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: September 2022

B A geometric interpretation of admissible SDF sets

B.1 Taxonomy of overspecification

It is pedagogically convenient to visualize the restrictions that a linear factor pricing model such as (1) imposes on the parameters $(a, \mathbf{b}, c)$. To do so, we repeat the analysis in section 2.2 assuming that the empirical researcher considers

$$m = a + b_p f_p + b_c f_c. \tag{B1}$$

These two pricing factors (f_p, f_c) can be motivated by a consumption CAPM with Epstein-Zin preferences, which correspond to the first two factors in the empirical SDF (3). Once again, let us begin by assuming that risk premia are given by the CAPM (4). The pricing errors of the empirical model (B1) would be

$$E(m\mathbf{r}) = \boldsymbol{\sigma}_p[\tau_p(a + \mu_p b_p + \mu_c b_c) + b_p] + \boldsymbol{\sigma}_c b_c, \tag{B2}$$

where μ_p and μ_c denote the population means of the empirical factors.

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 $\boldsymbol{\sigma}_c = \boldsymbol{\sigma}_p\kappa_{cp}$ so that the factor mimicking portfolios of f_c and f_p are proportional, and consequently both the CCAPM and the traditional CAPM will give rise to the same risk premia. Obviously, the (linearized) empirical counterparts of these two models will provide admissible SDFs (namely, $a_c\left[1 - (\kappa_{cp} + \tau_p\mu_c)^{-1}\tau_p f_c\right]$ 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 $f_c^* = f_c - \kappa_{cp} f_p$ and its mean $\mu_c^* = \mu_c - \kappa_{cp}\mu_p$, 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 an additional risk factor, f_s, as in the ICAPM (6). In this case, the pricing errors of the empirical model (B1) would be

$$E(m\mathbf{r}) = \boldsymbol{\sigma}_p[\tau_p(a + \mu_p b_p + \mu_c b_c) + b_p] + \boldsymbol{\sigma}_s\tau_s(a + \mu_p b_p + \mu_c b_c) + \boldsymbol{\sigma}_c b_c. \tag{B3}$$

Therefore, the moment conditions (1) will not be satisfied unless $\boldsymbol{\sigma}_c = \boldsymbol{\sigma}_p\kappa_{cp} + \boldsymbol{\sigma}_s\kappa_{cs}$. 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. 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 $\boldsymbol{\sigma}_c = \boldsymbol{\sigma}_p \kappa_{cp} + \boldsymbol{\sigma}_s \kappa_{sp}$ with $(\kappa_{cp}, \kappa_{cs})$ being the projection coefficients of f_c onto the true factors.

In this context, the value of κ_{cs} makes a big difference. If $\kappa_{cs} \neq 0$, the moment conditions (1) will be satisfied because the SDF specification in (B1) gives rise to an admissible empirical model perfectly compatible with the risk premia in (6).

Things are rather different when $\kappa_{cs} = 0$. Substituting $\boldsymbol{\sigma}_c = \boldsymbol{\sigma}_p \kappa_{cp}$ into the pricing errors of the empirical model (B3) immediately shows that the unique (up to scale) solution of the resulting system of linear equations will satisfy $b_p + \kappa_{cp} b_c = 0$ and $a + b_c \mu_c = 0$. Thus, the admissible empirical SDFs (B1) will be proportional to $f_c - \mu_c$, in marked contrast with the true model (6). This example provides a useful generalization of the useless factor example put forward by Kan and Zhang (1999) among others, who implicitly assume that $\kappa_{cp} = \kappa_{cs} = 0$ 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 (normalized) prices of risk are econometrically point identified.

Finally, we could have complete overspecification if the empirical researcher uses two other 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.

B.2 Geometric interpretation

Let us now turn to the geometric interpretation of the cases in the previous section, using $f_1 = f_p$ and $f_2 = f_c$.

Given (B1), the matrix $\mathbf{M}$ in (12) can then be expressed as

$$\mathbf{M} = [\ E(\mathbf{r}) \quad E(\mathbf{r}f_1) \quad E(\mathbf{r}f_2)\ ],$$

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. This situation corresponds to the Epstein-Zin empirical specification (B1) when the true model is the ICAPM in (6) but the factor mimicking portfolio for consumption growth is not spanned by the market and the factor mimicking portfolio for the state variable, in which case the pricing errors will be given by (B3).

(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,

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

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)

A different example in which the empirical Epstein - Zin specification (B1) holds arises when the true model is the CAPM in (4) but the market portfolio is not proportional to the mimicking portfolio for consumption growth, so that

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

An interesting feature of this example is that consumption growth does not appear in any admissible SDF. We discuss 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)

Let us now turn to situations with overspecification. 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 $E(\mathbf{r}) = E(\mathbf{r}f_2)\delta_2$, so that the two factor mimicking portfolios are proportional. 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 (B1), 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 admissible set will have a meaningful economic interpretation. Thus, the empirical model would be econometrically underidentified but only partially overspecified.

(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

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

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

(Figure B7: Admissible 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 as in the ICAPM (6), 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 yield zero pricing errors and have a meaningful economic interpretation. This is the usual example of a useless factor.

