continuously Updated GMM, Factor pricing models, Set estimation, Stochastic discount factor, Underidentification tests.

JEL: G12, C52.

*We would like to thank Craig Burnside, Cesare Robotti and Motohiro Yogo for kindly making their data available to us. We also thank Cesare Robotti, Enrique Salvador, Dacheng Xiu, participants at the Finance Forum (Zaragoza), the Midwest Finance Association Meeting (Chicago), the UK MMF Workshop on "Empirical Modelling of Financial Markets" (Brunel), the annual SoFiE Conference (Aarhus), the Computational and Financial Econometrics Conferences (Seville and London), the SanFI "New Methods for the Empirical Analysis of Financial Markets" Conference (Comillas), the World Congress (Montreal) and Asian Meetings (Xiamen) of the Econometric Society, the Spanish Economic Association Symposium (Alicante), the Virtual Workshop on Financial Econometrics (Durham), as well as audiences at CEU Cardenal Herrera (Elche), Chicago, Fundação Getulio Vargas (Rio), Geneva, Harvard Business School, LSE, McCombs School of Business (Austin), Queens College CUNY and Yale for helpful comments and suggestions. Of course, the usual caveat applies. Financial support from the Spanish Government through grant 2014-59262 (Sentana) is gratefully acknowledged. The second and third authors are also very grateful to SanFI and the Santander CEMFI Research Chair in Finance, respectively, for their funding.

1 Introduction

The most popular empirically oriented asset pricing models effectively assume the existence of a common stochastic discount factor (SDF) that is linear in some risk factors, which discounts uncertain payoffs differently across different states of the world. Those factors can be either the returns on some traded securities, non-traded economy wide sources of uncertainty related to macroeconomic variables, or a combination of the two. The empirical success of such models at explaining the so called CAPM anomalies was initially limited, but researchers have progressively entertained a broader and broader set of factors, which has resulted in several success claims. Harvey, Liu and Zhu (2016) contains a comprehensive list of references, cataloguing 315(!) different factors.

However, several authors have warned that some of those factors, or more generally linear combinations of them, could be uncorrelated with the vector of asset payoffs that they are meant to price, which would result in economically meaningless models (see Burnside (2016), Gospodinov, Kan, and Robotti (2017, 2019), Kleibergen and Zhan (2020) and the references therein). Further, those papers forcefully argue that such situations can lead to misleading econometric conclusions.

In this context, the purpose of our paper is to study the estimation of prices of risk and the testing of the cross-sectional restrictions imposed by overspecified linear factor pricing models. By overspecified models we mean those with at least one non-trivial SDF which is uncorrelated with the excess returns on the vector of test assets. We discuss in detail several textbook examples of this situation, which illustrate two important differences between our work and related studies. First, the presence of uncorrelated risk factors is sufficient but not necessary for overspecification. As a result, attempts to find out which factors are uncorrelated on an individual basis fail to provide a complete answer. Second, overspecification is necessary but not sufficient for the model parameters to be underidentified. Therefore, studying parameter identification by means of rank tests does not provide a full answer either.

Our point of departure from the existing literature is that we do not focus exclusively on the properties of the usual estimators and tests. Instead, we use the econometric framework in Arellano, Hansen and Sentana (2012).[1] Thus, we can identify a linear subspace of risk prices compatible with the cross-sectional asset pricing restrictions, a basis of which we can easily parametrize and efficiently estimate using standard GMM methods. When the dimension of the subspace is two or more, our approach effectively explores whether two or more asset pricing models simultaneously hold. For example, in the context of a linearized version of a model à la

[1] This use of their econometric framework in asset pricing was first explored by Manresa (2009).

Epstein and Zin (1989), we might consider whether the (linearized) CAPM and Consumption CAPM simultaneously hold, which would indicate that the risk prices of the empirical model that uses both factors would be underidentified.

We follow Peñaranda and Sentana (2015) in using single-step procedures, such as the continuously updated GMM estimator (CU-GMM) of Hansen, Heaton and Yaron (1996), to obtain numerically identical test statistics and risk price estimates for SDF and regression methods, with uncentred or centred moments, and symmetric or asymmetric normalizations. GEL methods such as Empirical Likelihood or Exponentially Tilted also share the numerical invariance properties of CU-GMM. However, given that these methods are often more difficult to compute than two-step estimators, and they may sometimes give rise to multiple local minima, in the appendix we propose simple, intuitive consistent parameter estimators that can be used as sensible initial values, and which will be efficient for elliptically distributed returns and factors. Interestingly, we show that these consistent initial values coincide with the GMM estimators recommended by Hansen and Jagannathan (1997), which use the second moment of returns as weighting matrix.

For simplicity of exposition, we initially focus on excess returns, but later on extend our analysis to cover gross returns too. Importantly, we show that single-step GMM procedures yield the same numerical results with both types of payoffs.

In addition to the usual overidentification test, which is informative about the existence of admissible SDFs, we propose simple tests that can diagnose economically meaningless but empirically relevant cases in which the expected values of all SDFs in the identified set are 0, which is equivalent to their being uncorrelated to the test assets. We refer to this situation as complete overspecification, which should not apply to credible empirical models. In addition, we explicitly relate our tests to the rank tests in the literature.

In our first empirical application, we investigate the potential overspecification of the three-factor consumption CAPM in Yogo (2006) using quarterly data from the popular Fama and French cross-section of excess returns on size and book-to-market sorted portfolios. Aside from its undisputable influence on the subsequent literature, an important characteristic of his model is that the chosen risk factors were theoretically motivated and not the result of either an extensive search or a reverse engineering process.

Nevertheless, the results we obtain with our novel inference procedures indicate that the admissible SDFs in the linearized version of this model lie on a two-dimensional subspace, so there is lack of identification, a situation that standard GMM cannot cope with. In addition, we cannot reject the null hypothesis that all those SDFs have zero means, which is tantamount to complete overspecification. These results are robust across several sample periods and cross-sections of stock returns. Importantly, our simulations suggest that these empirical findings are not due to lack of power. On the contrary, if anything, our proposed tests tend to overreject for the sample size of the data sets we use.

Our second empirical application evaluates the CAPM extension in Jagannathan and Wang (1996) to determine whether overspecification affects models without consumption risk factors too. Once again, we find evidence of underidentification when we use monthly returns on the size and book-to-market sorted portfolios. Furthermore, the identified two-dimensional space of valid SDFs is completely overspecified. However, we achieve the identification of a one-dimensional set of admissible SDFs if we add industry porfolios to the size and value sorted portfolios. Unfortunately, the identified set consists of scaled versions of an SDF that is uncorrelated with the extended cross-section of returns, and thereby economically unattractive.

Finally, we also apply our methodology to the Fama and French (1993) three factor model, whose pricing factors are the market portfolio and two portfolios that aim to capture the size and value effects. We find that the problem with this model is neither overspecification nor underidentification, but rather misspecification because its pricing errors are not zero.

The rest of the paper is organized as follows. Section 2 introduces linear factor pricing models, and characterizes their potential overspecification and underidentification by means of texbook asset pricing examples. Next, we present our econometric methodology in section 3, and compare it to existing approaches in section 4. Then, we apply our methods to some popular empirical asset pricing models in section 5. Finally, we summarize our conclusions and discuss some avenues for further research in section 6. Proofs of some formal results are presented in appendix A. We also include several supplemental appendices with some additional material: appendix B provides a geometrical interpretation of admissible SDFs sets, appendix C offers additional details on normalizations and starting values, and appendix D contains the results of our Monte Carlo experiments.

2 Overspecified Asset Pricing Models

2.1 Stochastic discount factors and moment conditions

Let $\mathbf{r}$ be a given $n \times 1$ vector of excess returns, whose means $E(\mathbf{r})$ we assume are not all equal to zero. Standard arguments such as lack of arbitrage opportunities or the first order conditions of a representative investor imply that

$$E(m\mathbf{r}) = \mathbf{0}$$

for some random variable m called SDF, which discounts uncertain payoffs in such a way that their expected discounted value equals their cost.

The standard approach in empirical finance is to model the SDF as an affine transformation of some $k < n$ observable risk factors $\mathbf{f}$.[2] In particular, researchers typically express the pricing equation as

$$E[(a + \mathbf{b}'\mathbf{f})\mathbf{r}] = \mathbf{0} \tag{1}$$

for some coefficients $(a, \mathbf{b})$, which we can refer to as the intercept and slopes of the affine SDF $m = a + \mathbf{b}'\mathbf{f}$.

We can also estimate the SDF mean $c = E(m)$ by adding the moment condition

$$E(a + \mathbf{b}'\mathbf{f} - c) = 0, \tag{2}$$

which pins down c for given $(a, \mathbf{b})$. As we will see in Section 3.2, the SDF mean plays a crucial role in testing for overspecification.

A non-trivial advantage of this approach is that (1) and (2) are linear in $(a, \mathbf{b}, c)$. Therefore, when there exist admissible parameter configurations other than the trivial one $(a, \mathbf{b}, c) = (0, \mathbf{0}, 0)$, we can at best identify a direction in $(a, \mathbf{b}, c)$ space, which leaves both the scale and sign of the SDF undetermined. One popular possibility would directly estimate the prices of risk $\delta = -\mathbf{b}/a$, which effectively fixes the SDF intercept to 1. Nevertheless, given that any asymmetric normalization is potentially restrictive, we prefer to use invariant estimation methods, such as CU-GMM (see Appendix C for further details).

In what follows, we consider models in which the elements of $\mathbf{f}$ are either non-traded or treated as such. In those cases, the pricing conditions (1) and (2) contain all the relevant information to estimate and test the asset pricing model. Nevertheless, it would be very easy to extend our analysis to explicitly deal with traded factors whose excess returns do not belong to the linear span of $\mathbf{r}$. In that case, we should add moment conditions such as

$$E[(a + \mathbf{b}'\mathbf{f})\mathbf{f}] = \mathbf{0}$$

to (1) and (2) to complete the asset pricing information that we should consider, as Lewellen, Nagel and Shanken (2010) suggest.

In the next section, we illustrate with some textbook asset pricing examples the different issues mentioned in the introduction that may affect an empirical SDF model such as (1).

[2] This ignores that m must be positive with probability 1 to avoid arbitrage opportunities, which would require non-linear specifications for m (see Hansen and Jagannathan (1991)).

2.2 A taxonomy of overspecification

For pedagogical purposes, let us begin by assuming that risk premia are given by $E(\mathbf{r}) = \boldsymbol{\sigma}_p\tau_p$, where τ_p captures the market price of risk and $\boldsymbol{\sigma}_p$ contains the covariances between the vector of excess returns and the single risk factor f_p, say the market portfolio, so that the CAPM holds. However, imagine an empirical researcher considers the following linearized version of the Consumption CAPM à la Epstein and Zin (1999):

$$m = a + b_p f_p + b_c f_c, \tag{3}$$

where f_c denotes consumption growth. The pricing errors of this empirical model would be

$$E\left(m\mathbf{r}\right) = \boldsymbol{\sigma}_p[\tau_p(a + \mu_p b_p + \mu_c b_c) + b_p] + \boldsymbol{\sigma}_c b_c, \tag{4}$$

where μ_p and μ_c denote the population means of the empirical factors and $\boldsymbol{\sigma}_c$ the vector of covariances between excess returns and consumption growth.

Given that the empirical model nests the true one, the CAPM solution $b_p = -a(1+\tau_p\mu_p)^{-1}\tau_p$ and $b_c = 0$ will trivially make these pricing errors zero regardless of the value of $\boldsymbol{\sigma}_c$. However, there will be (infinitely) many more solutions when $\boldsymbol{\sigma}_c = \boldsymbol{\sigma}_p\kappa_p$ so that the factor mimicking portfolios of f_c and f_p are proportional. An illustrative example of this situation arises in Breeden's (1979) consumption version of Merton's (1973) ICAPM: investors with log utility will optimally ignore changing investment opportunities, and consequently both the CCAPM and the traditional CAPM will give rise to the same risk premia. Obviously, the (linearised) empirical counterparts of these two models[3] will provide admissible SDFs (namely, $a_c\left[1 - (\kappa_p + \tau_p\mu_c)^{-1}\tau_p f_c\right]$ and $a_p\left[1 - (1 + \tau_p\mu_p)^{-1}\tau_p f_p\right]$), respectively), but there will be a continuum of other SDFs. In particular, defining $f_c^* = f_c - \kappa_p f_p$ and its mean $\mu_c^* = \mu_c - \kappa_p\mu_p$, the non-trivial SDFs that simply scale $f_c^* - \mu_c^*$ up or down will have zero covariance with the vector of excess returns $\mathbf{r}$. Therefore, the empirical model will be partially overspecified and econometrically underidentified.

Let us now consider a more general model in which risk premia depend on a second risk factor, f_s. One example would be a simplified version of the Intertemporal CAPM in which the wealth portfolio is equal to the market portfolio, and the default spread captures changes in state variables. In this context, the vector of risk premia will be given by

$$E(\mathbf{r}) = \boldsymbol{\sigma}_p\tau_p + \boldsymbol{\sigma}_s\tau_s, \tag{5}$$

where τ_s represents the price of risk of the second risk factor while $\boldsymbol{\sigma}_s$ contains the covariances between this factor and the vector of excess returns, which are such that $rank(\boldsymbol{\sigma}_p, \boldsymbol{\sigma}_s) = 2$. In

[3] To avoid collinearity between f_c and f_p, one can realistically assume that the consumption growth proxy includes measurement error uncorrelated to the vector of returns.

this case, the pricing errors of the empirical model (3) would be

$$E\left(m\mathbf{r}\right) = \boldsymbol{\sigma}_p[\tau_p(a + \mu_p b_p + \mu_c b_c) + b_p] + \boldsymbol{\sigma}_s\tau_s(a + \mu_p b_p + \mu_c b_c) + \boldsymbol{\sigma}_c b_c. \qquad (6)$$

Therefore, the moment conditions (1) will not be satisfied unless $\boldsymbol{\sigma}_c = \boldsymbol{\sigma}_p\kappa_p + \boldsymbol{\sigma}_s\kappa_s$. Intuitively, this condition requires that the factor mimicking portfolio of f_c is spanned by the factor mimicking portfolios of the true factors f_p and f_s.[4]

In this context, the value of κ_s makes a big difference. If $\kappa_s \neq 0$, the moment conditions (1) will be satisfied because the SDF specification in (3) gives rise to an admissible empirical model perfectly compatible with the risk premia in (5).

Things are rather different when $\kappa_s = 0$. Substituting $\boldsymbol{\sigma}_c = \boldsymbol{\sigma}_p\kappa_p$ into the pricing errors of the empirical model (6) immediately shows that the unique (up to scale) solution of the resulting system of linear equations will satisfy $b_p + \kappa_p b_c = 0$ and $a + b_c\mu_c^* = 0$. Thus, the admissible empirical SDFs (3) will be proportional to $f_c^* - \mu_c^*$, in marked contrast with the true model (5). This example provides a useful generalisation of the useless factor example put forward by Kan and Zhang (1999) among others, who implicitly assume that $\kappa_s = \kappa_p = 0$ so that $\boldsymbol{\sigma}_c = \mathbf{0}$. In particular, it implies that an empirical asset pricing model can be economically meaningless, in the sense that it generates uncorrelated SDFs, even though all its risk factors are correlated with the vector of excess returns and the (normalised) prices of risk are econometrically point identified.

Finally, we could have complete overspecification if the empirical researcher uses two factors, say f_c and f_d, which have zero covariances with the vector of excess returns $\mathbf{r}$. For example, she could use non-durable consumption growth together with durable consumption growth, as in Eichenbaum and Hansen (1989). In this case, the prices of risk will not be point identified either, and all admissible stochastic discount factors, which are linear combinations of $f_c - \mu_c$ and $f_d - \mu_d$, will have 0 covariance with the vector of excess returns. Similar problems arise in empirical models that use f_p, f_c and f_d as candidate risk factors.

[4] This condition nests Statement 1 in Lewellen, Nagel, and Shanken (2010), which says that the empirical model yields zero pricing errors if its factors are uncorrelated with the residual of the projection of the vector of returns onto the true factors. In our setting, one of the true factors already appears in the empirical model, so the Lewellen, Nagel, and Shanken (2010) condition simply requires that the projection residual and f_c be uncorrelated, namely $Cov(\mathbf{r} - \boldsymbol{\alpha} - \boldsymbol{\beta}_p f_p - \boldsymbol{\beta}_s f_s, f_c) = \mathbf{0}$, or equivalently $\sigma_{cc} - \beta_p\sigma_{pc} - \beta_s\sigma_{sc} = 0$. Given that $(\beta_p, \beta_s) = (\sigma_p, \sigma_s)\mathbf{V}^{-1}$, where $\mathbf{V}$ is the covariance matrix of the true factors f_p and f_s, we can write $\boldsymbol{\sigma}_c = \boldsymbol{\sigma}_p\kappa_p + \boldsymbol{\sigma}_s\kappa_s$ with (κ_p, κ_s) being the projection coefficients of f_c onto the true factors.

3 Econometric methodology

3.1 Set estimation

The pricing conditions (1) can be written in matrix notation as

$$\left[\begin{array}{cc} E(\mathbf{r}) & E(\mathbf{r}\mathbf{f}') \end{array}\right] \begin{pmatrix} a \\ \mathbf{b} \end{pmatrix} = \mathbf{M}\boldsymbol{\theta} = \mathbf{0}, \tag{7}$$

where $\mathbf{M}$ is an $n \times (k+1)$ matrix of first and second moments of data and $\boldsymbol{\theta}$ a $(k+1) \times 1$ parameter vector. For example, in the empirical specification (3) discussed in the previous section, $\mathbf{M} = \left[\begin{array}{ccc} E(\mathbf{r}) & E(\mathbf{r}f_p) & E(\mathbf{r}f_c) \end{array}\right]$.

The highest possible rank of $\mathbf{M}$ is its number of columns $k+1$ because $k < n$. In that case, though, the asset pricing model will not hold because the only value of $\boldsymbol{\theta}$ that satisfies (7) will be the trivial solution $\boldsymbol{\theta} = \mathbf{0}$. A case in point arose in the previous section when the factor mimicking portfolio of f_c is not spanned by the factor mimicking portfolios of the true factors f_p and f_s, so that the pricing errors (6) are not zero when $\boldsymbol{\sigma}_c \neq \boldsymbol{\sigma}_p\kappa_p + \boldsymbol{\sigma}_s\kappa_s$.

On the other hand, if the rank of $\mathbf{M}$ is k, then there will be a one-dimensional subspace of $\boldsymbol{\theta}$'s that satisfy the pricing conditions (7), in which case the solution $\boldsymbol{\theta}$ is unique up to scale, as we explained in section 2.1. Not surprisingly, $rank(\mathbf{M}) = k$ coincides with the usual identification condition required for standard GMM inference (see e.g. Hansen (1982) and Newey and McFadden (1994)). As we mentioned in the previous paragraph, the empirical specification (3) will be compatible with the true model (5) when $\boldsymbol{\sigma}_c = \boldsymbol{\sigma}_p\kappa_p + \boldsymbol{\sigma}_s\kappa_s$, although with different empirical and econometric implications depending on whether $\kappa_s = 0$; see section 3.2 for tests specifically aimed at detecting completely overspecified cases.