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 that simply scales $f_2 - \mu_2$ 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: Admissible 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 be admissible. The final example in the previous section provides an illustration of this situation 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

C.1 Normalizations

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, although the former is incompatible with $H_0 : c = 0$. Figure C1 illustrates the role of these normalizations in pinning down a single point on $(\mathbf{b}, c)$ space with 2 factors.

(Figure C1: Normalizations)

Similarly, the extended system of moment conditions (13) and (14) also requires normalizations. Although any asymmetric normalization may 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 single step estimators is that our inferences, including the DM tests, will be numerically invariant to the chosen normalization.

For 2-step and iterated methods, the most convenient normalizations are the asymmetric ones $a_i = 1$ $(i = 1, ..., d)$, because they make the moment conditions (13) and (14) linear in parameters, which leads to closed-form solutions to the first-order conditions, as illustrated in Propositions C1 and C2 below. In addition, the results in Newey and West (1987) imply that the Wald, Lagrange Multiplier and DM tests of linear homogeneous restrictions such as $H_0 : c_i = 0$ will be numerically identical for multi-step methods, as long as the GMM estimators of the restricted and unrestricted moments share the same weighting matrix. In this respect, our 2-step and iterated DM tests rely on the optimal weighting matrix under the null using the estimators in Proposition C2 as starting values. Given the fast convergence, we systematically stopped the calculations after 50 iterations.

In contrast, single-step methods involve a non-linear optimization procedure even when the moment conditions are linear in parameters. 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.

C.2 Efficient GMM estimation with elliptical distributions

C.2.1 Without complete overspecification

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 (13) 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 (13) as

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

and (14) 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 of each c_i is the sample mean of the corresponding SDF.

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

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

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

C.2.2 With complete overspecification

We can extend the previous 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 spanned by $\mathbf{x} = (\mathbf{r}', 1)'$. Specifically:

Proposition C2 *If* $(\mathbf{r}_t, \mathbf{f}_t)$ *is an i.i.d. elliptical random vector with bounded fourth moments such that (15) 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*

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

D Proofs

In the proofs of Propositions 1 and A1, 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 one uses the optimal weighting matrix. First, the inclusion of s additional unrestricted moment conditions 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 continuously differentiable bijective reparametrizations whose Jacobian matrix has full row rank in an open neighborhood 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.

D.1 Proposition 1

We find it convenient to express the pricing conditions (1) in terms of central moments in (16), 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 (16) to deal with a $d-$dimensional subspace of admissible SDFs. Thus, the centred SDF counterpart to (13) will be based on the moment conditions

$$E \begin{pmatrix} \mathbf{r} m_1 \\ \vdots \\ \mathbf{r} m_d \\ \mathbf{f} - \boldsymbol{\mu} \end{pmatrix} = \mathbf{0}, \qquad m_i = c_i + (\mathbf{f} - \boldsymbol{\mu})' \mathbf{b}_i, \tag{D1}$$

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 (D1). 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 (13) and (14).

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

D.2 Proposition A1

Let us start with the simple case of $d = 1$. The addition of the pricing of R in (A1) 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 (A1) 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. □

D.3 Proposition C1

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 $E(\mathbf{r}) = E(\mathbf{rf}_1')\boldsymbol{\delta}_1 = E(\mathbf{rf}_2')\boldsymbol{\delta}_2$, we can write these matrices as

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

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

In addition, let us define the matrices $\mathbf{Q}_i$ such that $E(\mathbf{rf}_1') = E(\mathbf{rf}_2')\mathbf{Q}_1$ and $E(\mathbf{rf}_2') = E(\mathbf{rf}_1')\mathbf{Q}_2$, 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 $E(\mathbf{rf}')$ together with the full column rank of $E(\mathbf{rf}_1')$ and $E(\mathbf{rf}_2')$. Thus, we can write

$$\mathbf{SC}' = \mathbf{D}\mathbb{Q},$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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, $E(\mathbf{mm}' \otimes \mathbf{r}')$ will continue to be the same function of $\mathcal{C}$ above. $\square$

D.4 Proposition C2

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 (15) become

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

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 $-E(\mathbf{xf}_i')$.

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

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

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

where κ is the coefficient of multivariate excess kurtosis.

The structure of $\mathbf{D}$ and $\mathbf{S}$ is similar to the structure of those matrices 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

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

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. $\square$

E Additional empirical results

E.1 Yogo's (2006) estimated risk premia with iterated GMM

Figure E1 reproduces the seeming alignment of the risk premia in the data with the risk premia generated by Yogo's (2006) model using exactly the estimation procedure based of the centred SDF moments (16) with the normalization $c = 1$ that he used.

(Figure E1: Risk premia from 2S-GMM)

In addition to the theoretical considerations we have discussed in section 5.1, we found that his results are sensitive to his choice of estimation method (2-step GMM) and the imposition of restrictions on the prices of risk. Specifically, if we use instead iterated GMM starting from the 2-step estimates, we encounter a cycle with four different solutions.

(Figure E2: Risk premia from IT-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 E3: Risk premia from IT-GMM, free coefficients)

These discrepancies highlight the advantages of the single-step GMM estimation procedures that we use with the uncentred SDF moment conditions (12), but they might also be a sign of overspecification.