In the previous section, though, we also encountered an underidentified model with log utility investors in which both the CCAPM and the traditional CAPM hold. As a result, it is of the utmost importance to use statistical inference tools that can successfully deal with situations in which $rank(\mathbf{M}) < k$.

Following Arellano, Hansen and Sentana (2012), we begin by specifying the dimension of the subspace of solutions to the pricing conditions (7), which we denote d, so that $rank(\mathbf{M}) = (k+1) - d$. Given that we maintain the hypothesis that $E(\mathbf{r}) \neq \mathbf{0}$, we could in principle consider ranks for $\mathbf{M}$ as low as 1 or, equivalently, any positive integer d up to a maximum value of k. The situation discussed at the end of the previous section, where the two empirical factors f_c and f_d have zero covariances with the vector of excess returns $\mathbf{r}$, provides an example of $rank(\mathbf{M}) = 1$.

When $d = 1$ we can rely on standard GMM to estimate a unique $\boldsymbol{\theta}$ (up to normalization) and use its associated J test to assess the validity of the asset pricing restrictions. However, when $d \geq 2$, we will have a multidimensional subspace of admissible SDFs even after fixing their scale. Nevertheless, we can efficiently estimate a basis of that subspace by replicating d times the moment conditions (7) as follows:

$$\left.\begin{array}{c} [\ E(\mathbf{r}) \quad E(\mathbf{rf}')\ ]\boldsymbol{\theta}_1 = \mathbf{0}, \\ [\ E(\mathbf{r}) \quad E(\mathbf{rf}')\ ]\boldsymbol{\theta}_2 = \mathbf{0}, \\ \vdots \\ [\ E(\mathbf{r}) \quad E(\mathbf{rf}')\ ]\boldsymbol{\theta}_d = \mathbf{0}, \end{array}\right\} \tag{8}$$

and imposing enough exclusion restrictions and normalizations on $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_d)$ to ensure the point identification of a basis of the null space of $\mathbf{M}$. Importantly, those exclusion restrictions effectively lead to the simultaneous estimation of several restricted versions of the asset pricing model (1). For instance, in the Epstein-Zin example of section 2.2, one would jointly estimate the CAPM and the (linearized) CCAPM.

In this setting, the familiar J test from the work of Sargan (1958) and Hansen (1982) for overidentification of the augmented model becomes a test for "underidentification" of the original model. The rationale is as follows: if we can identify a linear subspace of risk prices without statistical rejection, then the original asset pricing model is not well identified. In contrast, a statistical rejection provides evidence that the prices of risk in the original model are indeed point identified, unless of course the familiar J test continues to reject its overidentifying restrictions.

We can also add moment conditions to estimate $(c_1, c_2, \ldots, c_d)$, which characterize the expected values of the basis SDF's. Specifically, we can combine (8) with the moment conditions

$$\left.\begin{array}{c} [\ 1 \quad E(\mathbf{f}')\ ]\boldsymbol{\theta}_1 - c_1 = 0, \\ [\ 1 \quad E(\mathbf{f}')\ ]\boldsymbol{\theta}_2 - c_2 = 0, \\ \vdots \\ [\ 1 \quad E(\mathbf{f}')\ ]\boldsymbol{\theta}_d - c_d = 0, \end{array}\right\} \tag{9}$$

which are exactly identified for given values of $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_d)$.

3.2 Testing restrictions on admissible SDF sets

As we have just seen, our inference framework allows us to estimate the set of SDFs that is compatible with the pricing conditions (1). But we can also use it to test if the elements of this set satisfy some relevant restrictions.

A particularly important null hypothesis that empirical researchers would like to find evidence against is that all SDFs compatible with the data have zero means, a situation we have termed "complete overspecification". The rationale is as follows. Given that

$$E(\mathbf{r}m) = E(\mathbf{r})E(m) + Cov(\mathbf{r}, m) = \mathbf{0},$$

zero pricing errors implies $Cov(\mathbf{r}, m) = \mathbf{0}$ when $c = E(m) = 0$. As a result, there will be no element in the admissible SDF set that explains the cross-section of expected returns from a meaningful economic perspective, as we illustrated in the last two paragraphs of section 2.2 for $d = 1$ and $d = 2$ respectively. Those two examples represent completely overspecified models in which all the SDFs in the corresponding admissible set are uncorrelated with the asset payoffs, which renders them economically uninteresting.

In any given sample, though, the estimated values of the means of the admissible SDFs will not be 0. Given that the SDF means are associated to the parameters $(c_1, c_2, ..., c_d)$ by virtue of (9), a distance metric (DM) test of $H_0 : c_i = 0,\ i = 1, \ldots, d$ will give us a valid test of the null hypothesis of complete overspecification. As is well known, a DM test simply compares the GMM criterion functions (J statistics) with and without those constraints. We can trivially compute the criterion function without the zero mean constraints from the system (8), or equivalently, from the joint system that also considers the exactly identified moment conditions (9). In turn, we can construct the criterion function that imposes the zero mean constraints on all the SDFs from the system

$$E\begin{bmatrix} \mathbf{r}(1\ \mathbf{f}')\boldsymbol{\theta}_i \\ (1\ \mathbf{f}')\boldsymbol{\theta}_i \end{bmatrix} = E\left[\mathbf{x}(1\ \mathbf{f}')\boldsymbol{\theta}_i\right] = 0, \qquad i = 1, 2, \ldots, d, \tag{10}$$

where $\mathbf{x}' = (\mathbf{r}', 1)$, which is analogous to (8) for an extended vector of payoffs that includes a fictional unit safe payoff.[5]

Another interesting null hypothesis that we may also want to test is whether some particular pricing factor does not appear in any admissible SDF. Formally, the corresponding null hypothesis would be that the entry of b associated to this factor were zero in all the vectors $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_d)$. In section 2.2 we came across one such an example: the true model is the CAPM but a researcher estimates the Epstein-Zin model and the covariance of consumption growth with the vector of returns is not proportional to the covariace of the market. Again, a DM test based on single-step GMM procedures will be ideally suited for testing this restriction on the space of admissible SDFs.[6]

[5]If there really existed an unconditionally safe asset, an SDF that satisfied $E(\mathbf{x}m) = \mathbf{0}$ would allow for arbitrage opportunities in the extended payoff space. Although no such an asset exists in real life, if all the SDFs in the admissible set satisfied the moment conditions (10), then we would have a clear indication of the problematic economic interpretation of a completely overspecified model.

[6]A convenient property of DM tests is that they are invariant to normalisation. But if we chose the asymmetric normalisations $a_i = 1$, the moment conditions (8) and (9) would become linear in parameters, and the results in

4 Comparison to the existing literature

4.1 Equivalent approaches

There are two alternative popular approaches to test asset pricing models. One uses $Cov(\mathbf{r}, \mathbf{f})$ instead of $E(\mathbf{rf}')$ in explaining the cross-section of risk premia, while the other one relies on the regression of $\mathbf{r}$ onto a constant and $\mathbf{f}$.

To relate our approach to the so-called centred SDF approach, let us express the pricing conditions (1) in terms of central moments. Specifically, we can add and subtract $\mathbf{b}'\boldsymbol{\mu}$ from $a + \mathbf{b}'\mathbf{f}$ and recognize $c = a + \mathbf{b}'\boldsymbol{\mu}$ as the expected value of the proposed SDF. This allows us to re-write the pricing conditions (1) as

$$E\left\{ \begin{array}{c} [c + \mathbf{b}'(\mathbf{f} - \boldsymbol{\mu})]\,\mathbf{r} \\ \mathbf{f} - \boldsymbol{\mu} \end{array} \right\} = \mathbf{0}. \tag{11}$$

Thus, the unknown parameters become $(c, \mathbf{b}, \boldsymbol{\mu})$ instead of $(a, \mathbf{b})$, as we have added k extra moments to identify $\boldsymbol{\mu}$.

Similarly, if we define $\mathbf{B} = Cov(\mathbf{r}, \mathbf{f})[Var(\mathbf{f})]^{-1}$ and $\boldsymbol{\lambda} = Var(\mathbf{f})\mathbf{b}$, then we can write the pricing conditions (11) in terms of the following moment conditions:

$$E\left[\begin{array}{c} c\mathbf{r} - \mathbf{B}\boldsymbol{\lambda} \\ vec\{[\mathbf{Bf} - \mathbf{r}](\mathbf{f} - \boldsymbol{\mu})'\} \\ \mathbf{f} - \boldsymbol{\mu} \end{array} \right] = \left(\begin{array}{c} \mathbf{0}_n \\ \mathbf{0}_{nk} \\ \mathbf{0}_k \end{array} \right), \tag{12}$$

where the vectorized moment conditions correspond to the usual least squares normal equations.

We can then follow the same approach as in section 3.1 by replicating the first block of moment condtions in (11) or (12) after imposing the necessary exclusion restrictions on the prices of the factors, together with some chosen normalisation restrictions. Likewise, we can adapt the testing procedures we described in section 3.2 to these centred SDF and regression moment conditions too.

In this context, it is straightforward to extend the results in Proposition 2 of Peñaranda and Sentana (2015) so as to prove that all three approaches provide numerically equivalent test statistics, prices of risk estimates and pricing errors when one uses single-step GMM procedures. From the computational point of view, though, the advantage of our uncentred SDF approach is that it requires the estimation of a lower number of parameters from a lower number of moments.

Newey and West (1987b) woud imply that the Wald, Lagrange Multiplier and DM tests of linear homogeneous restrictions such as $H_0 : c_i = 0$ would be numerically identical for two-step GMM methods that shared the same weighting matrices. More generally, DM tests might be more reliable than Wald tests in non-standard situations with potential identification failures (see Dufour (1997) for closely related results in a likelihood context).

4.2 Rank tests

Burnside (2016) studies the identification of the prices of risk of the linear factor pricing model (11) by applying the tests proposed by Cragg and Donald (1997) and Kleibergen and Paap (2006) to assess the rank of $Cov(\mathbf{r}, \mathbf{f})$, which coincide with the expected Jacobian matrices of those GMM conditions. More recently, Kleibergen and Zhan (2020) apply the rank tests in Kleibergen and Paap (2006) to the matrix of regression coefficients $\mathbf{B}$ that appears in (12).

In this respect, we can prove the following result:

Proposition 1 *The CU version of the overidentification test of the original SDF moment conditions (8) and (9) after imposing the d restrictions* $c_1 = \ldots = c_d = 0$ *numerically coincides with the CU version of the test of the null hypothesis* $H_0 : \text{rank}[Cov(\mathbf{r}, \mathbf{f})] = k - d$.

In fact, it is possible to use the results in Theorem 1 of Al-Sadoon (2017) to prove that under standard regularity conditions, this J test statistic converges in probability to the rank test statistics in Cragg and Donald (1996, 1997) and Kleibergen and Paap (1997) under both the null hypothesis and sequences of local alternatives.[7]

Proposition 1 also implies that the DM test of $c_1 = ... = c_d = 0$ we introduced in the previous section can be interpreted as a test of the null hypothesis that $\text{rank}[Cov(\mathbf{r}, \mathbf{f})] = k - d$ under the maintained hypothesis that $\text{rank}(\mathbf{M}) = (k + 1) - d$. In those circumstances, $E(\mathbf{r})$ could not be spanned by $Cov(\mathbf{r}, \mathbf{f})$. As a result, the only admissible SDFs would be those economically meaningless random variables that exploit the rank failure in $Cov(\mathbf{r}, \mathbf{f})$ in setting to zero the pricing conditions (1). In contrast, the test of the rank of $Cov(\mathbf{r}, \mathbf{f})$ in Proposition 1 is often uninformative about the existence of economically meaningful SDFs precisely because it does not maintain any hypothesis on the rank of $\mathbf{M}$.[8]

To illustrate this subtle difference, it is once again convenient to look at some of the textbook models we discussed in section 2.2. In particular, when the CAPM holds but an empirical researcher adds a second useless factor, or when both the CAPM and the consumption CAPM simultaneously hold, the matrix $Cov(\mathbf{r}, \mathbf{f})$ has rank 1 instead of 2 while $\mathbf{M}$ has rank 1 instead of 3. As a result, $E(\mathbf{r})$ belongs to the span of $Cov(\mathbf{r}, \mathbf{f})$, which confirms that in those two examples there exist economically meaningful SDFs that correctly price $\mathbf{r}$. Therefore, the fact that the rank test of $H_0 : \text{rank}[Cov(\mathbf{r}, \mathbf{f})]$ will not reject does not imply the inexistence of SDFs correlated with the excess returns on the test assets.

However, if the ICAPM holds but the factor mimicking portfolios for f_c and f_p are proportional, then $\text{rank}[Cov(\mathbf{r}, \mathbf{f})] = 1$ instead of 2 but $E(\mathbf{r})$ cannot be spanned by $Cov(\mathbf{r}, \mathbf{f})$ because

[7] See Arellano, Hansen and Sentana (1992) for an analogous result relating their underidentification test to the minimum distance test in Crag and Donald (1993) in linear IV models.

[8] The only exception is the extreme case of $Cov(\mathbf{r}, \mathbf{f}) = \mathbf{0}$, which necessarily means $\text{rank}(\mathbf{M}) = 1$ when $E(\mathbf{r}) \neq \mathbf{0}$, making it impossible to find meaningful SDFs that can explain $E(\mathbf{r})$, as we discussed at the end of section 2.2.

the rank of $\mathbf{M}$ is 2, so that the only admissible SDFs must be uncorrelated to the vector of excess returns $\mathbf{r}$.

In contrast, our econometric methodology allows us to estimate a basis of the identified linear subspace of admissible SDFs, which can then use to test if all of them are uncorrelated to the test assets.

4.3 Gross returns

So far, we have exclusively focused on excess returns, while some of the empirical asset pricing literature looks at gross returns. In the rest of this section, though, we shall prove that the distinction turns out to be irrelevant for single-step GMM methods.

Let $\mathbf{R}$ denote a vector of gross returns on $N = n + 1$ assets. Without loss of generality, we can understand the vector of excess returns $\mathbf{r}$ that we have used so far as the difference between the gross returns of the last n assets and the first one, R say. In practice, this reference asset could be the real return on US T-bills, whose payoffs are not constant. The relevant pricing equation for $\mathbf{R}$ becomes:

$$E[\mathbf{R}(a + \mathbf{b}'\mathbf{f})] = \boldsymbol{\ell},$$

where $\boldsymbol{\ell}$ is a vector of N ones. Without loss of generality, we can re-write these moment conditions as the combination of the pricing of $\mathbf{r}$ in (1) with:

$$E\left[R(a + \mathbf{b}'\mathbf{f})\right] = 1. \tag{13}$$

In addition, we can continue to estimate the SDF mean from the moment condition (2). The addition of the pricing of R in (13) implies that we no longer require an arbitrary normalization of $(a, \mathbf{b}, c)$.

If we think of a dimension d of the subspace of admissible SDFs, then we need to replicate the previous moment conditions d times, and use some normalization. Regardless of the value of d, we can show that working with $\mathbf{R}$ instead of $\mathbf{r}$ does not change the empirical evaluation of an empirical asset pricing model.

Proposition 2 *The CU version of the overidentification restriction test for the joint system*

$$E[\mathbf{r}(a_i + \mathbf{b}_i'\mathbf{f})] = \mathbf{0}, \qquad E\left[R(a_i + \mathbf{b}_i'\mathbf{f})\right] = 1, \qquad i = 1, 2, ..., d,$$

is numerically identical to the one for

$$E[\mathbf{r}(a_i + \mathbf{b}_i'\mathbf{f})] = \mathbf{0}, \qquad i = 1, 2, ..., d,$$

and the same applies to the respective estimators of the ratios of $\mathbf{b}_i$ *to* a_i*.*

Intuitively, the addition of gross returns allows us to pin down a_i and the mean of each basis SDF, c_i, but otherwise, it simply re-scales this variable.

For analogous reasons, the CU rank test we introduced in Proposition 1 is also numerically invariant to the addition of the following d replicas of the gross return moment condition (13):

$$E\left[R(1\ \mathbf{f}')\boldsymbol{\theta}_i\right] = 0, \qquad i = 1, 2, \ldots, d. \tag{14}$$

In other words, the CU version of the overidentification test of the SDF moment conditions (8) and (9) that imposes the d overspecification restrictions $c_1 = \ldots = c_d = 0$ yields the same value irrespective of whether or not we add the moment conditions (14).

Given that the payoff space spanned by $(R, \mathbf{r}')$ and $\mathbf{R}$ coincide, it is tedious but straightforward to prove that the rank test in Proposition 1 will also be asymptotically equivalent to the Cragg and Donald (1997) test used by Gospodinov, Kan, and Robotti (2019) for the null hypothesis that $\text{rank}(\boldsymbol{\ell}, \mathfrak{B}) = k$, where $\mathfrak{B} = Cov(\mathbf{R}, \mathbf{f})[Var(\mathbf{f})]^{-1}$.[9] In fact, we can easily generalize their approach to test that $\text{rank}(\boldsymbol{\ell}, \mathfrak{B}) = (k+1) - d$ for values of d bigger than one in a normalization-invariant way.

4.4 Identified sets

Another important difference with many of the aformentioned papers that look at models with possibly unnecesary factors is that they focus on the implications of those rank failures for standard GMM procedures, which assume point identification, while we propose alternative inference procedures that explicitly handle set identification.

In this respect, our procedure is closer to Kleibergen and Zhan (2020), who propose an alternative methodology to make inferences about the vector of risk prices regardless of the identification strength. Specifically, they construct identification-robust confidence intervals for those prices of risk by inverting the Wald test statistic of zero intercepts in the multivariate regression framework in (12). Thus, their confidence regions will be unbounded when there are identification problems with the prices of risk.

This methodology is certainly useful to detect identification problems, but it does not precisely characterise their source. In contrast, we can directly estimate the set of admissible SDFs compatible with the returns at hand, and test some of their properties, while at the same time providing a J test as a diagnostic on plausible values of d. In addition, our methods are easier to apply with more than one pricing factor.

[9] They assumed conditional homoskedasticity, but we can easily robustify our CU GMM tests.

5 Empirical Applications

5.1 A reassement of Yogo (2006)

As is well known, Yogo's theoretical model extends the CCAPM by assuming recursive preferences over a consumption bundle of nondurable and durable goods.[10] Therefore, in the linearized version of his model, the SDF will be an affine function of three factors: the market return, and the consumption growth of nondurables and durables, so that we can write the empirical SDF as:

$$m = a\left(1 - \delta_p f_p - \delta_c f_c - \delta_d f_d\right). \tag{15}$$

In practice, the log-growth rate of US real per capita consumption of nondurables and services and durables are identified with f_c and f_d, respectively. In turn, the return on wealth - proxied by the (log) return on the value-weighted U.S. stock market measured in real terms - is associated with f_p.

We initially evaluate this model with the original data, which corresponds to quarterly excess returns on the Fama-French cross-section of 25 size and book-to-market sorted portfolios from 1951 to 2001 (see Fama and French (1993) for further details).[11] In addition to the insightful nature of Yogo's (2006) theoretically motivated SDF specification, his results became very influential because he failed to reject the asset pricing restrictions, aligning the risk premia in the data with the risk premia generated by his model.

(Figure 1: Risk premia from 2S-GMM)

Nevertheless, the theoretical results in Burnside (2016) and Gospodinov, Kan, and Robotti (2019) indicate that a high cross-sectional R^2 may spuriously arise in models with useless factors too.

As an aside, we find that the results in Figure 1 depend on the estimation method (2-step GMM) and the imposition of some restrictions on the prices of risk.[12] Specifically, if we use instead iterated GMM starting from the 2-step estimates, we encounter a cycle with four different solutions.

(Figure 2: Risk premia from IT-GMM)

[10] Eichenbaum and Hansen (1990) were the first authors to empirically entertain the idea that it might be necessary to look at different consumption measures to successfully explain asset risk premia. Yogo's (2006) empirical model goes one important step further by combining their ideas with those in Epstein and Zin (1989).

[11] Note that although the market return is a traded factor, we do not add its pricing condition to (1) because it can effectively be generated as a portfolio of the cross-section of excess returns that we want to price.

[12] In Figures 1 to 3, we follow Yogo (2006) in using the moment conditions of the centred SDF approach mentioned in section 4.1, which are given by equation (11), with the normalization $c = 1$. As explained in Peñaranda and Sentana (2015), this matters for 2-step and iterated GMM, but not for CU-GMM.

Convergence does not improve if we free up the price of risk coefficients: iterated GMM enters yet another cycle of three different solutions.

(Figure 3: Risk premia from IT-GMM, free coefficients)

These discrepancies highlight the advantages of single-step GMM estimation procedures, but they might also be a sign of overspecification. For that reason, we apply our methodology to the same data. Specifically, we use the moment conditions (8) with $d = 1$, 2 and 3 to test for one, two and three-dimensional linear subsets of valid SDFs, respectively. In all cases, we augment those moment conditions with the exactly identified moment conditions (9) to obtain the associated SDF means. As we mentioned in section 3.2, we can then assess whether the model is completely overspecified by testing the joint significance of those means.

Table 1 shows the results of our overspecification analysis of the model. For reporting purposes, we display estimates of the SDF parameters using the popular SDF normalization $a = 1$, but our results are numerically invariant to this choice. We also report the usual J tests, as well as the criterion function under the restriction of zero SDF means, which is equivalent to a rank test for $Cov(\mathbf{r}, \mathbf{f})$ from Proposition 1. The p-values of the different J tests are shown in parenthesis.

(Table 1: Empirical evaluation of Yogo model 1951-2001)

We complement the J tests with significance tests for the prices of risk. In particular, to the right of the point estimates we include in parenthesis the p-value of the DM test of the null hypothesis of a zero parameter value. All the results correspond to a weighting matrix à la Newey and West (1987a) with one lag, but we obtained qualitatively similar conclusions when we used a VARHAC procedure also with one lag.[13]

The first, second and third panels of Table 1 refer to SDF sets of dimension 1, 2 and 3, respectively. As can be seen, we estimate the different subspaces for risk prices and SDFs using single-step GMM methods choosing those exclusion restrictions which are arguably easiest to interpret in each context. In the case of $d = 2$, in particular, we present the results for the simple normalization of the prices of risk given by $(\delta_p, \delta_c, 0)$ and $(\delta_p, 0, \delta_d)$.[14] Since the first factor is the market, we can interpret those SDFs as two variants of the linearized Epstein and Zin (1989) model, one with nondurable consumption and another with durable consumption. In contrast,

[13]Den Haan and Levin's (1997) VARHAC procedure assumes that the moment conditions have a finite VAR representation, which they exploit to estimate the required long-run covariance matrix.

[14]This normalization is identified as long as $\delta_c \neq 0$ and $\delta_d \neq 0$. In this respect, Table 1 shows that the DM tests that we proposed in section 3.3 reject that either $\delta_c = 0$ or $\delta_d = 0$.

in the case of $d = 3$ we present the results for the simple normalization $(\delta_p, 0, 0)$, $(0, \delta_c, 0)$ and $(0, 0, \delta_d)$, which effectively imposes that each factor can separately explain risk premia.

The results for the one-dimensional set entirely agree with the results in Yogo (2006), who finds that (i) the J test of two-step GMM does not reject his model for these 25 size- and value-sorted portfolios and (ii) durable consumption provides the only non-zero price of risk. In this respect, the usual overidentification test reported in the first column of Table 1 does not reject the null hypothesis that there exists an SDF affine in the three factors that can price the cross-section of securities (p-value=53.7%).

However, the validity of the asymptotic distribution of this J test crucially depends on the model parameters being point identified. For that reason, we also report the overidentification test for $d = 2$. As explained before, this test assesses whether there is a linear subspace of dimension 2 of admissible SDFs that can price the cross section of risk premia. We obtain a p-value of 13.4%, which suggests that the linearized version of Yogo's (2006) model in (15) is likely to be underidentified. In contrast, the overidentification test corresponding to $d = 3$ is strongly rejected, which reinforces the conclusion that the admissible SDFs (15) lie on a two-dimensional subspace. In turn, the DM tests show that all three factors are statistically significant, although the statistical significance of the market price of risk does not necessarily mean that the market porfolio is economically relevant in this model. In fact, the variability of the SDF basis is mainly driven by the two consumption measures.

Nevertheless, both consumption measures have low correlation with the vector of excess returns, which explains why the DM test of the null hypothesis that all the admissible SDFs have zero means when $d = 2$ has a p-value of 49.4%. This suggests that the seeming pricing ability of this set of SDFs simply exploits the lack of correlation of its elements with $\mathbf{r}$. In other words, the vector of risk premia does not appear to lie in the span of the covariance matrix of excess returns and factors, which suggests the model is completely overspecified.

Our results are in line with Burnside (2016), who finds that the matrix $Cov(\mathbf{r}, \mathbf{f})$ for this combination of test assets and pricing factors has rank 1 only. As Proposition 1 shows, an asymptotically equivalent rank test is given in the second panel of Table 1 by the J test that imposes zero SDF means. Given that its p-value is 15.1%, we do not reject either the null hypothesis that rank$[Cov(\mathbf{r}, \mathbf{f})] = 1$. But our results go further, in that they show that all the SDFs compatible with the linearized version of Yogo's (2006) model in (15) are economically meaningless because they are uncorrelated to the excesss returns on the test assets.

5.2 Robustness exercises

One potential concern with our GMM procedures is that the number of moments involved may be too large relative to the sample size. For that reason, we assess the reliability of the empirical results in Table 1 in two different ways: using a sample with a longer time span, and also with a smaller but more varied cross-section of test assets.

In the first case, we use the same data as Burnside (2016), whose sample period covers 1949-2012 (256 observations). Panel A of Table 2 shows that the findings in Table 1 still hold.[15] Specifically, we continue to find that the admissible SDFs lie on a two-dimensional subspace, and that they all have zero means, which confirms that the model is completely overspecifed.

(Table 2: Empirical evaluation of Yogo model 1949-2012)

We also find that the p-value of the test of rank$[Cov(\mathbf{r}, \mathbf{f})] = 1$ is 0.216. Not surprisingly, when we regress the estimated SDFs for $d = 2$ on the cross-section of returns, the corresponding R^2s are very low: 9.6% for the first element of the SDF basis and 9.0% for the second one.

Importantly, the complete overspecification of the model implies that the estimated SDF coefficients are not meaningful prices of risk that explain risk premia, but rather weights of linear combinations of factors uncorrelated with returns. In any event, when $d = 2$ the variability of the SDF basis is mainly driven by the two consumption measures, as in Table 1, the market portfolio having again a rather marginal role.

Lewellen, Nagel, and Shanken (2010) emphasized that empirical evaluations of asset pricing models that only look at size and book-to-market sorted portfolios may not be sufficiently informative, recommending the addition of industry portfolios. For that reason, in Panel B of Table 2 we repeat our empirical analysis, this time combining the set of five industry portfolios to the set of six size and book-to-market sorted portfolios in Ken French's data library, thereby avoiding a very large number of moment conditions.

Nevertheless, our results confirm that overspecification and underidentification problems are still prevalent. The test of rank$[Cov(\mathbf{r}, \mathbf{f})] = 1$ has a p-value of 0.164, so we cannot trust the results of the one-dimensional set. In fact, we can again conclude that the admissible SDFs (15) lie on a two-dimensional subspace, whose elements all have zero means. This is confirmed when we regress the estimated two-dimensional basis of SDFs on the cross-section of excess returns, as the R^2's are 3.7% and 11.6%, respectively.

[15] We follow Burnside (2016) in using real excess returns, while Yogo (2006) used nominal excess returns. Given that the effect of inflation is second order for excess returns, the choice between nominal and real returns is inconsequential for our results.

Finally, we can also confirm that our findings are not simply due to lack of power of our tests. Specifically, we apply the same methodology to another well known three factor model, the Fama-French model, whose pricing factors are the market portfolio and two portfolios that aim to capture the size and value effects; see Fama and French (1993) for details. For this model, the J statistic associated to a one-dimensional set is 60.55 with the 25 sorted portfolios, and 39.53 with the 11 sorted and industry portfolios, whose p-values are very close to zero. The corresponding J statistics for a two-dimensional set reject their null hypothesis too. Similarly, the rank test of Proposition 1 has a zero p-value in all cases. Therefore, the problem with this model is neither overspecification nor underidentification, but rather misspecification.

5.3 A reassesment of Jagannathan-Wang (1996)

Next, we re-evaluate the popular extension of the CAPM in Jagannathan and Wang (1996), who tried to capture the wealth portfolio by including a proxy for the return on human capital in addition to the market portfolio, and gave a role to conditioning information by adding the default spread as a third factor. Specifically, the SDF of their model is

$$m = a\left(1 - \delta_p f_p - \delta_l f_l - \delta_s f_s\right), \tag{16}$$

where f_p is the excess return on the value-weighted stock market index from Ken French's website, using the one-month T-bill rate from Ibbotson Associates as the nominally safe return, f_l is the growth rate in per capita labor income as a proxy for the human capital return, defined as the difference between total personal income and dividend payments divided by the total population (from the Bureau of Economic Analysis),[16] and f_s is the lagged default premium, measured as the yield spread between Baa- and Aaa-rated corporate bonds.

Panel A of Table 3 evaluates the Jagannathan-Wang model with monthly excess returns on the Fama-French cross-section of 25 size and book-to-market sorted portfolios. In the case of $d = 2$, we display results for the basis $(\delta_p, \delta_l, 0)$ and $(\delta_p, 0, \delta_s)$, so that we are simultaneously estimating a conditional CAPM with the default spread as the relevant state variable, and a traditional CAPM in which the return to the wealth portfolio is proxied by a linear combination of labor income and the return to the market portfolio.

We use the 647 observations from 1959:02 to 2012:12 in an earlier version of Gospodinov, Kan, and Robotti (2019), whose rank tests, like those in Kleibergen and Paap (2006), point out identification problems with this model. Not suprisingly, we find a two-dimensional subspace of valid SDFs, with a p-value of 0.199 for the corresponding J test.

[16] Following Jagannathan and Wang (1996), we use a two-month moving average for the purpose of minimizing measurement error.

(Table 3: Empirical evaluation of Jagannathan-Wang model 1959-2012)

In this case, the DM test cannot reject that the market portfolio does not enter any of the SDFs. This fact, combined with the low correlation between the two other pricing factors with the vector of excess returns explains the p-value of 0.116 for the null hypothesis that the mean of all admissible SDFs is zero. As we mentioned before, this is equivalent to those SDFs being uncorrelated to the test assets. To verify this claim, we regress the estimated SDFs on the vector of excess returns, finding that the R^2 are 5.3% and 4.3% for the two elements of the SDF basis. More formally, the p-value of the test that checks that rank$[Cov(\mathbf{r}, \mathbf{f})] = 1$ is 0.150.

Panel B of Table 3 combines the Fama-French cross-section of 6 size and book-to-market sorted portfolios with their set of 5 industry portfolios. In this case, our tests reject the existence of a valid two-dimensional set of SDFs, so there seems to be identifying information in the extended cross-section. We also find that the admissible SDF depends on labor income and the default spread, albeit the former is marginally significant.

Nevertheless, our results suggest that the one-dimensional set of admissible SDFs in Table 3, Panel B is completely overspecified, as we cannot reject the null hypothesis that their normalised version has zero mean. Again, we can verify that the estimated SDF is uncorrelated with the extended test assets by regressing it on the 11 excess returns, which gives rise to an R^2 of 1.6%.

Finally, we can once more confirm that our findings are not simply due to lack of power by applying our methodology to the Fama and French (1993) three factor model using the same return data. Thus, we find that the J statistic for a one-dimensional set is equal to 48.34 when we use the 11 sorted and industry portfolios, whose p-value is close to zero.

6 Conclusions

We study the estimation of prices of risk and the testing of the cross-sectional restrictions imposed by overspecified linear factor pricing models in which there is at least one non-trivial SDF which is uncorrelated with the excess returns of the test assets chosen by the researcher. We provide several textbook examples of this situation, which is necessary but not sufficient for the model parameters to be underidentified. In addition, we also emphasize the distinction between a model with uncorrelated pricing factors, which is necessarily overidentified, from a model with uncorrelated SDFs.

Unlike most previous studies, which focus on the non-standard asymptotic properties of the usual estimators and tests, our methods directly estimate the linear subspaces of prices of risk and associated SDFs compatible with the pricing restrictions of the model, which we can easily express in terms of linear moment conditions and efficiently estimate using standard GMM methods. In this regard, a non-trivial advantage of our procedures is that they have standard asymptotic distributions.

We use single-step GMM procedures, and in particular continuously updated GMM, to obtain identical test statistics and risk price estimates for SDF and regression methods, with uncentered or centred moments, and symmetric or asymmetric normalizations. Another non-trivial advantage of these methods is that they yield exactly the same conclusions for excess returns and gross returns.

We also propose simple tests to detect economically unattractive but empirically relevant situations in which the expected values of all SDFs in the identified set are 0, which is equivalent to their being uncorrelated to the test assets. In our opinion, researchers could convince readers that their results are meaningful by systematically reporting that they reject the restrictions implicit in these completely overspecified models.

More concretely, we recommend empirical researchers that when they evaluate asset pricing models they enrich the usual tables in two dimensions:

- An additional row with the estimates and significance test of the SDF mean. This would clarify whether a model which is not statistically rejected explains the cross-section of risk premia in an economically meaningful way.
- An additional column with the J tests and estimated risk prices (and the SDF mean test above) for a basis of every conceivable linear space of admissible SDFs.[17] This would shed light on the degree of underidentification.

As usual, the J test for $d = 1$ should not reject in an empirical model that prices returns. But in addition, both the DM test for $H_0 : c = 0$ and the J tests for $d \geq 1$ should reject. In this ideal situation, not only would the model be econometrically identified but it would also explain the cross-section of risk premia in a meaningful way. Still, the model could be useful even though the J test for $d = 2$ does not reject, as long as the zero mean hypothesis on the SDF basis is rejected.

In our first empirical application, we follow these recommendations to investigate the potential overspecification of the three-factor CCAPM model in Yogo (2006), which combines two macroeconomic factors: non-durable and durable consumption, and a stock market factor.

We evaluate the linearized version of this model with the original data, which corresponds to excess returns on the Fama-French cross-section of twenty-five size and book-to-market sorted portfolios from 1951 to 2001. Our results indicate that the admissible SDFs lie on a two-dimensional subspace. In addition, we cannot reject the null hypothesis that model is completely

[17]For example, if the proposed asset pricing model includes two factors, there should be two columns: $d = 1$ and $d = 2$. Likewise, with three factors, there should be three columns: $d = 1$, $d = 2$ and $d = 3$.

overspecified. Importantly, our results hold both when we update the sample period and when we consider the Fama-French cross-section of six size- and value-sorted portfolios jointly with their five industry portfolios. In addition, our simulations suggest that these empirical findings are not due to lack of power. On the contrary, if anything, our proposed tests tend to overreject for the sample sizes of our datasets.

In turn, our second empirical application evaluates the CAPM extension of Jagannathan and Wang (1996). When we use the twenty-five size and book-to-market sorted portfolios, we find evidence of underidentification, and the two-dimensional space of valid SDFs is completely overidentified. On the other hand, when we add industry porfolios to size- and value-sorted portfolios, we achieve identification, but complete overspecification remains because the (normalised) admissible SDF is uncorrelated with the extended cross-section of returns.

Finally, we also apply our methodology to the Fama and French (1993) three factor model, whose pricing factors are the market portfolio and two portfolios that aim to capture the size and value effects. We find that the problem with this model is neither overspecification nor underidentification, but rather misspecification.

Our econometric methodology is *positive* in nature, in the sense that our main objective has been to complement the diagnostics that researchers typically report in support of their preferred linear factor pricing specification so as to increase the empirical credibility of their results. Nevertheless, it might be interesting to combine our procedures with *normative* econometric methods that some researchers use to come up with an acceptable specification. Three recent proposals are Harvey, Liu and Zhu (2016), Bryzgalova (2016) and Kozak, Nagel and Santosh (2020). The application of our proposed diagnostics to models that have been selected after an implicit or explicit specification search raises multiple testing issues that we leave for future research. In fact, it might be possible to use the diagnostics that we have proposed in this paper in a factor selection procedure.

Another interesting avenue would be to consider bootstrap versions of our tests to improve their finite sample reliability. We could also apply our methods to other other portfolio sortings such as profitability and investment in Fama and French (2015), or other popular empirical asset pricing models such as the CCAPM extension of Lettau and Ludvigson (2001). In fact, we could also consider more general conditional settings with the sieve managed portfolios of Peñaranda and Sentana (2016). Finally, although our paper has focused on linear asset pricing models because they are pervasive in empirical asset pricing, one could explore the extension of our methodology to nonlinear models with multiple risk factors along the lines of sections 5 and 6 in Arellano, Hansen and Sentana (2012). We are currently pursuing some of these extensions.

References

Al-Saddoon, M.M. (2017): “A unifying theory of tests of rank”, *Journal of Econometrics* 199, 49-62.

Arellano, M. (2003): *Panel data econometrics.* Oxford University Press.

Arellano, M., L.P. Hansen and E. Sentana (2012): “Underidentification?”, *Journal of Econometrics* 170, 256-280.

Breeden, D.T. (1979): “An intertemporal asset pricing model with stochastic consumption and investment opportunities”, *Journal of Financial Economics* 7, 265-296.

Bryzgalova, S. (2016): “Spurious factors in linear asset pricing models”, Stanford GSB working paper.

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

Cragg, J.G. and S.G. Donald (1993): “Testing identifiability and specification in instrumental variable nodels”, *Econometric Theory* 9, 222-240.

Cragg, J.G. and S.G. Donald (1996): “On the asymptotic properties of LDU-based tests of the rank of a matrix”, *Journal of the American Statistical Association* 91, 1301-1309.

Cragg, J.G. and S.G. Donald (1997): “Inferring the rank of a matrix”, *Journal of Econometrics* 76, 223-250.