E.2 Evaluation of submodels

In this section, we report the results of analyzing the different empirical asset pricing models associated to the basis of the space of admissible SDFs as if they were empirical models on their own.

Specifically, in the case of the original Yogo (2006) data, Table E1 reports the separate evaluation of each submodel in the second and third blocks of columns of Tables 1 and 2. As can be seen, we find that the SDFs that correspond to the two versions of the Epstein-Zin model are uncorrelated with the cross-section of asset returns when $d = 2$, which is in line with our simultaneous results in Table 1. In addition, we find that the traditional CAPM is clearly rejected when $d = 3$, while each of the consumption factors appears to be useless. In this respect, the $R^{2\prime}s$ in the regressions of each factor onto the vector of excess returns are 0.983, 0.099 and 0.177 for the market portfolio, durable and nondurable consumption, respectively.

(Table E1: Submodels of Yogo model 1951-2001)

We repeat the same exercise for the Jagannathan-Wang (1999) mode analyzed in Tables 5 and 6. When $d = 2$, the results in Panel A of Table E2 indicate that the two submodels that we use as a basis to characterize the identified set of admissible SDFs are economically meaningless when we focus on size and book-to-market sorted portfolios. Similarly, we find that the traditional CAPM is clearly rejected when $d = 3$, while the additional factors (labor income and default premium) appear to be useless on their own. In contrast, in Panel B we only find one uncorrelated two-factor model when we add industry portfolios because the correlation of their returns with labor income is statistically significant.

(Table E2: Submodels of Jagannathan-Wang model 1959-2012)

We would like to emphasize that most of these submodel results can be inferred directly from the results in section 5. For example, the conclusions about the Jagannathan-Wang model with industry portfolios follow from the fact that our methodology pins down a one-dimensional set of admissible SDFs that is uncorrelated with the cross-section in which only the coefficient of the default premium is statistically significant (see Panel B of Table 5). Therefore, although this model is econometrically identified, the fact that it is not rejected is due to a useless factor: the default premium.

Looking at each individual submodel separately, though, substantially complicates inferences, as the number of simultaneous tests increases very quickly, which in turn increases the chances of falsely rejecting one of the multiple null hypotheses. For that reason, we recommend using the simultaneous procedures in the main text.

E.3 Fama and French 3-factor model

Next, we apply our proposed methodology to the popular Fama-French 3-factor model, whose pricing factors are all traded. As is well known, the factors are the market portfolio and two portfolios that aim to capture the size and value effects; see Fama and French (1993) for details. When we use the quarterly data in section 5.2, we find that the J statistics associated to a one-dimensional set are 60.55 and 39.53 for the 25 size- and value-sorted portfolios and the 11 sorted and industry portfolios, respectively, whose p-values are very close to zero. Similarly, the corresponding J statistics for two-dimensional SDF sets reject their null hypothesis too. In addition, the rank test of Proposition 1 has a zero p-value in all cases.

We obtain entirely analogous results when we consider the monthly data in section 5.3. Therefore, the problem with this model is neither overspecification nor underidentification, but rather lack of admissible SDFs.

F Monte Carlo Evidence

In this appendix, we assess the finite sample size and power properties of the testing procedures we have discussed in the main text by means of several extensive Monte Carlo exercises. The exact design of our experiments is described below, and corresponds to three-factor empirical models in section 2.2 and our empirical applications. In an earlier version (see Manresa, Peñaranda and Sentana (2017)), we present analogous results for the two-factor models in appendix B. Unlike in section 2.2, 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.

F.1 Data generating process

Consider the following unrestricted joint data generating process (DGP) for the $k+n$ random vector $(\mathbf{f}, \mathbf{r})$:

$$\mathbf{f} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{F1a}$$

$$\mathbf{r} = \boldsymbol{\mu}_r + \mathbf{B}_r(\mathbf{f} - \boldsymbol{\mu}) + \mathbf{u}_r, \qquad \mathbf{u}_r \sim N(\mathbf{0}, \boldsymbol{\Omega}_{rr}), \tag{F1b}$$

with $cov(\mathbf{f}, \mathbf{u}_r) = \mathbf{0}$, so that $\mathbf{B}_r$ is the $n \times k$ matrix of least squares projection coefficients characterized by the beta vectors

$$\mathbf{B}_r = \begin{pmatrix} \boldsymbol{\beta}_1 & \ldots & \boldsymbol{\beta}_k \end{pmatrix}.$$

By premultiplying $\mathbf{f}$ and $\boldsymbol{\mu}$ by $\boldsymbol{\Sigma}^{-1/2}$ and postmultiplying $\mathbf{B}_r$ by $\boldsymbol{\Sigma}^{1/2}$, where $\boldsymbol{\Sigma}^{1/2}$ is one of the square roots of the positive definite matrix $\boldsymbol{\Sigma}$, we can alternatively express (F1) so that the covariance matrix of the k factors is the identity matrix. In addition, given that the only thing that matters for asset pricing tests is the linear span of $\mathbf{r}$, we can substantially reduce the number of parameters characterizing the conditional DGP for $\mathbf{r}$ in (F1b) without loss of generality by premultiplying $\mathbf{r}$, $\boldsymbol{\mu}_r$ and $\mathbf{B}_r$ by $\boldsymbol{\Omega}_{rr}^{-1/2}$, where $\boldsymbol{\Omega}_{rr}^{1/2}$ is one of the square roots of $\boldsymbol{\Omega}_{rr}$, so that the residual covariance matrix becomes the identity matrix. In this respect, note that positive definite matrices of dimension higher than 1 have a continuum of square root matrices, which are all orthogonal transformations of each other, the usual lower triangular Cholesky matrix being just one such example.