Den Haan, W.J. and A.T. Levin (1997): “A practitioner’s guide to robust covariance estimation”, in G.S. Maddala and C.R. Rao (eds.) *Handbook of Statistics* vol. 15, 299-342, Elsevier.

Dufour, J.-M. (1997): “Some impossibility theorems in econometrics with applications to structural and dynamic models”, *Econometrica* 65, 1365-1388.

Eichenbaum, M. and L.P. Hansen (1990): “Estimating models with intertemporal substitution using aggregate time series data”, *Journal of Business and Economic Statistics* 8, 53-69.

Epstein, L.G. and S.E. Zin (1989): “Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework”, *Journal of Political Economy* 57, 937-969.

Fama, E.F. and K.R. French (1993): “Common risk factors in the returns on stock and bonds”, *Journal of Financial Economics* 33, 3-56.

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

Gospodinov, N., R. Kan, and C. Robotti (2017): “Spurious inference in reduced-rank asset-pricing models”, *Econometrica* 85, 1613-1628.

Gospodinov, N., R. Kan, and C. Robotti (2019): "Too good to be true? Fallacies in evaluating risk factor models ", *Journal of Financial Economics* 132, 451-471.

Hansen, L.P. (1982): "Large sample properties of generalized method of moments estimators", *Econometrica* 50, 1029-1054.

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

Hansen, L.P. and R. Jagannathan (1991): "Implications of security market data for models of dynamic economies", *Journal of Political Economy* 99, 225-262.

Hansen, L.P. and R. Jagannathan (1997): "Assessing specification errors in stochastic discount factor models", *Journal of Finance* 52, 557-590.

Harvey, C. R., Y. Liu, and H. Zhu (2016): "... and the cross-section of expected returns", *Review of Financial Studies* 29, 5-68.

Jagannathan, R. and Z. Wang (1996): "The conditional CAPM and the cross-section of expected returns", *Journal of Finance* 51, 3-53.

Kan, R. and C. Zhang (1999): "GMM tests of stochastic discount factor models with useless factors", *Journal of Financial Economics* 54, 103-127.

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

Kleibergen, F. and Z. Zhan (2020): "Robust inference for consumption-based asset pricing", *Journal of Finance* 75, 507-550.

Kozak, S., S. Nagel, and S. Santosh (2020): "Shrinking the cross section", *Journal of Financial Economics* 135, 271-292.

Lettau, M., and S. Ludvigson (2001): "Resurrecting the (C)CAPM: a cross-sectional test when risk premia are time-varying", *Journal of Political Economy* 109, 1238-1287.

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

Manresa, E. (2009): "Application of rank tests to linear factor asset pricing models", unpublished Master thesis, CEMFI.

Merton, R.C. (1973): "An intertemporal capital asset pricing model", *Econometrica* 41, 867-887.

Newey, W.K. and D.L. McFadden (1994): "Large sample estimation and hypothesis testing", in R.F Engle and D.L. McFadden (eds.) *Handbook of Econometrics* vol. IV, 2111-2245, Elsevier.

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

Newey, W.K. and K.D. West (1987b): "Hypothesis testing with efficient method of moments estimation", *International Economic Review* 28, 777-787.

Peñaranda, F. and E. Sentana (2015): "A unifying approach to the empirical evaluation of asset pricing models", *Review of Economics and Statistics* 97, 412-435.

Peñaranda, F. and E. Sentana (2016): "Duality in mean-variance frontiers with conditioning information", *Journal of Empirical Finance* 38, 762-785.

Sargan, J.D. (1958): "The estimation of economic relationships using instrumental variables", *Econometrica* 26, 393-415.

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

Appendices

A Proofs

In the proofs, we follow Peñaranda and Sentana (2015) in exploiting three important properties of CU estimators and related single-step GMM procedures in an overidentifed GMM system in which we use the optimal weighting matrix. First, the inclusion of s additional unrestricted moment conditios with s new parameters does not affect the estimators of the original parameters or the value of the overidentification restrictions test (see e.g. Arellano (2003)). Second, the CU estimators and associated overidentification test are numerically invariant to parameter-dependent full-rank linear transformations of the influence functions (see Hansen, Heaton and Yaron (1996)). Third, CU is numerically invariant to continously differentiable bijective reparametrizations whose Jacobian matrix has full row rank in an open neighbourhood of the true values, in the sense that the overidentification restriction test is numerically identical and the reparametrized CU estimators are simply the result of applying the transformation to the original ones.

Proposition 1

We find it convenient to express the pricing conditions (1) in terms of central moments in (11), which is numerically inconsequential for single-step procedures such as CU-GMM (see Proposition 2 in Peñaranda and Sentana (2015) for a formal result).

As we explained in Section 4.1, we need to replicate d times the pricing conditions in (11) to deal with a $d-$dimensional subspace of admissible SDFs. Thus, the centred SDF counterpart to (8) will be based on the moment conditions

$$E\begin{pmatrix} \mathbf{r}m_1 \\ \vdots \\ \mathbf{r}m_d \\ \mathbf{f}-\boldsymbol{\mu} \end{pmatrix} = \mathbf{0}, \qquad m_i = c_i + (\mathbf{f}-\boldsymbol{\mu})'\,\mathbf{b}_i, \tag{A1}$$

where the basis $(\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_d)$ includes the necessary exclusion restrictions on the factors to guarantee its identification up to the normalization of each column.

Let us denote by J the CU-GMM value of the overidentifying restrictions test with free $(c_1, c_2, ..., c_d)$ in (A1). Similarly, let us denote by J_0 the CU-GMM value of the corresponding overidentifying restrictions test after imposing $c_1 = ... = c_d = 0$. In this context, it is straightforward to see that the overidentification test based on J_0 is trivially a rank test on $Cov(\mathbf{r}, \mathbf{f})$ because it is testing the existence of d linear combinations of the columns of this covariance matrix with weights $\mathbf{b}_i$ that are equal to zero

$$E\begin{pmatrix} \mathbf{r}\,(\mathbf{f}-\boldsymbol{\mu})'\,\mathbf{b}_1 \\ \vdots \\ \mathbf{r}\,(\mathbf{f}-\boldsymbol{\mu})'\,\mathbf{b}_d \\ \mathbf{f}-\boldsymbol{\mu} \end{pmatrix} = \mathbf{0}.$$

By the invariance properties of single-step GMM methods, it is easy to prove that we would obtain the same value for the overidentification test from the moment conditions (8) and (9).

Finally, note that our DM test of the null hypothesis $c_1 = ... = c_d = 0$ is based on $J_0 - J$.□

Proposition 2

Let us start with the simple case of $d = 1$. The addition of the pricing of R in (13) to the pricing of $\mathbf{r}$ in (1) implies that we no longer require an arbitrary normalization of $(a, \mathbf{b})$. As Peñaranda and Sentana (2015) prove in their Proposition 3, though, the empirical evidence obtained by single-step methods applied to $\mathbf{R}$ is consistent with the analogous evidence obtained from $\mathbf{r}$ alone. In particular, the overidentification restriction test for the joint system (1) and (13) is numerically identical to the one for (1) alone, and the ratio of the estimates of $\mathbf{b}$ to a obtained from the moment conditions for excess returns coincides with the same ratio obtained using all the assets.

The same comments apply to those situations with $d > 1$. The only difference is that they involve several SDFs, namely

$$E\begin{bmatrix} \mathbf{r}(a_1 + \mathbf{b}_1'\mathbf{f}) \\ R(a_1 + \mathbf{b}_1'\mathbf{f}) - 1 \\ \vdots \\ \mathbf{r}(a_d + \mathbf{b}_d'\mathbf{f}) \\ R(a_d + \mathbf{b}_d'\mathbf{f}) - 1 \end{bmatrix} = \mathbf{0}.$$

But since we add one moment and one parameter for each dimension, the equivalence between the results for excess and gross returns we have just discussed for $d = 1$ continues to hold for any d. □

Table 1: Empirical evaluation of Yogo model 1951-2001

	One-dimensional Set		Two-dimensional Set			Three-dimensional Set		
Market	0.200	(0.805)	-3.888	0.514	(0.002)	4.793	0	0
Nondur.	24.765	(0.458)	222.902	0	(0.000)	0	115.687	0
Durables	92.229	(0.035)	0	99.333	(0.000)	0	0	121.320
Mean	0.014	(0.790)	-0.099	0.034	(0.494)	0.852	0.421	-0.029
Criterion	20.743	(0.537)		56.687	(0.134)		215.144	(0.000)
Criterion $c = 0$	20.814	(0.592)		58.098	(0.151)			

Notes. This table displays estimates of the SDF parameters, as well as the J and J_0 tests (with free and constrained SDF means) with p-values in parenthesis (). The number of degrees of freedom of these test are $22(= 26 - 4)$ and 23 for $d = 1$, $46(= 52 - 6)$ and 48 for $d = 2$, and $72(= 78 - 6)$ and 75 for $d = 3$. The first, second and third blocks of columns report the results for sets of SDFs of dimension 1, 2 and 3, respectively. We display the estimates for a particular normalization, but our results are numerically invariant to the chosen normalization because we use CU-GMM. The J tests are complemented with significance tests of some SDF parameters. In particular, the p-value of the distance metric test of the null hypothesis of zero parameters is reported in parenthesis to the right of the estimates. Distance metric tests are not reported when the p-value of the J test is lower than 0.01. The payoffs of the test assets correspond to 25 nominal excess returns of size and book-to-market sorted portfolios on a quarterly basis.

Table 2 : Empirical evaluation of Yogo model 1949-2012

	One-dimensional Set		Two-dimensional Set			Three-dimensional Set		
Panel A. 25 size and book-to-market sorted portfolios								
Market	0.766	(0.673)	-1.878	0.882	(0.000)	12.882	0	0
Nondur.	-6.452	(0.834)	192.583	0	(0.000)	0	169.191	0
Durables	106.144	(0.024)	0	97.810	(0.000)	0	0	110.143
Mean	0.003	(0.972)	0.052	0.008	(0.757)	0.411	0.065	-0.075
Criterion	18.278	(0.689)		54.818	(0.175)		165.053	(0.000)
Criterion $c = 0$	18.279	(0.742)		55.375	(0.216)			
Panel B. 6 size and book-to-market sorted portfolios, and 5 industry portfolios								
Market	-4.864	(0.089)	-1.165	1.422	(0.002)	7.743	0	0
Nondur.	415.901	(0.001)	194.803	0	(0.000)	0	184.681	0
Durables	-89.991	(0.400)	0	84.501	(0.000)	0	0	119.130
Mean	0.053	(0.800)	0.079	0.093	(0.142)	0.640	0.090	-0.087
Criterion	5.941	(0.654)		22.165	(0.225)		85.715	(0.000)
Criterion $c = 0$	6.005	(0.739)		26.066	(0.164)			

Notes. This table displays estimates of the SDF parameters, as well as the J and J_0 tests (with free and constrained SDF means) with p-values in parenthesis (). In Panel A, the number of degrees of freedom of these test are $22(= 26 - 4)$ and 23 for $d = 1$, $46(= 52 - 6)$ and 48 for $d = 2$, and $72(= 78 - 6)$ and 75 for $d = 3$, while in Panel B they are $8(= 12 - 4)$ and 9 for $d = 1$, $18(= 24 - 6)$ and 20 for $d = 2$, and $30(= 36 - 6)$ and 33 for $d = 3$. The first, second and third blocks of columns report the results for sets of SDFs of dimension 1, 2 and 3, respectively. We display the estimates for a particular normalization, but our results are numerically invariant to the chosen normalization because we use CU GMM. The J tests are complemented with significance tests of some SDF parameters. In particular, the p-value of the distance metric test of the null hypothesis of zero parameters is reported in parenthesis to the right of the estimates. Distance metric tests are not reported when the p-value of the J test is lower than 0.01. The payoffs of the test assets correspond to 25 real excess returns of size and book-to-market sorted portfolios at the quarterly frequency (Panel A), and 6 size- and value-sorted portfolios plus 5 industry portfolios (Panel B).

Table 3: Empirical evaluation of Jagannathan-Wang model 1959-2012

	One-dimensional Set		Two-dimensional Set			Three-dimensional Set		
Panel A. 25 size and book-to-market sorted portfolios								
Market	-0.012	(0.995)	1.130	-0.745	(0.119)	6.925	0	0
Labor	-146.74	(0.020)	268.127	0	(0.000)	0	327.991	0
Premium	45.261	(0.587)	0	99.848	(0.000)	0	0	81.699
Mean	-0.114	(0.093)	-0.217	0.033	(0.116)	0.918	-0.530	0.226
Criterion	22.964	(0.404)		53.847	(0.199)		208.090	(0.000)
Criterion $c = 0$	25.778	(0.311)		58.157	(0.150)			
Panel B. 6 size and book-to-market sorted portfolios, and 5 industry portfolios								
Market	-0.253	(0.544)	1.442	-0.736		4.845	0	0
Labor	58.613	(0.066)	259.386	0		0	347.978	0
Premium	72.463	(0.001)	0	93.845		0	0	76.529
Mean	0.002	(0.968)	-0.203	0.079		0.958	-0.623	0.244
Criterion	12.570	(0.128)		41.588	(0.001)		121.842	(0.000)
Criterion $c = 0$	12.572	(0.183)		44.857	(0.001)			

Notes. This table displays estimates of the SDF parameters, as well as the J and J_0 tests (with free and constrained SDF means) with p-values in parenthesis (). In Panel A, the number of degrees of freedom of these test are $22(= 26 - 4)$ and 23 for $d = 1$, $46(= 52 - 6)$ and 48 for $d = 2$, and $72(= 78 - 6)$ and 75 for $d = 3$, while in Panel B they are $8(= 12 - 4)$ and 9 for $d = 1$, $18(= 24 - 6)$ and 20 for $d = 2$, and $30(= 36 - 6)$ and 33 for $d = 3$. The first, second and third blocks of columns report the results for sets of SDFs of dimension 1, 2 and 3, respectively. We display the estimates for a particular normalization, but our results are numerically invariant to the chosen normalization because we use CU GMM. The J tests are complemented with significance tests of some SDF parameters. In particular, the p-value of the distance metric test of the null hypothesis of zero parameters is reported in parenthesis to the right of the estimates. Distance metric tests are not reported when the p-value of the J test is lower than 0.01. The payoffs of the test assets correspond to 25 nominal excess returns of size and book-to-market sorted portfolios at the monthly frequency (Panel A), and 6 size- and value-sorted portfolios plus 5 industry portfolios (Panel B).

Figure 1: Risk premia from 2S-GMM

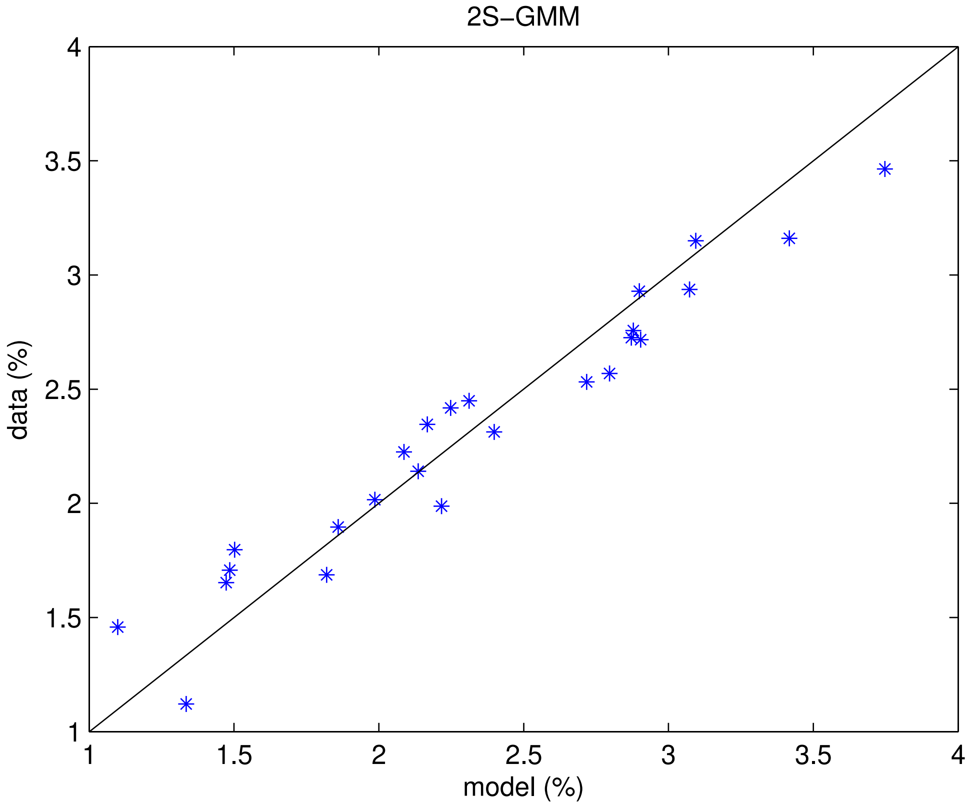


Figure 2: Risk premia from IT-GMM

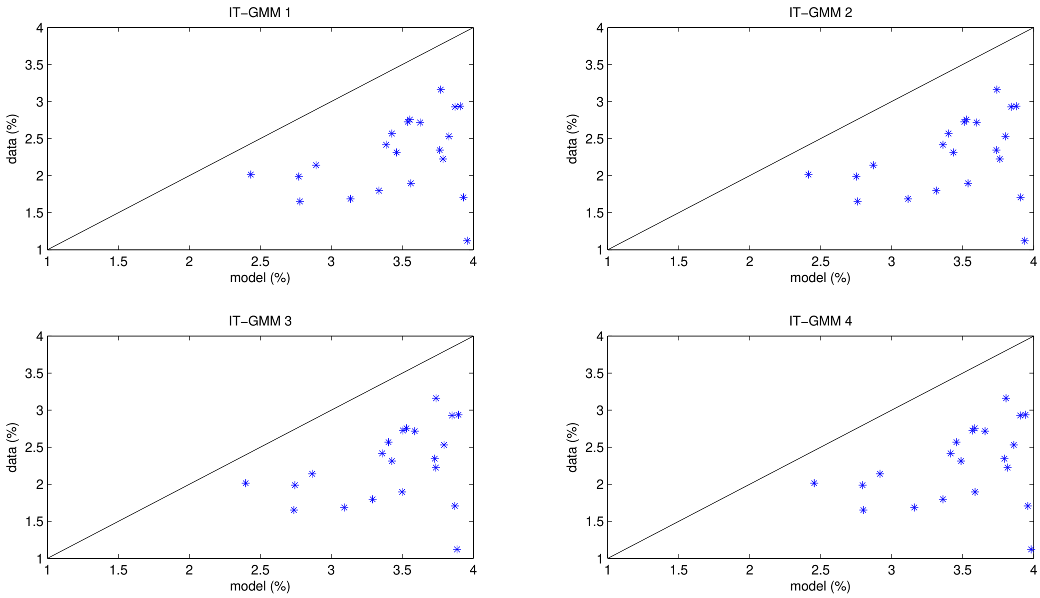
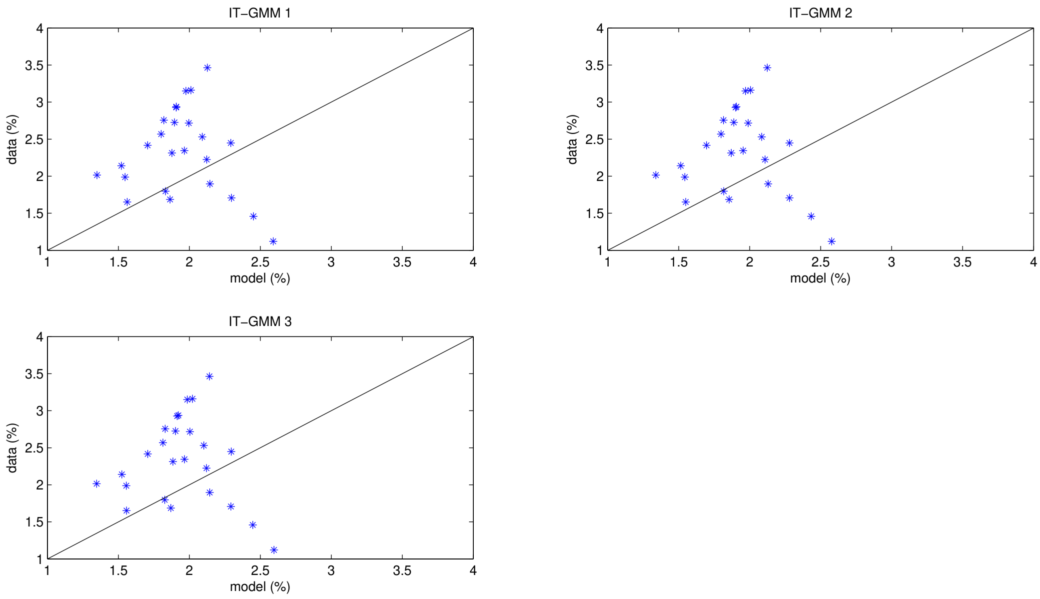