Next, we can exploit the singular value decomposition of the matrix of regression coefficients of the resulting system, $\boldsymbol{\Omega}_{rr}^{-1/2}\mathbf{B}_r\boldsymbol{\Sigma}^{1/2} = \mathbf{B}_r^* = \mathbf{U}^*\boldsymbol{\Lambda}^*\mathbf{V}^{*\prime}$, where $\mathbf{U}^*$ and $\mathbf{V}^*$ are orthonormal matrices of dimensions n and k, respectively, and $\boldsymbol{\Lambda}^*$ is an $n \times k$ matrix in which all the elements except the k along its main diagonal are 0. Specifically, if we further premultiply the assets by $\mathbf{U}^{*\prime}$ and the factors by $\mathbf{V}^{*\prime}$, we end up with a version of (F1b) in which the only non-zero betas of the n portfolios on the k risk factors will appear in positions $(1,1)\ldots,(k,k)$, so that both the true factors and their mimicking portfolios will now be orthogonal to each other.

Finally, we can further premultiply the returns on the resulting portfolios by a bordered Householder matrix (Householder, 1964) that leaves the k mimicking portfolios unchanged but sets to 0 the risk premia of portfolios $k+2,\ldots,n$, which nevertheless not only continue to have zero betas but also remain uncorrelated to the mimicking portfolios because Household matrices are orthonormal. Thus, the risk premia of the first three assets will reflect the risk premia of the factor mimicking portfolios while the risk premia of the $k+1$ asset, which also has zero betas and is orthogonal to the rest by construction, will fully characterize the mispricing of the original set of test assets by those factors. As we explain in the next section, this mispricing is very closely related to the Hansen - Jagannathan (1994) distance.

As a result, we can use without loss of generality the following simplified DGP for excess returns

$$\mathbf{r} = \mu_{r1}\mathbf{e}_1 + \mu_{r2}\mathbf{e}_2 + \mu_{r3}\mathbf{e}_3 + \mu_{r4}\mathbf{e}_4 + \beta_{11}\mathbf{e}_1 (f_1 - \mu_1) + \beta_{22}\mathbf{e}_2 (f_2 - \mu_2) + \beta_{33}\mathbf{e}_3 (f_3 - \mu_3) + \mathbf{u}_r,$$

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

where the vectors $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4)$ are the first four columns of the identity matrix, and

$$\mathbf{f} \sim N(\boldsymbol{\mu}, \mathbf{I}_3).$$

F.2 Calibration of first and second moments

We set the values of the three elements of $\boldsymbol{\mu}$ to 1. In turn, we calibrate the parameters that define $\mathbf{r}$ as follows. First, we define the (squared) HJ distance for this three-factor model as the minimum with respect to (a normalized version of) $\boldsymbol{\phi}$ of the quadratic form

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

where

$$\mathbb{M}\boldsymbol{\phi} = \begin{bmatrix} E(\mathbf{r}) & Cov(\mathbf{r}, \mathbf{f}) \end{bmatrix} \begin{pmatrix} c \\ \mathbf{b} \end{pmatrix}.$$

Note that $\mathbb{M}\boldsymbol{\phi} = \mathbf{M}\boldsymbol{\theta}$ for the appropriate $\boldsymbol{\theta}$ and rank($\mathbb{M}$) = rank($\mathbf{M}$), where $\mathbf{M}$ and $\boldsymbol{\theta}$ are defined in (12). Therefore, the centred SDF representation in this appendix is equivalent to the uncentred SDF used in the main text.

The 4×4 weighting matrix

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

$$= \begin{pmatrix} E(\mathbf{r})' Var^{-1}(\mathbf{r}) E(\mathbf{r}) & E(\mathbf{r})' Var^{-1}(\mathbf{r}) Cov(\mathbf{r}, \mathbf{f}) \\ Cov(\mathbf{r}, \mathbf{f})' Var^{-1}(\mathbf{r}) E(\mathbf{r}) & Cov(\mathbf{r}, \mathbf{f})' Var^{-1}(\mathbf{r}) Cov(\mathbf{r}, \mathbf{f}) \end{pmatrix} = \begin{pmatrix} \sigma_{00} & \sigma_{01} & \sigma_{02} & \sigma_{03} \\ \sigma_{01} & \sigma_{11} & 0 & 0 \\ \sigma_{02} & 0 & \sigma_{22} & 0 \\ \sigma_{03} & 0 & 0 & \sigma_{33} \end{pmatrix}$$

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

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

while the other three are the centred factor mimicking portfolios

$$r_i = \mathbf{r}'Var^{-1}(\mathbf{r})\,Cov(\mathbf{r}, f_i), \qquad i = 1, 2, 3.$$

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

The first entry σ_{00} of $\mathbb{W}$ is the variance of r_0 or, equivalently, the squared maximum Sharpe ratio. The other three diagonal entries $(\sigma_{11}, \sigma_{22}, \sigma_{33})$ are the variances of (r_1, r_2, r_3) or, equivalently, the R^2 of their respective factor mimicking regressions. Finally, we can pin down the three covariances $(\sigma_{01}, \sigma_{02}, \sigma_{03})$ between r_0 and (r_1, r_2, r_3) by the factor mimicking portfolios' Sharpe ratios because the portfolio with the maximum Sharpe ratio is such that $Cov(r_0, r) = E(r)$ for any r. In this way, we have seven parameters that are easy to interpret and calibrate, from which we can obtain the seven parameters that our DGP requires for $\mathbf{r}$, namely $(\mu_{r1}, \mu_{r2}, \mu_{r3}, \mu_{r4})$ and $(\beta_{11}, \beta_{22}, \beta_{33})$.

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, r_3) with coefficients (c, b_1, b_2, b_3) to have zero variance. Thus, the rank of the weighting matrix $\mathbb{W}$ controls the dimension of the admissible set of SDFs. We define 4 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 seven 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, b_3). This design will have two variants: one with nonzero c in the linear combination (c, b_1, b_2, b_3), and a second one with $c = 0$. In the former variant, we make the fourth column of $\mathbb{W}$ linearly dependent from the other columns by changing a single parameter
$$\sigma_{03} = \left[\sigma_{33}\left[\sigma_{00} - \frac{\sigma_{01}^2}{\sigma_{11}} - \frac{\sigma_{02}^2}{\sigma_{22}}\right]\right]^{0.5}.$$
with respect to the design $d = 0$. In the latter variant, we make the third factor mimicking portfolio equal to zero (an uncorrelated factor) by changing two parameters
$$\sigma_{03} = \sigma_{33} = 0.$$

- Design $d = 2$: The matrix $\mathbb{W}$ has two rank failures defined by a two-dimensional subspace of vectors (c, b_1, b_2, b_3). We start from the parameters used above to impose $d = 1$ and $c = 0$, that is, $\sigma_{03} = \sigma_{33} = 0$. Once again, this design will have two variants: one with nonzero c in the additional linear combination (c, b_1, b_2, b_3), and a second one with $c = 0$. In the former variant, we make the third column of $\mathbb{W}$ linearly dependent from the first two columns by changing a single parameter
$$\sigma_{02} = \left[\sigma_{22}\left[\sigma_{00} - \frac{\sigma_{01}^2}{\sigma_{11}}\right]\right]^{0.5}.$$

In the latter variant, we make the second factor mimicking portfolio equal to zero by changing two parameters

$$\sigma_{02} = \sigma_{22} = 0.$$

Now both the second and third factors are uncorrelated with the cross-section of returns.

- Design $d = 3$: The matrix $\mathbb{W}$ has three rank failures defined by a three-dimensional subspace of vectors (c, b_1, b_2, b_3). We start from the parameters used above to impose $d = 2$ and $c = 0$. This design will also have two variants: one with nonzero c in the additional linear combination (c, b_1, b_2, b_3), and a second one with $c = 0$. In the former variant, we make the second column of $\mathbb{W}$ linearly dependent from the first column by changing a single parameter

  $$\sigma_{01} = [\sigma_{11}\sigma_{00}]^{0.5}.$$

  In the latter variant, we make the first factor mimicking portfolio equal to zero by changing two parameters

  $$\sigma_{01} = \sigma_{11} = 0.$$

  Now the three factors are useless.

Given its lack of empirical relevance, though, in the interest of space we do not report the results for $d = 3$, which are available on request. As for the $d = 0$ design, whose results are also available on request, we find that our procedures have a lot of power when the admissible set of SDFs consists of the trivial element $m = 0$ only, as expected.

In view of the fact that many empirical papers assessing linear factor pricing models rely on monthly returns, finally we have calibrated the values of the parameters to the dataset we used in section 5.3, whose exact values are available upon request. Thus, we simulate $5,000$ samples for each design with $n = 25$, $k = 3$ and $T = 660$.

F.3 Computational details

As we mentioned in appendix C, 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 explore the parameter space by computing the criterion function by means of the auxiliary OLS regressions described in appendix B of Peñaranda and Sentana (2012) using as starting values five different random perturbations of the consistent estimators in Propositions C1 and C2, together with another five different random perturbations of the consistent first-step estimators that use the identity as weighting matrix.

Given that single-step methods are invariant to different parametrizations of the SDF, we use the uncentred 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 $\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_d$ only. Once this is done, the joint criterion function can be minimized with respect to $c_1, \ldots, c_d$ only, keeping $\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_d$ fixed at their continuously updated estimates and using the sample means of the estimated SDF basis as consistent starting values.