Figure 3: Risk premia from IT-GMM, free coefficients



Supplemental Appendices for Empirical Evaluation of Overspecified Asset Pricing Models

Elena Manresa
New York University, 19 West 4th St, New York, NY 10012, USA
<elena.manresa@nyu.edu>

Francisco Peñaranda
Queens College CUNY, 65-30 Kissena Blvd, Flushing, NY 11367, USA
<francisco.penaranda@qc.cuny.edu>

Enrique Sentana
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<sentana@cemfi.es>

Revised: June 2020

B A geometric interpretation of admissible SDF sets

It is pedagogically convenient to think about the restrictions a linear factor pricing model such as (1) imposes on the parameters $(a, \mathbf{b}, c)$ as we increase the number of assets we consider. For simplicity, we focus on the case of two pricing factors (f_1, f_2), as in Section 2.2, where these empirical factors are the market portfolio and nondurable consumption, or durable and nondurable consumptions. Either way, the matrix $\mathbf{M}$ in (7) can then be expressed as

$$\mathbf{M} = \left[\begin{array}{ccc} E(\mathbf{r}) & E(\mathbf{r}f_1) & E(\mathbf{r}f_2) \end{array}\right],$$

for an $n \times 1$ vector of excess returns. Admissible SDFs are defined by $\mathbf{M}\boldsymbol{\theta} = \mathbf{0}$. If there exists a solution to these equations, then we say that the empirical model holds.

When $n = 1$, there is always a two dimensional linear space of admissible solutions, which can be regarded as the dual set to the combination line of expected excess returns and covariances with the risk factors that can be generated by leveraging r_1 up or down.

(Figure B1: One asset)

When $n = 2$, the two dimensional space generated by each asset will generally be different, so their intersection will be a straight line.

(Figure B2: Two assets)

Occasionally, though, the two linear subspaces might coincide. This will happen when the two assets are collinear in the space of expected excess returns and covariances with the risk factors, an issue we will revisit when we discuss Figures B6 and B7 below.

Three assets is the minimum number required to be able to reject the model. The reason is the following. If an empirical asset pricing model does not hold, the three linear subspaces associated to each of the assets will only intersect at the origin. We may then say that there is financial markets "segmentation", in the sense that there is no single SDF within the model that can price all the assets. One such example would be the Epstein-Zin empirical specification considered in section 2.2 when the true model is the ICAPM but the factor mimicking portfolio for consumption growth is not spanned by the market and the factor mimicking portfolio for the state variable.

(Figure B3: Three segmented asset markets)

If on the other hand the proposed empirical asset pricing model holds, the intersection will be a linear subspace of positive dimension. This requires that the three assets are coplanar in the space of expected excess returns and covariances with the risk factors, so that they all lie on the security market plane $E(\mathbf{r}) = E(\mathbf{r}f_1)\delta_1 + E(\mathbf{r}f_2)\delta_2$. Therefore,

$$\mathbf{M} = \left[\begin{array}{cc} E(\mathbf{r}f_1) & E(\mathbf{r}f_2) \end{array}\right] \begin{bmatrix} \delta_1 & 1 & 0 \\ \delta_2 & 0 & 1 \end{bmatrix}.$$

When this happens, we may say that there is financial markets "integration". The same example discussed in the previous paragraph will give rise to this situation when the factor mimicking portfolio for consumption growth is spanned by the market and the factor mimicking portfolio for the state variable.

(Figure B4: Three integrated asset markets)

Another example in which the empirical Epstein - Zin specification (3) in section 2.2 holds would arise when the true model is the CAPM but the market portfolio is not proportional to the consumption growth mimicking portfolio, so that

$$\mathbf{M} = \begin{bmatrix} E(\mathbf{r}f_1) & E(\mathbf{r}f_2) \end{bmatrix} \begin{bmatrix} \delta_1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

An interesting feature of this example is that consumption growth does not appear in any admissible SDF. We discussed tests for such a hypothesis in section 3.2. Formally, the null hypothesis would be that the entry of b associated to this factor is equal to zero in all the basis vectors $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_d)$.

(Figure B5: An unpriced second factor)

We can also use this graphical framework to represent the other different situations that we discuss in section 2.2. Specifically, assume that both the CAPM and the (linearized) CCAPM hold, in the sense that excess returns on the market and consumption growth can price on their own a cross-section of excess returns, i.e. $E(\mathbf{r}) = E(\mathbf{r}f_1)\delta_1$ and $E(\mathbf{r}) = E(\mathbf{r}f_2)\delta_2$. As a consequence,

$$\mathbf{M} = E(\mathbf{r})\begin{pmatrix} 1 & 1/\delta_1 & 1/\delta_2 \end{pmatrix},$$

for the (linearized) Epstein-Zin model (3), which means that we can find a two-dimensional subspace of SDFs whose parameters satisfy $\mathbf{M}\boldsymbol{\theta} = \mathbf{0}$. Nevertheless, except for a linear subspace of dimension 1, most SDFs in the admisibble set will have a meaningful economic interpretation.

(Figure B6: Two single factor models)

A closely related situation would be as follows. Consider a two-factor model with a useless factor such that $Cov(\mathbf{r}, f_2) = \mathbf{0}$, so that

$$\mathbf{M} = \begin{bmatrix} E(\mathbf{r}) & E(\mathbf{r}f_1) & E(\mathbf{r})\mu_2 \end{bmatrix},$$

where μ_2 is the population mean of the second empirical factor. If f_1 is a valid pricing factor on its own, so that $E(\mathbf{r}) = E(\mathbf{r}f_1)\delta_1$, then rank$(\mathbf{M}) = 1$ because

$$\mathbf{M} = E(\mathbf{r})\begin{pmatrix} 1 & 1/\delta_1 & \mu_2 \end{pmatrix}.$$

Consequently, this overspecified pricing model will be economically meaningful but parametrically underidentified.

(Figure B7: Valid and attractive model with a useless factor)

In contrast, if $E(\mathbf{r})$ and $E(\mathbf{r}f_1)$ are linearly independent because the true model involves a second risk factor, then the model parameters will be econometric identified because rank$(\mathbf{M}) = 2$, and we can still rely on standard GMM inference. However, in these circumstances there can be no admissible SDF affine in the two empirical factors that can both explain cross-sectional risk premia and have a meaningful economic interpretation. This is the usual example of a useless factor, which we also discuss in section 2.2.

Indeed, when $Cov(\mathbf{r}, f_2) = \mathbf{0}$ but $E(\mathbf{r}) \neq \mathbf{0}$, the SDF conditions (1) will trivially hold for any $m \propto (f_2 - \mu_2)$ because they will all satisfy $\mathbf{M}\boldsymbol{\theta} = \mathbf{0}$. As a result, the admissible SDFs will have $b_1 = 0$ and $c = E(m) = 0$. Thus, this overspecified model will be econometrically identified but economically unattractive.

(Figure B8: Valid but unattractive model with a useless factor)

Finally, there will also be a two-dimensional subspace of SDFs whose parameters satisfy $\mathbf{M}\boldsymbol{\theta} = \mathbf{0}$ when there are two useless factors, i.e. $Cov(\mathbf{r}, f_1) = Cov(\mathbf{r}, f_2) = \mathbf{0}$. Hence,

$$\mathbf{M} = E(\mathbf{r})\begin{pmatrix} 1 & \mu_1 & \mu_2 \end{pmatrix},$$

and any SDF which is a linear combination of $f_1 - \mu_1$ and $f_2 - \mu_2$ will work. The final example in section 2.2 provides an illustration with durable and nondurable consumption growth.

(Figure B9: Two useless factors)

The special feature of this completely overspecified case is that $c = 0$ for all admissible SDFs, so there is not only underidentification but also the absence of any economic meaningful specification.

C Normalizations and starting values

We saw in section 2.1 that the parameter vector $(a, \mathbf{b}, c)$ that appears in (1) and (2) is only identified up to scale. As forcefully argued by Hillier (1990) for single equation IV models, this suggests that we should concentrate our efforts in estimating the identified direction. However, empirical researchers often prefer to estimate points rather than directions, and for that reason they typically focus on some asymmetric scale normalization, such as $(1, \mathbf{b}/a, c/a)$. In this regard, note that $\boldsymbol{\delta} = -\mathbf{b}/a$ can be interpreted as prices of risk since we may rewrite (1) as $E(\mathbf{r}) = E(\mathbf{r}\mathbf{f}')\boldsymbol{\delta}$. Other normalizations, such as $(a/c, \mathbf{b}/c, 1)$ or $\mathbf{b}'\mathbf{b} + c^2 = 1$ are also possible.

(Figure C1: Normalizations)

Similarly, the extended system of moment conditions (8) and (9) also requires normalizations. Although any asymmetric normalization may be be problematic in certain circumstances (see section 4.4 in Peñaranda and Sentana (2015) for further details in the case of a single pricing factor), in the presentation of our empirical results we use a popular SDF normalization that fixes the first element of each $\boldsymbol{\theta}_i$ to 1. Additionally, we need to impose enough zero restrictions on the prices of risk to achieve identification. Alternatively, we could make a $d \times d$ block of (a permutation of) the matrix $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_d)$ equal to the identity matrix of order d. Either way, the advantage of CU-GMM and other GEL estimators is that our inferences will be numerically invariant to the chosen normalization.

Nevertheless, one drawback of these single-step methods is that they involve a non-linear optimization procedure even though the moment conditions are linear in parameters, which may result in multiple local minima. For that reason, we propose to use as starting value a computationally simple intuitive estimator that is always consistent, but which would become efficient when the returns and factors are *i.i.d.* elliptical. This family of distributions includes the multivariate normal and Student t distributions as special cases, which are often assumed in theoretical and empirical finance.

Let us define $(\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_d)$ as the vectors of factors that enter each one of the SDFs in (8) after imposing the necessary restrictions that guarantee the point identification of the basis of risk prices $(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, ..., \boldsymbol{\delta}_d)$, where $\boldsymbol{\delta}_i$ contains only those prices of risk which have not been set to 0 for identification purposes, so that the corresponding Jacobian matrices $E(\mathbf{r}\mathbf{f}_i')$ have full rank.

As a result, we can re-write (8) as

$$E\left[(1 - \mathbf{f}_1'\boldsymbol{\delta}_1)\mathbf{r}\right] = \mathbf{0}, \qquad i = 1, 2, .., d, \tag{C1}$$

and (9) as

$$E(1 - \mathbf{f}_i'\boldsymbol{\delta}_i - c_i) = 0, \qquad i = 1, 2, .., d. \tag{C2}$$

Let $\mathbf{r}_t$ and $\mathbf{f}_t$ denote the values of the excess returns on the n assets and the k factors at time t. We can then prove that

Proposition C1 *If $(\mathbf{r}_t, \mathbf{f}_t)$ is an i.i.d. elliptical random vector with bounded fourth moments such that (C1) holds, then:*
a) The most efficient GMM estimator of $\boldsymbol{\delta}_i$ $(i = 1, \ldots, d)$ from the system (C1) will be given by

$$\mathring{\boldsymbol{\delta}}_{iT} = \left(\sum_{t=1}^{T} \tilde{\mathbf{r}}_{it}^{+}\tilde{\mathbf{r}}_{it}^{+\prime}\right)^{-1} \sum_{t=1}^{T} \tilde{\mathbf{r}}_{it}^{+}, \tag{C3}$$

where $\tilde{\mathbf{r}}_{it}^{+}$ are the relevant elements of the sample factor mimicking portfolios

$$\tilde{\mathbf{r}}_t^{+} = \left(\sum_{s=1}^{T} \mathbf{f}_s\mathbf{r}_s'\right)\left(\sum_{s=1}^{T} \mathbf{r}_s\mathbf{r}_s'\right)^{-1} \mathbf{r}_t. \tag{C4}$$

b) When we combine the moment conditions (C1) with (C2), the most efficient GMM estimator of each $\boldsymbol{\delta}_i$ is the same as in a), and the most efficient GMM estimator each c_i is the sample mean of the corresponding SDF.

Proof. We develop most of the proof for the case $d = 2$ to simplify the expressions, but explain the extension to $d > 2$ at the end.

a) When $d = 2$, the moment conditions (C1) become

$$E(\mathbf{m} \otimes \mathbf{r}) = E\begin{pmatrix} m_1\mathbf{r} \\ m_2\mathbf{r} \end{pmatrix} = E\begin{bmatrix} (1 - \mathbf{f}_1'\boldsymbol{\delta}_1)\,\mathbf{r} \\ (1 - \mathbf{f}_2'\boldsymbol{\delta}_2)\,\mathbf{r} \end{bmatrix} = \mathbf{0}.$$

We know from Hansen (1982) that the optimal moments correspond to the linear combinations

$$\mathbf{D}'\mathbf{S}^{-1}\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} m_{1t}\mathbf{r}_t \\ m_{2t}\mathbf{r}_t \end{pmatrix},$$

where $\mathbf{D}$ is the expected Jacobian and $\mathbf{S}$ the corresponding long-run variance

$$\mathbf{S} = avar\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{pmatrix} m_{1t}\mathbf{r}_t \\ m_{2t}\mathbf{r}_t \end{pmatrix}\right].$$

In this setting, the expected Jacobian trivially is

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{pmatrix}, \qquad \mathbf{D}_i = -E\left(\mathbf{r}\mathbf{f}_i'\right).$$

Since we assume that the chosen normalization $(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$ is identified, $\mathbf{D}$ has full column rank, which in turn implies that both $\mathbf{D}_1$ and $\mathbf{D}_2$ must have full column rank too.

When $(\mathbf{r}_t, \mathbf{f}_t)$ is an i.i.d. elliptical random vector with bounded fourth moments, we can tediously show that the long-run covariance matrix of the influence functions will be

$$\mathbf{S} = \mathcal{A} \otimes E\left(\mathbf{r}\mathbf{r}'\right) - \mathcal{B} \otimes E(\mathbf{r})\,E(\mathbf{r})',$$

$$\mathcal{A} = (1+\kappa)\,V(\mathbf{m}) + E(\mathbf{m})\,E(\mathbf{m})', \qquad \mathcal{B} = \kappa V(\mathbf{m}) + 2(1-\kappa)\,E(\mathbf{m})\,E(\mathbf{m})',$$

where κ is the coefficient of multivariate excess kurtosis (see Fang, Kotz and Ng (1990)).