F.4 One-dimensional set of admissible SDFs

Table F1 displays the rejection rates of the continuously updated, 2-step and iterated versions of our proposed tests when the empirical model contains only one (up to scale) admissible SDF compatible with the original moment conditions (1). Specifically, Panel A contains the rejection rates when the SDF has a nonzero mean, while Panel B reports the corresponding figures when the model is completely overspecified. As we explained in appendix F.2, we achieve complete overspecification by imposing that one of the factors is uncorrelated with the cross-section of returns, which effectively makes it a useless factor. In each panel, we report the Monte Carlo rejection rates for nine different tests: the overspecification tests for the moment conditions (13) for $d = 1$, $d = 2$, and $d = 3$, their augmented variants in (15), and the corresponding DM tests for zero SDF means.

(Table F1: Rejection rates for a one-dimensional set of admissible SDFs ($T = 660$))

For the design in Panel A, we would expect the J test for $d = 1$ to yield rejection rates close to size, while the J test for $d = 2$ and $d = 3$, as well as the J tests that additionally impose that $c = 0$ regardless of d and the associated DM tests should show substantial power. And in fact, our simulation results confirm that this is indeed the case. In addition, we find that continuously updated tests have more reliable finite sample sizes than either 2-step or iterated GMM, as expected from footnote 2.

In contrast, for the design in Panel B, we would expect the J test of (15) with $d = 1$ to yield rejection rates close to size, while the J test of (13) and (15) should show substantial power for $d = 2$ and $d = 3$. And again, our Monte Carlo results are in line with these predictions. The only noticeable result is that the DM test of $H_0 : c = 0$ when $d = 1$ is too liberal, which suggests that our finding of an overspecified model in Panel B of Table 5 cannot be attributed to these distortions. Finally, we also find that the continuously updated tests have more reliable finite sample sizes than either 2-step or iterated GMM. In addition, their pattern of rejections is in line with the results reported in Manresa, Peñaranda and Sentana (2017) despite now using 25 assets rather than 6 and three factors instead of two.

F.5 Two-dimensional set of admissible SDFs

Table F2 shares the format of Table F1 to display the rejection rates of the tests discussed in the previous section when there is a two-dimensional set of admissible SDFs compatible with the original moment conditions (1). Panel A reports those 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 force two of the factors to be uncorrelated with the cross-section of returns, as we explained in appendix F.2.

(Table F2: Rejection rates for a two-dimensional set of admissible SDFs ($T = 660$))

In Panel A, standard GMM asymptotic theory suggests that we would expect the rejection rates of the J test of the moment conditions (13) to be close to the nominal size for $d = 2$, while the same test for $d = 3$, as well as the J tests that additionally impose that $c = 0$ and the associated DM tests should display substantial power for $d \geq 2$. And while most of these predictions are confirmed by our simulations, we also observe that the continuously updated version of the J test of the moment conditions (13) is too liberal, while the 2-step and iterated versions too conservative. This suggests that the lack of rejections that we saw in the middle blocks of columns of Tables 1, 3 and 5 cannot be attributed to these distortions.

We also find that the J test of (1) massively underrejects, as one would expect from the results in Cragg and Donald (1993) because the parameters of this linear set of moment conditions are underidentified. In contrast, the J test of (15) with $d = 1$ shows rejection rates close to nominal size because there is always a single (up to scale) zero-mean linear combination of the pricing factors in this partially overspecified model. Not surprisingly, the combination of these two results implies that the DM test for $c = 0$ when $d = 1$ shows a high rejection rate.

In turn, Panel B of Table F2 reports the rejection rates when the empirical model is completely overspecified. As expected, the continuously updated version of the J test of (13) has rejection rates close to nominal size when $d = 2$, while the corresponding test of (15) and the associated DM tests are too liberal, so once again, the lack of rejections that we saw in the middle blocks of columns of Tables 1, 3 and 5 are unlikely to be due to these distortions. On the other hand, our results indicate the excessive conservative nature of the 2-step and iterated versions of these tests. Nevertheless, we find systematic rejections of the different tests for $d = 3$.

Finally, it is worth mentioning that the continuously updated versions of the J tests of (1) and (15) with $d = 1$ underreject in this design, as they should because the parameters of both sets of moment conditions become underidentified in this completely overidentified situation.

Once again, our continuously updated results are in line with the ones we reported in Manresa, Peñaranda and Sentana (2017), although not surprisingly the size distortions tend to be larger with many more assets and three factors instead of two.

F.6 Selection of the dimension of the admissible set of SDFs

Although the underidentification tests put forward by Arellano, Hansen and Sentana (2012) were not intended as the basis for a consistent estimator of the dimension of the identified SDF set, we have recycled the Monte Carlo results that we have just discussed to tentatively analyze the performance of a very simple dimension selection procedure whose rational would be as follows. In line with our discussion of the empirical tables in section 5, we may select $d = 1$ if the J tests associated to this dimension fail to reject but the corresponding tests for $d = 2$ succeed. Similarly, we could choose $d = 2$ if the J tests associated to this dimension do not reject but those for $d = 3$ do so. Finally, in our three-factor specification it would seem natural to select $d = 0$ when all those tests reject and $d = 3$ when none of them does.

We focus on the completely overspecified designs in Panels B of Tables F.1 and F.2, which seem the most relevant ones in our empirical applications. In this respect, we find that when we apply the rule described in the previous parargraph to the continuously updated versions of the overidentified tests of (15) using 1% as the threshold for the p-values, we select $d = 0$, $d = 1$, $d = 2$ and $d = 3$ with relative (%) frequencies 1.09, 98.32, 0.59 and 0, respectively, when the identified set of SDFs is of dimension 1. In turn, the corresponding relative frequencies become 0, 3.3, 96.7 and 0 when the true dimension d is 2. Therefore, our methodology shows some promise to consistently estimate the degree of underidentification of an empirical asset pricing model in practice, although further research would be necessary for different combinations of n, k and T and alternative parameter configurations.

Additional references

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.

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

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.

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.

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

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.

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.

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.

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

Manresa, E., F. Peñaranda and E. Sentana (2017): "Empirical evaluation of overspecified asset pricing models", CEMFI Working Paper 1711.

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

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.

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.

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

Table E1: Submodels of Yogo model 1951-2001

	Market, Nondur.		Market, Durables		Market		Nondur.		Durables	
CU-GMM										
Mean	-0.13	(0.17)	0.02	(0.79)	0.67	(0.00)	0.02	(0.86)	-0.02	(0.75)
Criterion	25.19	(0.34)	21.30	(0.56)	78. 72	(0.00)	35.26	(0.07)	22.47	(0.55)
Criterion0	27.11	(0.30)	21.37	(0.62)	104.24	(0.00)	35.29	(0.08)	22.56	(0.60)
2S-GMM										
Mean	0.05	(0.51)	0.06	(0.20)	0.87	(0.00)	0.13	(0.05)	0.04	(0.40)
Criterion	29.99	(0.15)	19.01	(0.70)	84.19	(0.00)	33.00	(0.10)	21.45	(0.61)
Criterion0	30.43	(0.17)	20.68	(0.66)	899.25	(0.00)	36.87	(0.06)	22.15	(0.63)
IT-GMM										
Mean	0.06	(0.41)	0.07	(0.16)	0.65	(0.00)	0.13	(0.04)	0.04	(0.36)
Criterion	30.47	(0.14)	20.18	(0.63)	43.87	(0.01)	34.97	(0.07)	22.64	(0.54)
Criterion0	31.16	(0.15)	22.19	(0.57)	131.27	(0.00)	39.33	(0.03)	23.49	(0.55)

Notes: This table displays the J and J_0 tests (with free and constrained SDF mean) with p-values in parenthesis () for each individual submodel in Tables 1 and 2. We display the results for the same normalization as in those tables, which CU-GMM is numerically invariant to. 2S-GMM and IT-GMM refer to 2-step and iterated procedures. The J tests are complemented with significance tests of a zero SDF mean. In particular, the p-value of the distance metric test of the null hypothesis of zero parameter is reported in parenthesis to the right of the estimate. 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 E2: Submodels of Jagannathan-Wang model 1959-2012

	Market, Labor		Market, Premium		Market		Labor		Premium	
Panel A. 25 size and book-to-market sorted portfolios										
CU-GMM										
Mean	-0.22	(0.09)	0.02	(0.71)	0.98	(0.00)	-0.31	(0.01)	0.03	(0.58)
Criterion	23.26	(0.45)	28.36	(0.20)	104.57	(0.00)	24.01	(0.46)	30.70	(0.16)
Criterion0	26.10	(0.35)	28.50	(0.24)	126.38	(0.00)	31.39	(0.18)	31.00	(0.19)
2S-GMM										
Mean	0.07	(0.40)	0.05	(0.16)	0.99	(0.00)	0.03	(0.74)	0.07	(0.06)
Criterion	27.65	(0.23)	28.70	(0.19)	104.43	(0.00)	36.00	(0.06)	30.81	(0.160)
Criterion0	28.35	(0.25)	30.70	(0.16)	13688.97	(0.00)	36.10	(0.07)	34.30	(0.10)
IT-GMM										
Mean	0.07	(0.42)	0.06	(0.17)	0.90	(0.00)	0.02	(0.77)	0.07	(0.08)
Criterion	26.34	(0.29)	26.95	(0.26)	40.78	(0.02)	33.26	(0.10)	28.37	(0.25)
Criterion0	27.00	(0.31)	28.81	(0.23)	371.31	(0.00)	33.34	(0.12)	31.51	(0.17)
Panel B. 6 size and book-to-market sorted portfolios, and 5 industry portfolios										
CU-GMM										
Mean	-0.10	(0.40)	0.04	(0.46)	0.98	(0.00)	-0.30	(0.03)	0.06	(0.34)
Criterion	23.18	(0.01)	15.95	(0.07)	77.14	(0.00)	27.27	(0.00)	17.82	(0.06)
Criterion0	23.90	(0.01)	16.50	(0.09)	152.61	(0.00)	32.28	(0.00)	18.73	(0.07)
2S-GMM										
Mean	0.08	(0.33)	0.07	(0.13)	0.99	(0.00)	0.08	(0.34)	0.08	(0.05)
Criterion	24.22	(0.00)	15.38	(0.08)	77.58	(0.00)	33.17	(0.00)	16.82	(0.08)
Criterion0	25.17	(0.01)	17.65	(0.06)	13288.56	(0.00)	34.09	(0.00)	20.80	(0.04)
IT-GMM										
Mean	0.08	(0.34)	0.07	(0.15)	0.93	(0.00)	0.08	(0.34)	0.08	(0.06)
Criterion	23.69	(0.01)	14.60	(0.10)	21.53	(0.02)	33.28	(0.00)	15.62	(0.11)
Criterion0	24.60	(0.01)	16.68	(0.08)	262.01	(0.00)	34.21	(0.00)	19.04	(0.06)