To relate the optimal moments to the factor mimicking portfolios

$$\mathbf{r}_i^+ = \mathbf{C}_i\mathbf{r}, \qquad \mathbf{C}_i = E\left(\mathbf{r}\mathbf{f}_i'\right)' E^{-1}\left(\mathbf{r}\mathbf{r}'\right),$$

it is convenient to define the matrix

$$\mathbf{C}' = \begin{pmatrix} \mathbf{C}_1' & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2' \end{pmatrix},$$

on the basis of which we can compute

$$\mathbf{S}\mathbf{C}' = \left[\mathcal{A} \otimes E\left(\mathbf{r}\mathbf{r}'\right) - \mathcal{B} \otimes E(\mathbf{r})\,E(\mathbf{r})'\right]\begin{pmatrix} \mathbf{C}_1' & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2' \end{pmatrix}$$

$$= \begin{pmatrix} \mathcal{A}_{11}E\left(\mathbf{r}\mathbf{f}_1'\right) & \mathcal{A}_{12}E\left(\mathbf{r}\mathbf{f}_2'\right) \\ \mathcal{A}_{12}E\left(\mathbf{r}\mathbf{f}_1'\right) & \mathcal{A}_{22}E\left(\mathbf{r}\mathbf{f}_2'\right) \end{pmatrix} - \begin{pmatrix} \mathcal{B}_{11}E(\mathbf{r})\,E(\mathbf{r})'\,\mathbf{C}_1' & \mathcal{B}_{12}E(\mathbf{r})\,E(\mathbf{r})'\,\mathbf{C}_2' \\ \mathcal{B}_{12}E(\mathbf{r})\,E(\mathbf{r})'\,\mathbf{C}_1' & \mathcal{B}_{22}E(\mathbf{r})\,E(\mathbf{r})'\,\mathbf{C}_2' \end{pmatrix}.$$

Given that the existence of two valid SDFs implies that $E(\mathbf{r}) = E(\mathbf{r}\mathbf{f}_1')\boldsymbol{\delta}_1 = E(\mathbf{r}\mathbf{f}_2')\boldsymbol{\delta}_2$, we can write these matrices as

$$\mathbf{S}\mathbf{C}' = \begin{pmatrix} \mathcal{A}_{11}E\left(\mathbf{r}\mathbf{f}_1'\right) & \mathcal{A}_{12}E\left(\mathbf{r}\mathbf{f}_2'\right) \\ \mathcal{A}_{12}E\left(\mathbf{r}\mathbf{f}_1'\right) & \mathcal{A}_{22}E\left(\mathbf{r}\mathbf{f}_2'\right) \end{pmatrix} - \begin{pmatrix} \mathcal{B}_{11}E\left(\mathbf{r}\mathbf{f}_1'\right)\boldsymbol{\delta}_1\boldsymbol{\delta}_1'\mathbf{G}_1 & \mathcal{B}_{12}E\left(\mathbf{r}\mathbf{f}_2'\right)\boldsymbol{\delta}_2\boldsymbol{\delta}_2'\mathbf{G}_2 \\ \mathcal{B}_{12}E\left(\mathbf{r}\mathbf{f}_1'\right)\boldsymbol{\delta}_1\boldsymbol{\delta}_1'\mathbf{G}_1 & \mathcal{B}_{22}E\left(\mathbf{r}\mathbf{f}_2'\right)\boldsymbol{\delta}_2\boldsymbol{\delta}_2'\mathbf{G}_2 \end{pmatrix},$$

$$\mathbf{G}_i = E\left(\mathbf{r}\mathbf{f}_i'\right)' E^{-1}\left(\mathbf{r}\mathbf{r}'\right) E\left(\mathbf{r}\mathbf{f}_i'\right).$$

In addition, let us define the matrices $\mathbf{Q}_i$ such that $E(\mathbf{r}\mathbf{f}_1') = E(\mathbf{r}\mathbf{f}_2')\mathbf{Q}_1$ and $E(\mathbf{r}\mathbf{f}_2') = E(\mathbf{r}\mathbf{f}_1')\mathbf{Q}_2$, which are related by $\mathbf{Q}_2 = \mathbf{Q}_1^{-1}$. The existence of these matrices is guaranteed by the lack of full column rank of $E(\mathbf{r}\mathbf{f}')$ together with the full column rank of $E(\mathbf{r}\mathbf{f}_1')$ and $E(\mathbf{r}\mathbf{f}_2')$. Thus, we can write

$$\mathbf{S}\mathbf{C}' = \mathbf{D}\mathbb{Q},$$

$$\mathbb{Q} = -\begin{pmatrix} \mathcal{A}_{11}\mathbf{I}_1 - \mathcal{B}_{11}\boldsymbol{\delta}_1\boldsymbol{\delta}_1'\mathbf{G}_1 & \mathbf{Q}_2\left(\mathcal{A}_{12}\mathbf{I}_1 - \mathcal{B}_{12}\boldsymbol{\delta}_2\boldsymbol{\delta}_2'\mathbf{G}_2\right) \\ \mathbf{Q}_1\left(\mathcal{A}_{12}\mathbf{I}_2 - \mathcal{B}_{12}\boldsymbol{\delta}_1\boldsymbol{\delta}_1'\mathbf{G}_1\right) & \mathcal{A}_{22}\mathbf{I}_2 - \mathcal{B}_{22}\boldsymbol{\delta}_2\boldsymbol{\delta}_2'\mathbf{G}_2 \end{pmatrix}.$$

The assumption that $\mathbf{D}'\mathbf{S}^{-1}$ has full row rank guarantees that the same is true for $\mathbf{C}$, so that $\mathbb{Q}$ will be invertible. Therefore, we have found that

$$\mathbf{D}'\mathbf{S}^{-1} = \mathbb{Q}'^{-1}\mathbf{C}.$$

In other words, the rows of $\mathbf{D}'\mathbf{S}^{-1}$ are spanned by the rows of $\mathbf{C}$, which confirms that the factor mimicking portfolios span the optimal instrumental variables.

As a result, the optimal moments can be expressed as

$$\begin{pmatrix} \mathbf{C}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 \end{pmatrix} \frac{1}{T}\sum_{t=1}^{T} \begin{pmatrix} m_{1t}\mathbf{r}_t \\ m_{2t}\mathbf{r}_t \end{pmatrix} = \frac{1}{T}\sum_{t=1}^{T} \begin{pmatrix} \mathbf{r}_{1t}^{+}m_{1t} \\ \mathbf{r}_{2t}^{+}m_{2t} \end{pmatrix} = \mathbf{0},$$

which proves that the optimal estimator of each vector of risk prices simply uses the corresponding factor mimicking portfolios. This estimator is infeasible because we do not know $\mathbf{C}_i$, but under standard regularity conditions we can replace $\mathbf{r}_{it}^{+}$ by its sample counterpart in (C4) without affecting the asymptotic distribution.

b) When $d = 2$, the joint system of moments (C1) and (C2)

$$E\left(\mathbf{h}\right) = E\begin{pmatrix} \mathbf{m}\otimes\mathbf{r} \\ \mathbf{m} - \mathbf{c} \end{pmatrix},$$

consists of

$$E\left(\mathbf{m}\otimes\mathbf{r}\right) = E\begin{pmatrix} m_1\mathbf{r} \\ m_2\mathbf{r} \end{pmatrix} = E\begin{bmatrix} \left(1 - \mathbf{f}_1'\boldsymbol{\delta}_1\right)\mathbf{r} \\ \left(1 - \mathbf{f}_2'\boldsymbol{\delta}_2\right)\mathbf{r} \end{bmatrix} = \mathbf{0},$$

$$E\left(\mathbf{m} - \mathbf{c}\right) = E\begin{pmatrix} m_1 - c_1 \\ m_2 - c_2 \end{pmatrix} = E\begin{bmatrix} 1 - \mathbf{f}_1'\boldsymbol{\delta}_1 - c_1 \\ 1 - \mathbf{f}_2'\boldsymbol{\delta}_2 - c_1 \end{bmatrix} = \mathbf{0},$$

with the parameters being

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\delta} \\ \mathbf{c} \end{pmatrix}, \qquad \boldsymbol{\delta} = \begin{pmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \end{pmatrix}, \qquad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$

The optimal moments correspond to the linear combinations

$$\mathcal{D}'\mathcal{S}^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbf{h}_t,$$

where $\mathcal{D}$ is the expected Jacobian and $\mathcal{S}$ the corresponding long-run variance

$$\mathcal{S} = avar\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{h}_t\right].$$

In this setting, the expected Jacobian can be decomposed as

$$\mathcal{D} = \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbb{D} & -\mathbf{I}_2 \end{pmatrix},$$

where $\mathbb{D}$ contains the Jacobian of $\mathbf{m}-\mathbf{c}$ with respect to $\boldsymbol{\delta}$, and $\mathbf{I}_2$ is the identity matrix of order 2. The long-run variance for i.i.d. returns and factors can be decomposed as

$$\mathcal{S} = \begin{pmatrix} \mathbf{S} & E\left(\mathbf{mm}'\otimes\mathbf{r}\right) \\ E\left(\mathbf{mm}'\otimes\mathbf{r}'\right) & Var\left(\mathbf{m}\right) \end{pmatrix}.$$

Once again, we can exploit the structure of the optimal moments to show that the optimal estimator of $\boldsymbol{\delta}$ satisfies the moment conditions

$$\mathbf{D}'\mathbf{S}^{-1}\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{m}_t\otimes\mathbf{r}_t\right) = \mathbf{0}.$$

Hence, the optimal estimator of $\mathbf{c}$ will satisfy the moment conditions

$$\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{m}_t-\mathbf{c}\right) - E\left(\mathbf{mm}'\otimes\mathbf{r}'\right)\mathbf{S}^{-1}\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{m}_t\otimes\mathbf{r}_t\right) = \mathbf{0}.$$

Obviously, as the additional moments $E\left(\mathbf{m}-\mathbf{c}\right) = \mathbf{0}$ are exactly identified, the moment conditions that define the optimal estimator of $\boldsymbol{\delta}$ coincide with the conditions in point a), and consequently the same estimator is obtained. The optimal estimator of $\mathbf{c}$ is equal to

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{m}_t - E\left(\mathbf{mm}'\otimes\mathbf{r}'\right)\mathbf{S}^{-1}\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{m}_t\otimes\mathbf{r}_t\right),$$

with $\mathbf{m}_t$ evaluated at the optimal estimator of $\boldsymbol{\delta}$.

When $(\mathbf{r}_t,\mathbf{f}_t)$ is an i.i.d. elliptical random vector with bounded fourth moments, we can show that

$$E\left(\mathbf{mm}'\otimes\mathbf{r}'\right) = \mathcal{C}\otimes E\left(\mathbf{r}\right)', \qquad \mathcal{C} = Var\left(\mathbf{m}\right) - E\left(\mathbf{m}\right)E\left(\mathbf{m}\right)'.$$

There are two valid SDFs: $E(\mathbf{r}) = E(\mathbf{r}\mathbf{f}_1')\,\boldsymbol{\delta}_1 = E(\mathbf{r}\mathbf{f}_2')\,\boldsymbol{\delta}_2$. Hence, we can write

$$E(\mathbf{m}\mathbf{m}' \otimes \mathbf{r}') = \begin{pmatrix} \mathcal{C}_{11}E(\mathbf{r})' & \mathcal{C}_{12}E(\mathbf{r})' \\ \mathcal{C}_{12}E(\mathbf{r})' & \mathcal{C}_{22}E(\mathbf{r})' \end{pmatrix} = \begin{pmatrix} \mathcal{C}_{11}\boldsymbol{\delta}_1'E(\mathbf{r}\mathbf{f}_1')' & \mathcal{C}_{12}\boldsymbol{\delta}_2'E(\mathbf{r}\mathbf{f}_2')' \\ \mathcal{C}_{12}\boldsymbol{\delta}_1'E(\mathbf{r}\mathbf{f}_1')' & \mathcal{C}_{22}\boldsymbol{\delta}_2'E(\mathbf{r}\mathbf{f}_2')' \end{pmatrix}.$$

Let us focus on the optimal estimator of c_1. We can express it as

$$\begin{aligned} &\frac{1}{T}\sum_{t=1}^{T} m_{1t} - \begin{pmatrix} \mathcal{C}_{11}\boldsymbol{\delta}_1' & \mathcal{C}_{12}\boldsymbol{\delta}_2' \end{pmatrix} \begin{pmatrix} E(\mathbf{r}\mathbf{f}_1')' & \mathbf{0} \\ \mathbf{0} & E(\mathbf{r}\mathbf{f}_2')' \end{pmatrix} \mathbf{S}^{-1}\frac{1}{T}\sum_{t=1}^{T}(\mathbf{m}_t \otimes \mathbf{r}_t) \\ &= \frac{1}{T}\sum_{t=1}^{T} m_{1t} + \begin{pmatrix} \mathcal{C}_{11}\boldsymbol{\delta}_1' & \mathcal{C}_{12}\boldsymbol{\delta}_2' \end{pmatrix} \mathbf{D}'\mathbf{S}^{-1}\frac{1}{T}\sum_{t=1}^{T}(\mathbf{m}_t \otimes \mathbf{r}_t), \end{aligned}$$

where the second term must be zero by definition of the optimal estimator of $\boldsymbol{\delta}$. A similar argument can be applied to the optimal estimator of c_2. Thus, we can conclude that

$$\hat{\mathbf{c}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{m}_t$$

will be the optimal estimator of the SDF means in an elliptical setting.

Finally, we can easily extend our proof to $d > 2$ because the structures of $\mathbf{D}$, $\mathbf{S}$, and $\mathbf{C}$ are entirely analogous. Specifically, $\mathbf{S}$ will continue to be the same function of $\mathcal{A}$ and $\mathcal{B}$ above, although the dimension of these matrices becomes d instead of 2. In turn, $\mathbf{D}$ and $\mathbf{C}$ will remain block-diagonal, but with d blocks instead of 2 along the diagonal. Lastly, $E(\mathbf{m}\mathbf{m}' \otimes \mathbf{r}')$ will continue to be the same function of $\mathcal{C}$ above. $\square$

Intuitively, Proposition C1 states that the optimal GMM estimator in an elliptical setting is such that it prices without error the factor mimicking portfolios in any given sample. The optimal instrumental variables are defined by the Jacobian and the long-run covariance matrix of the GMM influence functions. In general, the Jacobian depends on the cross-moments between returns and factors. Under the elliptical assumption of Proposition C1, the long-run covariance matrix depends only on the first and second moments of returns on the one hand, and the first and second moments of the SDFs on the other (and their coefficient of multivariate excess kurtosis). Moreover, under the maintained hypothesis that the asset pricing model holds, we can relate the first moments of returns in that covariance matrix to the cross-moments between returns and factors. The proof above shows that these properties of the Jacobian and the long-run covariance matrix imply that the factor mimicking portfolios span the optimal "instrumental variables".

Although the elliptical family is rather broad (see Fang, Kotz and Ng (1990)), it is important to stress that (C3) will remain consistent under correct specification even if the assumptions of serial independence or a multivariate elliptical distribution do not hold in practice.

In addition, we can provide a rather different justification for (C3). Specifically, we can prove that $\mathring{\boldsymbol{\delta}}_{iT}$ in (C3) coincides with the GMM estimator that we would obtain if we used as weighting matrix the second moment of the vector of excess returns $\mathbf{r}$. In other words, $\mathring{\boldsymbol{\delta}}_{iT}$ minimizes the sample counterpart to the Hansen and Jagannathan (1997) (HJ) distance

$$E\left[\left(1-\mathbf{f}_i'\boldsymbol{\delta}_i\right)\mathbf{r}\right]'\left[E\left(\mathbf{r}\mathbf{r}'\right)\right]^{-1}E\left[\left(1-\mathbf{f}_i'\boldsymbol{\delta}_i\right)\mathbf{r}\right]$$

irrespective of the distribution of returns and the validity of the asset pricing model. The reason is that the first order condition of this minimization is

$$E\left(\mathbf{f}_i\mathbf{r}'\right)\left[E\left(\mathbf{r}\mathbf{r}'\right)\right]^{-1}E\left[\left(1-\mathbf{f}_i'\delta_i\right)\mathbf{r}\right]=\mathbf{0},$$

which is equivalent to the exact pricing of the factor mimicking portfolios in Proposition C1.

We can extend these results to the case when we want to test complete overspecification by imposing that $c_i = 0$ for $i = 1, \ldots, d$. Again, normalization-invariant procedures are crucial to avoid obtaining different results for different basis of the admissible SDF set. But given the numerical complications that they may entail, we again propose to use as starting value a computationally simple intuitive estimator that is always consistent, but which would become efficient when the returns and factors are *i.i.d.* elliptical. In fact, we can prove that the optimal estimator of the prices of risk continues to have the same structure as in Proposition C1 if we define the factor mimicking portfolios over the extended payoff space. Specifically:

Proposition C2 *If $(\mathbf{r}_t, \mathbf{f}_t)$ is an i.i.d. elliptical random vector with bounded fourth moments such that (10) holds, then the most efficient GMM estimator of $\boldsymbol{\delta}_i$ $(i = 1, \ldots, d)$ will be given by*

$$\dot{\boldsymbol{\delta}}_{iT}=\left(\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}^{+}\tilde{\mathbf{x}}_{it}^{+\prime}\right)^{-1}\sum_{t=1}^{T}\tilde{\mathbf{x}}_{it}^{+},\tag{C5}$$

where $\tilde{\mathbf{x}}_{it}^{+}$ are the relevant elements of the sample factor mimicking portfolios

$$\tilde{\mathbf{x}}_{it}^{+}=\left(\sum_{s=1}^{T}\mathbf{f}_s\mathbf{x}_s'\right)\left(\sum_{s=1}^{T}\mathbf{x}_s\mathbf{x}_s'\right)^{-1}\mathbf{x}_t.\tag{C6}$$

Proof. Once again, we develop most of the proof for the case $d = 2$ to simplify the expressions, but explain the extension to $d > 2$ at the end.

When $d = 2$, the moment conditions (10) become

$$E\left(\mathbf{m}\otimes\mathbf{x}\right)=E\begin{pmatrix}m_1\mathbf{x}\\ m_2\mathbf{x}\end{pmatrix}=E\begin{bmatrix}\left(1-\mathbf{f}_1'\boldsymbol{\delta}_1\right)\mathbf{x}\\ \left(1-\mathbf{f}_2'\boldsymbol{\delta}_2\right)\mathbf{x}\end{bmatrix}=\mathbf{0}.$$

The optimal moments correspond to the linear combinations

$$\mathbf{D}'\mathbf{S}^{-1}\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}m_{1t}\mathbf{x}_t\\ m_{2t}\mathbf{x}_t\end{pmatrix},$$

where $\mathbf{D}$ is the expected Jacobian and $\mathbf{S}$ the corresponding long-run variance. In this setting, the expected Jacobian is block-diagonal with blocks $-E\left(\mathbf{x}\mathbf{f}_i'\right)$.

When $(\mathbf{r}_t, \mathbf{f}_t)$ is an i.i.d. elliptical random vector with bounded fourth moments, and $E(\mathbf{m}) = \mathbf{0}$, we can use the results in the proof of Proposition C1 to show that the long-run covariance matrix of the influence functions will be

$$\mathbf{S} = \mathcal{A} \otimes E\left(\mathbf{x}\mathbf{x}'\right) - \mathcal{B} \otimes E(\mathbf{x})\, E(\mathbf{x})',$$

$$\mathcal{A} = (1+\kappa)\, E\left(\mathbf{m}\mathbf{m}'\right), \qquad \mathcal{B} = \kappa E\left(\mathbf{m}\mathbf{m}'\right),$$

where κ is the coefficient of multivariate excess kurtosis.

The structure of $\mathbf{D}$ and $\mathbf{S}$ is similar to their structures in the proof of Proposition C1. Therefore, we can follow the same argument to conclude that if we define the factor mimicking portfolios on the extended payoff space as

$$\mathbf{x}_i^+ = \mathbf{C}_i \mathbf{x}, \qquad \mathbf{C}_i = E\left(\mathbf{x}\mathbf{f}_i'\right)' E^{-1}\left(\mathbf{x}\mathbf{x}'\right),$$

then the sample version of the optimal moments can be written as

$$\begin{pmatrix} \mathbf{C}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 \end{pmatrix} \frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} m_{1t}\mathbf{x}_t \\ m_{2t}\mathbf{x}_t \end{pmatrix} = \frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} \mathbf{x}_{1t}^+ m_{1t} \\ \mathbf{x}_{2t}^+ m_{2t} \end{pmatrix}.$$

This expression proves that the optimal estimator of each vector of risk prices simply uses the corresponding factor mimicking portfolios. Once again, this estimator is infeasible because we do not know $\mathbf{C}_i$, but under standard regularity conditions we can replace $\mathbf{x}_{it}^+$ by its sample counterpart in (C6) without affecting the asymptotic distribution.

As in the case of Proposition C1, we can easily extend our proof to $d > 2$ because the structure of $\mathbf{D}$, $\mathbf{S}$, and $\mathbf{C}$ is entirely analogous. Specifically, $\mathbf{S}$ will continue to be the same function of $\mathcal{A}$ and $\mathcal{B}$ above, although the dimension of these matrices becomes d instead of 2. In turn, $\mathbf{D}$ and $\mathbf{C}$ will remain block-diagonal, but with d blocks instead of 2 along the diagonal. $\square$

D Monte Carlo Evidence

In this appendix, we assess the finite sample size and power properties of the testing procedures discussed above by means of several extensive Monte Carlo exercises. The exact design of our experiments is described below, and corresponds to a two-factor empirical model like the one in section 2.2, which reduces the number of variants we need to consider. Unlike in that section, though, we do not explicitly assume the existence of some underlying true factors, relying instead in the concept of HJ distance. Nevertheless, given that the number of mean, variance and correlation parameters for returns and empirical factors is large, we have simplified the data generating process (DGP) as much as possible without losing generality, so that in the end we only had to select a handful of parameters whose interpretation is very simple.

D.1 Data generating process

In this appendix, we extend the design of the single factor Monte Carlo experiment in Peñaranda and Sentana (2015) to a two-factor model. An unrestricted (i.i.d.) Gaussian data generating process (DGP) for $(\mathbf{f}, \mathbf{r})$ is

$$\mathbf{f} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

$$\mathbf{r} = \boldsymbol{\mu}_r + \mathbf{B}_r(\mathbf{f} - \boldsymbol{\mu}) + \mathbf{u}_r, \qquad \mathbf{u}_r \sim N(\mathbf{0}, \boldsymbol{\Omega}_{rr}),$$

where the $n \times 2$ matrix $\mathbf{B}_r$ is defined by the two beta vectors

$$\mathbf{B}_r = \begin{pmatrix} \beta_1 & \beta_2 \end{pmatrix}.$$

Without loss of generality, we construct the two factors so that their covariance matrix is the identity matrix. In addition, given that we use the simulated data to test that an affine function of $\mathbf{f}$ is orthogonal to $\mathbf{r}$, the only thing that matters is the linear span of $\mathbf{r}$. As a result, we can substantially reduce the number of parameters characterizing the conditional DGP for $\mathbf{r}$ by means of the following steps:

1. a Cholesky transformation of $\mathbf{r}$ which effectively sets the residual variance $\boldsymbol{\Omega}_{rr}$ equal to the identity matrix,
2. a Householder transformation that makes the second to the last entries of the vector of risk premia $\boldsymbol{\mu}_r$ equal to zero (see Householder (1964)),
3. another Householder transformation that makes the third to the last entries of β_1 equal to zero,
4. a final third Householder transformation that makes the fourth to the last entries of β_2 equal to zero.

As a result, our simplified DGP for excess returns will be

$$\mathbf{r} = \mu_r\mathbf{e}_1 + (\beta_{11}\mathbf{e}_1 + \beta_{21}\mathbf{e}_2)(f_1 - \mu_1) + (\beta_{12}\mathbf{e}_1 + \beta_{22}\mathbf{e}_2 + \beta_{32}\mathbf{e}_3)(f_2 - \mu_2) + \mathbf{u}_r,$$

$$\mathbf{u}_r \sim N(\mathbf{0}, \mathbf{I}_n),$$

where $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ are the first, second, and third columns of the identity matrix, and

$$\mathbf{f} \sim N(\boldsymbol{\mu}, \mathbf{I}_2).$$

D.2 Model restrictions

We set the values of the two elements of $\boldsymbol{\mu}$ to 1. In turn, we calibrate the six parameters that define $\mathbf{r}$ as follows. First, we define the HJ distance for this two-factor model as the minimum with respect to ϕ of the quadratic form

$$\phi'\mathbb{M}'Var^{-1}(\mathbf{r})\mathbb{M}\phi,$$

where

$$\mathbb{M}\phi = \left[\begin{array}{cc} E(\mathbf{r}) & Cov(\mathbf{r}, \mathbf{f}) \end{array} \right] \left(\begin{array}{c} c \\ \mathbf{b} \end{array} \right).$$

Note that $\mathbb{M}\phi = \mathbf{M}\boldsymbol{\theta}$ and rank($\mathbb{M}$) = rank($\mathbf{M}$), where $\mathbf{M}$ and $\boldsymbol{\theta}$ are defined in (7). Therefore, the centred SDF representation in this appendix is equivalent to the uncentred SDF used in the main text.

The 3×3 weighting matrix

$$\mathbb{W} = \mathbb{M}' Var^{-1}(\mathbf{r}) \mathbb{M}$$

$$= \left(\begin{array}{cc} E(\mathbf{r})' Var^{-1}(\mathbf{r}) E(\mathbf{r}) & E(\mathbf{r})' Var^{-1}(\mathbf{r}) Cov(\mathbf{r}, \mathbf{f}) \\ \cdot & Cov(\mathbf{r}, \mathbf{f})' Var^{-1}(\mathbf{r}) Cov(\mathbf{r}, \mathbf{f}) \end{array} \right) = \left(\begin{array}{ccc} \sigma_{00} & \sigma_{01} & \sigma_{02} \\ \cdot & \sigma_{11} & \sigma_{12} \\ \cdot & \cdot & \sigma_{22} \end{array} \right)$$

can be interpreted as the variance matrix of three noteworthy portfolios. The first one yields the maximum Sharpe ratio

$$r_0 = \mathbf{r}' Var^{-1}(\mathbf{r}) E(\mathbf{r}),$$

while the other two are the centred factor mimicking portfolios

$$r_i = \mathbf{r}' Var^{-1}(\mathbf{r}) Cov(\mathbf{r}, f_i), \qquad i = 1, 2.$$

Note that if we minimize the above quadratic form subject to the symmetric normalization $\boldsymbol{\phi}'\boldsymbol{\phi} = 1$, then this HJ distance will be equal to the minimum eigenvalue of the covariance matrix $\mathbb{W}$.

The first entry of $\mathbb{W}$ is the variance of r_0 or, equivalently, the squared maximum Sharpe ratio. The other two diagonal entries are the variances of (r_1, r_2) or, equivalently, the R^2 of their respective regressions. Finally, the three different off-diagonal elements correspond to the covariances between these three portfolios, which we can pin down by their correlations. In this way, we have six parameters that are easy to interpret and calibrate, from which we can obtain the six parameters that our DGP requires for $\mathbf{r}$.

Below we start from the free design and progressively add more and more constraints. In addition, we can interpret the constraints that the different models impose as forcing certain linear combinations of (r_0, r_1, r_2) with coefficients (c, b_1, b_2) to have zero variance. We define 3 designs (with some variants) indexed by the dimension of the subspace of prices of risk d.

- Design $d = 0$: The matrix $\mathbb{W}$ has full rank. We need to give values to the six parameters with the interpretations mentioned before, and we calibrate their values to the data. The rest of designs require constraints on the matrix $\mathbb{W}$, which we impose by means of small changes in that matrix.

- Design $d = 1$: The matrix $\mathbb{W}$ has one rank failure defined by a one-dimensional subspace of vectors (c, b_1, b_2). At least one of the factors must enter the SDF to avoid risk neutrality, so we can assume that $b_2 \neq 0$. Thus, we can choose a linear combination $(c^*, b_1^*, -1)$ with zero variance. We can achieve the same goal by expressing r_2 as

$$r_2 - \nu_2 = c^*(r_0 - \nu_0) + b_1^* (r_1 - \nu_1),$$

with $\nu_j = E(r_j)$, and changing the last column of matrix $\mathbb{W}$ to

$$\sigma_{02} = c^* \sigma_{00} + b_1^* \sigma_{01},$$
$$\sigma_{12} = c^* \sigma_{01} + b_1^* \sigma_{11},$$
$$\sigma_{22} = c^{*2} \sigma_{00} + b_1^{*2} \sigma_{11} + 2c^* b_1^* \sigma_{01}.$$

We keep the three parameters that define the covariance matrix of (r_0, r_1) equal to the values they take in design $d = 0$. This design will have two variants: one with nonzero c in the linear combination (c, b_1, b_2), and a second one with $c^* = 0$. In the former variant, we choose c^* and b_1^* to keep the same σ_{02} and σ_{22} as in the design $d = 0$. In the second variant, we chose $c^* = b_1^* = 0$, which is equivalent to an uncorrelated factor, so that $\sigma_{02} = \sigma_{12} = \sigma_{22} = 0$.

- Design $d = 2$: The matrix $\mathbb{W}$ has two rank failures defined by a two-dimensional subspace of vectors (c, b_1, b_2). We maintain the linear combination $(c^*, b_1^*, -1)$ with zero variance from design $d = 1$, and add a second linear combination $(c^{**}, -1, 0)$ with zero variance. Equivalently, we can express r_1 as

$$r_1 - \nu_1 = c^{**} (r_0 - \nu_0),$$

and modify the matrix $\mathbb{W}$ accordingly

$$\sigma_{01} = c^{**} \sigma_{00},$$
$$\sigma_{11} = c^{**2} \sigma_{00},$$

with $(\sigma_{02}, \sigma_{12}, \sigma_{22})$ satisfying the same equations as in design $d = 1$. We keep σ_{00} equal to the value in design $d = 0$. This design will again have two variants: one with nonzero c in the linear combinations (c, b_1, b_2), and a second one with $c^* = c^{**} = 0$. In the former variant, we choose c^{**} to keep the same σ_{11} as in the design $d = 0$. In the second variant, we have two uncorrelated factors, and hence all entries of $\mathbb{W}$ except σ_{00} are equal to 0.

D.3 Numerical details

We calibrated the values of the parameters to some of the datasets mentioned in the empirical section, and they are available upon request. We use $n = 6$ and $T = 200$. This number of test assets coincides with the dimension of the simplest version of the Fama-French portfolios sorted acording to size and value, while the sample size represents fifty years of quarterly data, as in Yogo (2006). Further, we also run simulations with $T = 600$, which corresponds to fifty years of monthly data, as in our evaluation of the Jagannathan and Wang (1996) model. In all instances, we simulate $10,000$ samples for each design.

The main practical difficulty is that we have to rely on numerical optimization methods to maximize the non-linear CU-GMM criterion function even though the moment conditions are linear in the parameters. For that reason, we compute the criterion by means of the auxiliary OLS regressions described in appendix B of Peñaranda and Sentana (2012). We achieve substantial gains in numerical reliability by using the consistent estimators in Propositions C1 and C2 as starting values.

Given that single-step methods are invariant to different parametrizations of the SDF, we use the uncentered version in (C1) because it is the most parsimonious in terms of parameters. Nevertheless, one could exploit the numerical equivalence of the different approaches mentioned in section 4.1, as well as the different normalizations, to check that a global minimum has been reached.

In view of the exactly identified nature of the moment conditions (C2), further speed gains can be achieved by minimizing the original moment conditions (C1) with respect to $\delta_1, \ldots, \delta_d$ only. Once this is done, the joint criterion function can be minimized with respect to $c_1, \ldots, c_d$ only, keeping $\delta_1, \ldots, \delta_d$ fixed at their CUEs and using the sample means of the estimated SDF basis as consistent starting values.

D.4 Two-dimensional set of admissible SDFs

Table D1 displays the rejection rates of the J and DM tests when there is a two-dimensional set of admissible SDFs. In our two factor setting, this means that any of the factors can price the cross-section of returns on its own. Our standard asymptotic theory implies that we expect rejection rates close to size for the J test for $d = 2$. In contrast, the usual J test for $d = 1$ should under-reject because of its generic lack of identification. The only exception arises when $c = 0$, in which case there will be a unique linear combination of the factors that yields an admissible SDF with zero mean, even though the two SDFs that we use in this design have nonzero means. Thus, the J test for $d = 1$ that imposes a zero SDF mean should yield rejection rates close to size too.

Panel A reports the rejection rates when most SDFs in the admissible set have nonzero means, while Panel B shows the corresponding figures when the asset pricing model is completely overspecified. To achieve this, we use two factors that are uncorrelated with the cross-section of returns as the DGP of Panel B.

In each panel, we report the Monte Carlo rejection rates for 6 tests: the J tests for $d = 2$ and $d = 1$, their variants restricted to have zero SDF means, and the corresponding DM tests.

(Table D1: Rejection rates for a two-dimensional set of admissible SDFs ($T = 200$))

The first result we can see in Panel A of Table 5 is that the J test for $d = 2$ performs well, showing only a slight overrejection under the null, and considerable power against $c = 0$. As expected, the J test for $d = 1$ massively under-rejects when we do not impose the restriction that $c = 0$, while it has rejection rates close to size if we do.

On the other hand, Panel B of Table D1 confirms that the J test for $d = 2$ underrejects, the restricted J test performs well, with only a slight overrejection, and the corresponding DM test overrejects. This last overrejection indicates that, if this DM test does not reject in our empirical application with quarterly data, it is not due to lack of power. In that respect, Table D2 shows that this DM test no longer shows any noticeable size distortions for $T = 600$.

(Table D2: Rejection rates for a two-dimensional set of admissible SDFs ($T = 600$))

D.5 One-dimensional set of admissible SDFs

Table D3 displays the rejection rates of the J and DM tests when the empirical model contains only one (up to scale) admissible SDF. In that case, we expect that the J test for $d = 1$ yields rejection rates close to size, while the J test for $d = 2$ should now show substantial power.

Once again, Panel A contains the rejection rates when the SDF has a nonzero mean, while Panel B reports the corresponding figures when the model is overspecified. To achieve this, we impose that one the factors is uncorrelated with the cross-section of returns as the DGP of Panel B. This is the well-known case of a useless factor.

(Table D3: Rejection rates for a one-dimensional set of admissible SDFs ($T = 200$))

As expected, Panel A of Table D3 confirms that the J test for $d = 1$ performs well while the J test for $d = 2$ has power indeed. Therefore, our finding of an overspecified model in the empirical application with quarterly data cannot be due to lack of power of this second test.

In Panel B of Table D3, the J test for $d = 2$ shows considerable power. Further, the J test for $d = 1$ underrejects, the restricted J test performs well, and the corresponding DM test overrejects. As in the previous section, Table D4 shows that this DM test no longer shows any noticeable size distortions for $T = 600$.

(Table D4: Rejection rates for a one-dimensional set of admissible SDFs ($T = 600$))

Finally, we also simulated a design where the admissible set of SDFs consists of the trivial element $m = 0$. In this case, all the tests that we study should reject their respective null hypotheses. Our results, which are available upon request, confirm the power of our proposed procedures under such a design.

Additional references

Fang, K.-T., S. Kotz and K.-W. Ng (1990): *Symmetric multivariate and related distributions*, Chapman and Hall.

Hillier, G.H. (1990): "On the normalization of structural equations: properties of direct estimators", *Econometrica* 58, 1181-1194.

Householder, A.S. (1964): *The theory of matrices in numerical analysis*, Blaisdell Publishing Co.

Peñaranda, F. and E. Sentana (2012): "Spanning tests in portfolio and stochastic discount factor mean-variance frontiers: a unifying approach", *Journal of Econometrics* 170, 303-324.

Table D1: Rejection rates for a two-dimensional set of admissible SDFs ($T = 200$)

	Nominal size		
	10	5	1
Panel A. Some SDFs have nonzero mean			
J d=2	13.65	7.03	1.61
J d=2, c=0	99.62	99.62	99.62
DM c=0	99.62	99.62	99.62
J d=1	1.03	0.37	0.01
J d=1 c=0	11.78	3.07	1.20
DM c=0	39.97	28.01	10.89
Panel B. All SDFs have zero mean			
J d=2	8.97	4.49	0.71
J d=2, c=0	14.26	7.73	1.72
DM c=0	21.46	13.69	4.34
J d=1	0.75	0.15	0.00
J d=1, c=0	0.89	0.31	0.00
DM c=0	8.03	3.37	0.30

Note: This table displays the rejection rates of CU J tests, their variants restricted to zero SDF means, and the corresponding DM tests, as described in Section 3. The rates are shown in percentage for the asymptotic critical values at 10, 5, and 1%. 10,000 samples of 6 excess returns are simulated under two variants of a two-dimensional set of admissible SDFs. Panel A reports the results for the first variant, when most of these SDFs have nonzero means, and Panel B reports the results for the second variant, when the asset pricing model is completely overspecified.

Table D2: Rejection rates for a two-dimensional set of admissible SDFs ($T = 600$)

	Nominal size		
	10	5	1
Panel A. Some SDFs have nonzero mean			
J d=2	10.80	5.55	1.13
J d=2, c=0	99.55	99.55	99.55
DM c=0	99.55	99.55	99.55
J d=1	0.78	0.16	0.00
J d=1 c=0	10.36	5.20	1.23
DM c=0	37.94	26.15	9.82
Panel B. All SDFs have zero mean			
J d=2	10.15	5.03	0.99
J d=2, c=0	11.58	5.98	1.28
DM c=0	13.63	7.58	1.63
J d=1	0.77	0.16	0.00
J d=1, c=0	0.82	0.13	0.00
DM c=0	5.94	2.09	0.19

Note: This table displays the rejection rates of CU J tests, their variants restricted to zero SDF means, and the corresponding DM tests, as described in Section 3. The rates are shown in percentage for the asymptotic critical values at 10, 5, and 1%. 10,000 samples of 6 excess returns are simulated under two variants of a two-dimensional set of admissible SDFs. Panel A reports the results for the first variant, when most of these SDFs have nonzero means, and Panel B reports the results for the second variant, when the asset pricing model is completely overspecified.

Table D3: Rejection rates for a one-dimensional set of admissible SDFs ($T = 200$)

	Nominal size		
	10	5	1
Panel A. Some SDFs have nonzero mean			
J d=2	99.17	98.03	92.79
J d=2, c=0	99.99	99.99	99.99
DM c=0	99.97	99.97	99.97
J d=1	10.06	4.99	0.98
J d=1 c=0	97.92	95.39	84.25
DM c=0	99.24	98.71	95.57
Panel B. All SDFs have zero mean			
J d=2	68.55	56.95	33.30
J d=2, c=0	99.85	99.85	99.85
DM c=0	99.84	99.84	99.84
J d=1	6.29	2.66	0.33
J d=1, c=0	11.65	5.96	1.20
DM c=0	20.51	12.92	3.88

Note: This table displays the rejection rates of CU J tests, their variants restricted to zero SDF means, and the corresponding DM tests, as described in Section 3. The rates are shown in percentage for the asymptotic critical values at 10, 5, and 1%. 10,000 samples of 6 excess returns are simulated under two variants of a one-dimensional set of admissible SDFs. Panel A reports the results for the first variant, when the SDF has a nonzero mean, and Panel B reports the results for the second variant, when the asset pricing model is overspecified.

Table D4: Rejection rates for a one-dimensional set of admissible SDFs ($T = 600$)

	Nominal size		
	10	5	1
Panel A. Some SDFs have nonzero mean			
J d=2	100	100	100
J d=2, c=0	100	100	100
DM c=0	100	100	100
J d=1	9.94	5.00	1.06
J d=1 c=0	100	100	100
DM c=0	100	100	100
Panel B. All SDFs have zero mean			
J d=2	99.06	97.97	92.69
J d=2, c=0	100	100	100
DM c=0	100	100	100
J d=1	8.87	4.26	0.82
J d=1, c=0	10.50	5.25	1.11
DM c=0	12.79	6.63	1.60

Note: This table displays the rejection rates of CU J tests, their variants restricted to zero SDF means, and the corresponding DM tests, as described in Section 3. The rates are shown in percentage for the asymptotic critical values at 10, 5, and 1%. 10,000 samples of 6 excess returns are simulated under two variants of a one-dimensional set of admissible SDFs. Panel A reports the results for the first variant, when the SDF has a nonzero mean, and Panel B reports the results for the second variant, when the asset pricing model is overspecified.

Figure B1: One asset

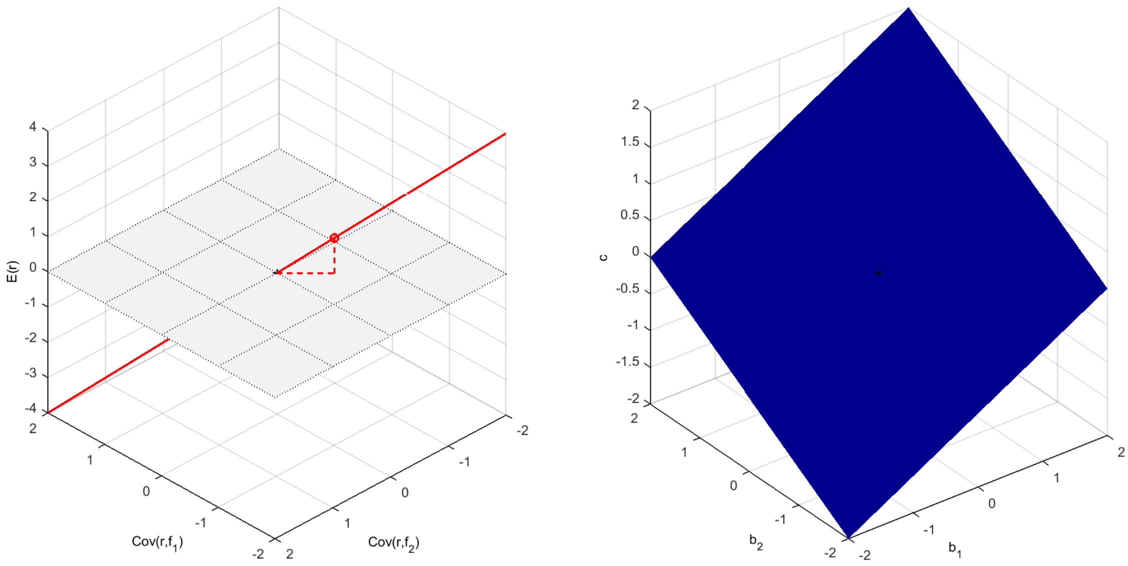


Figure B2: Two assets

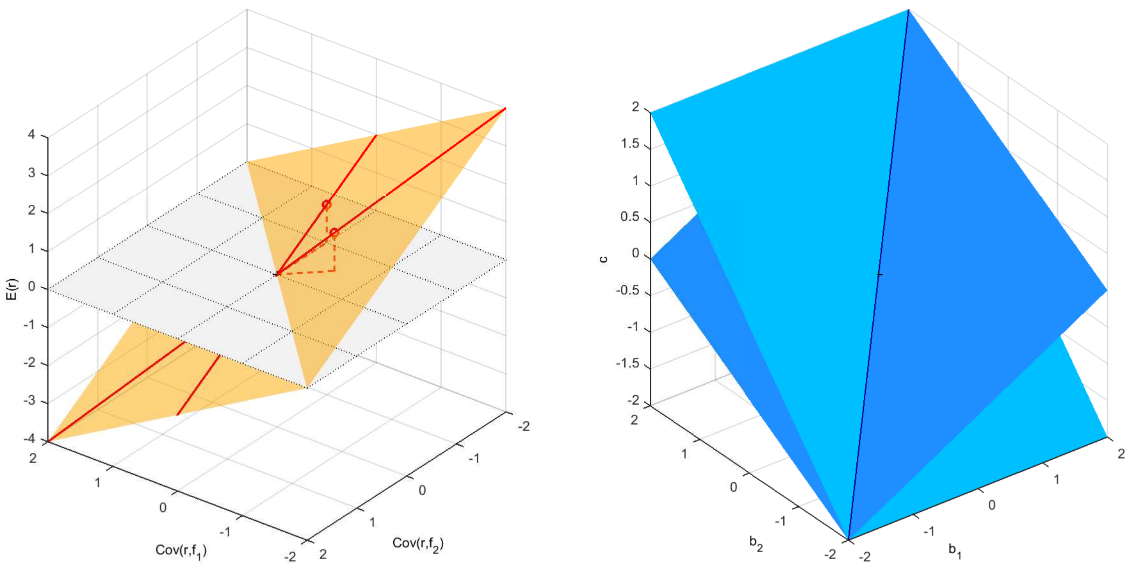


Figure B3: Three segmented asset markets

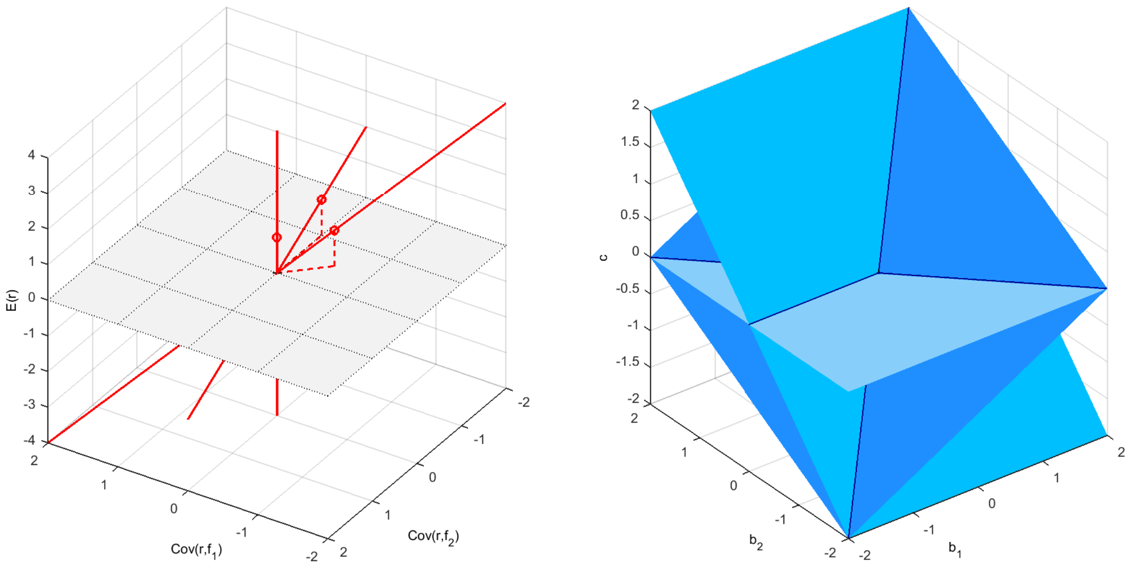


Figure B4: Three integrated asset markets

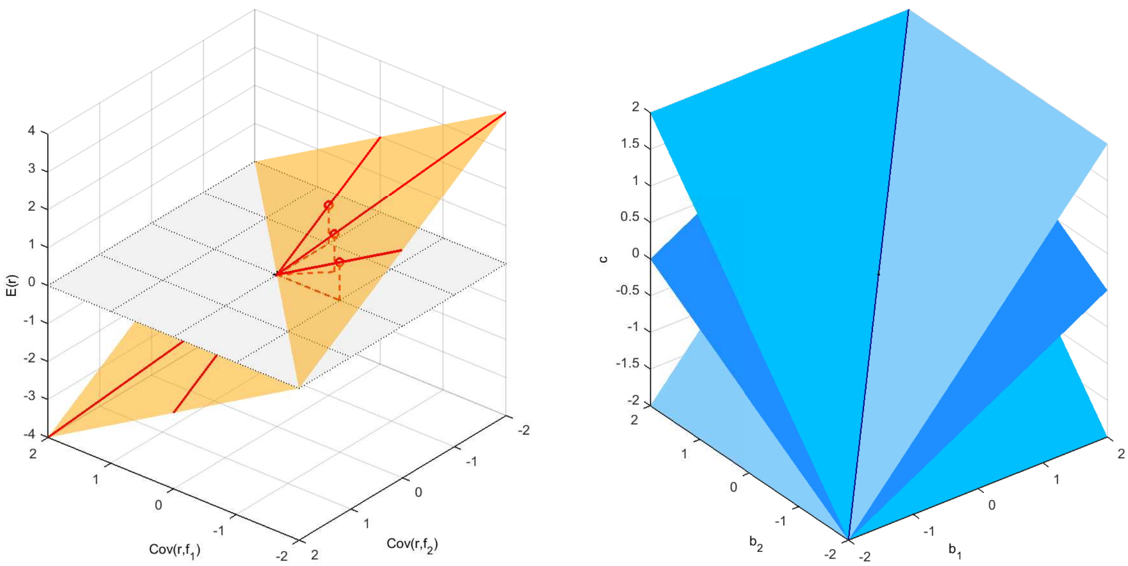


Figure B5: An unpriced second factor

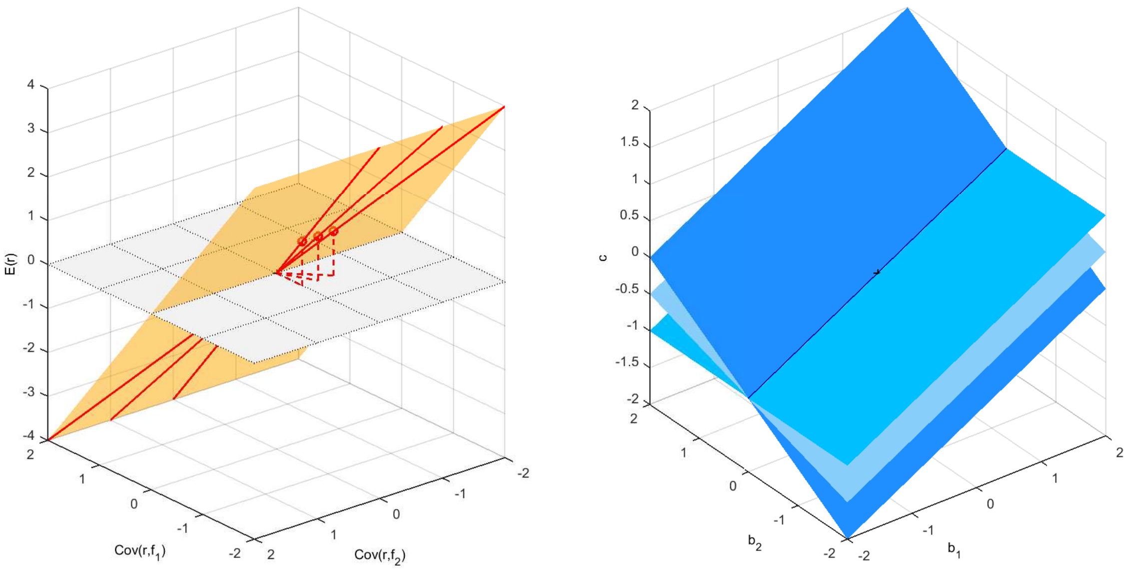


Figure B6: Two single factor models

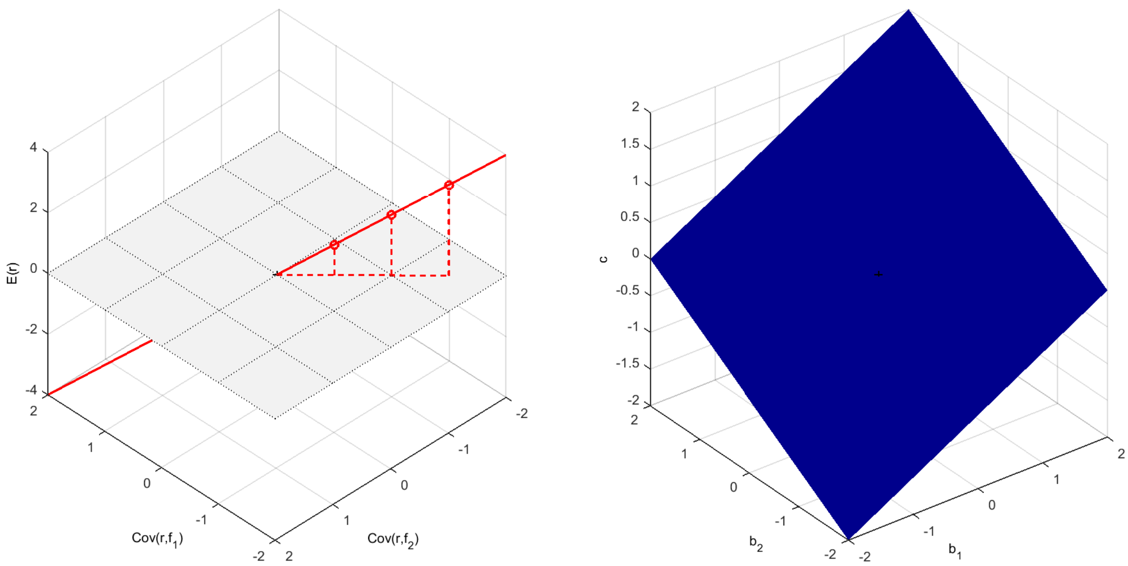


Figure B7: Valid and attractive model with a useless factor

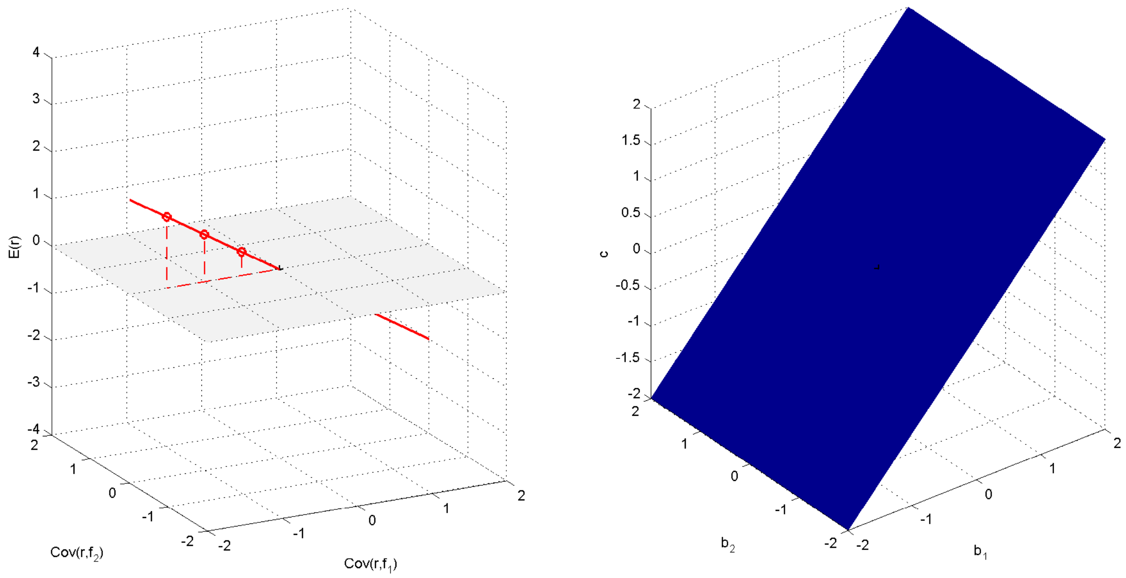


Figure B8: Valid but unattractive model with a useless factor

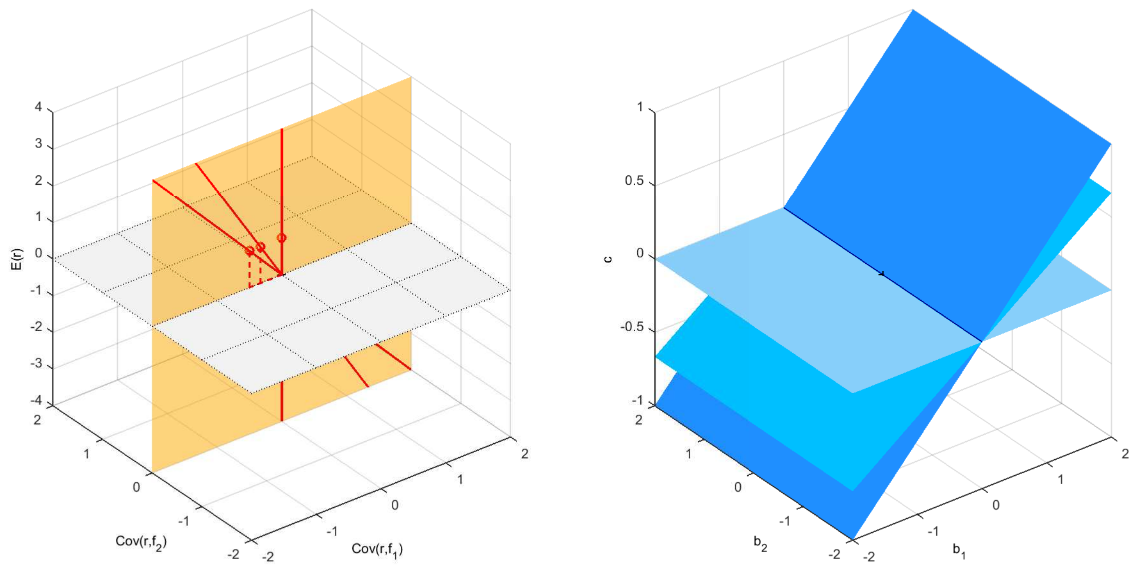


Figure B9: Two useless factors

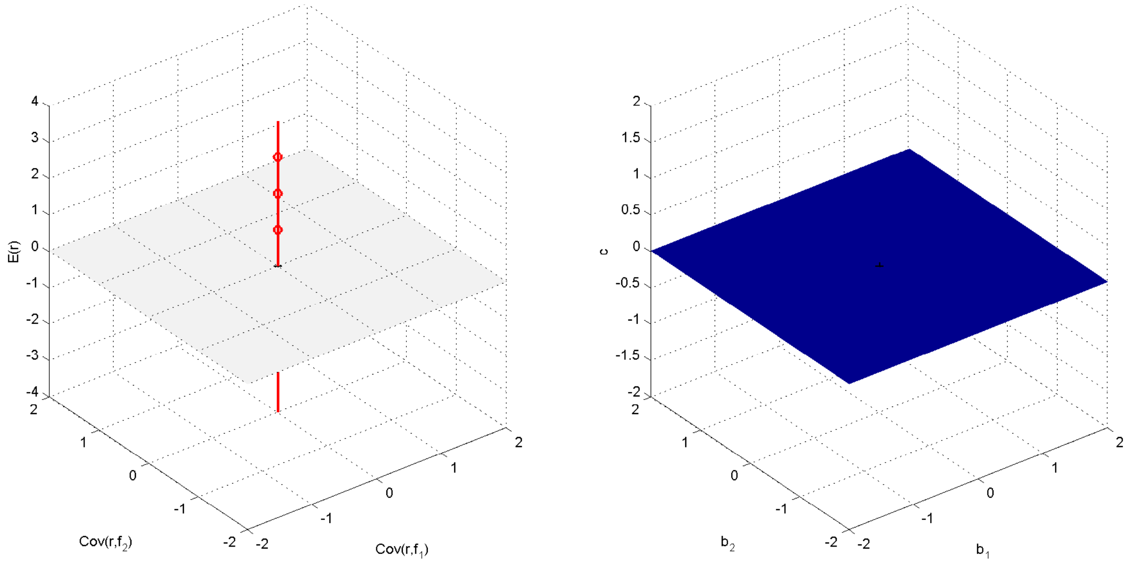


Figure C1: Normalizations