Note: This table displays the J and J_0 tests (with free and constrained SDF mean) with p-values in parenthesis () for each individual submodel in Tables 5 and 6. We display the results for the same normalization as in those tables, which CU-GMM is numerically invariant to. 2S-GMM and IT-GMM refer to 2-step and iterated procedures. The J tests are complemented with significance tests of a zero SDF mean. In particular, the p-value of the distance metric test of the null hypothesis of zero parameter is reported in parenthesis to the right of the estimate. 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 F1: Rejection rates (%) for a one-dimensional set of admissible SDFs ($T = 660$)

	CU			2S			IT		
				Nominal size					
	1	5	10	1	5	10	1	5	10
Panel A. Correct specification									
J d=1	0.88	5.21	10.13	0.38	2.96	6.82	0.52	2.72	6.58
J d=1, c=0	99.58	99.98	99.98	100	100	100	99.98	100	100
DM c=0	99.92	99.98	99.98	100	100	100	100	100	100
J d=2	98.73	98.85	99.96	6.44	19.02	29.56	7.78	21.58	32.20
J d=2, c=0	100	100	100	100	100	100	100	100	100
DM c=0	100	100	100	100	100	100	100	100	100
J d=3	100	100	100	89.58	95.36	97.30	85.18	93.52	95.86
J d=3, c=0	100	100	100	100	100	100	100	100	100
DM c=0	100	100	100	100	100	100	100	100	100
Panel B. Complete overspecification									
J d=1	0.26	2.56	6.40	0.34	1.60	3.68	0.32	1.74	3.68
J d=1, c=0	1.09	5.60	10.66	11.20	26.28	36.60	11.38	26.02	36.72
DM c=0	9.24	21.48	30.35	71.92	88.66	93.52	71.80	88.48	93.58
J d=2	43.55	68.58	78.94	5.88	17.60	27.34	5.78	17.26	27.00
J d=2, c=0	99.41	98.98	100	99.96	100	100	99.96	100	100
DM c=0	97.13	98.35	98.93	100	100	100	100	100	100
J d=3	99.97	100	100	89.64	95.50	97.32	85.82	93.72	96.16
J d=3, c=0	100	100	100	100	100	100	100	100	100
DM c=0	100	100	100	100	100	100	100	100	100

Note: This table displays the rejection rates of J tests, their variants restricted to zero SDF means, and the corresponding DM tests, as described in Section 3. The tests are computed for CU, two-step and iterated GMM. The rates are shown in percentage for the asymptotic critical values at 1, 5, and 10%. 5,000 samples of 25 excess returns and 3 factors 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 F2: Rejection rates (%) for a two-dimensional set of admissible SDFs ($T = 660$)

	CU			2S			IT		
				Nominal size					
	1	5	10	1	5	10	1	5	10
Panel A. Partial overspecification									
J d=1	0.00	0.19	0.78	0.00	0.02	0.12	0.00	0.02	0.14
J d=1, c=0	1.21	5.81	11.73	12.56	26.90	37.62	12.54	26.52	37.48
DM c=0	36.69	53.43	62.08	93.70	98.66	99.52	93.58	98.68	99.48
J d=2	2.74	10.27	17.79	0.02	0.34	1.04	0.02	0.36	1.16
J d=2, c=0	99.27	99.92	99.98	99.88	100	100	99.90	100	100
DM c=0	99.98	99.98	100	100	100	100	100	100	100
J d=3	99.97	99.97	100	73.42	87.14	91.18	63.42	80.06	86.64
J d=3, c=0	100	100	100	100	100	100	100	100	100
DM c=0	100	100	100	100	100	100	100	100	100
Panel B. Complete overspecification									
J d=1	0.00	0.04	0.40	0.04	0.66	2.14	0.08	0.66	2.34
J d=1, c=0	0.00	0.12	0.84	1.74	7.16	13.40	1.64	7.28	13.26
DM c=0	4.06	13.63	22.05	35.72	60.62	72.54	35.32	60.22	72.36
J d=2	1.14	5.31	11.90	0.00	0.10	00.40	0.00	0.10	0.46
J d=2, c=0	3.29	11.01	19.80	7.90	21.60	31.98	8.14	21.54	32.20
DM c=0	15.37	30.58	41.26	91.34	97.80	98.98	91.42	97.74	98.98
J d=3	61.51	81.14	88.30	74.64	87.74	92.30	66.14	82.14	88.54
J d=3, c=0	100	100	100	100	100	100	100	100	100
DM c=0	100	100	100	100	100	100	100	100	100

Note: This table displays the rejection rates of J tests, their variants restricted to zero SDF means, and the corresponding DM tests, as described in Section 3. The tests are computed for CU, two-step and iterated GMM. The rates are shown in percentage for the asymptotic critical values at 1, 5, and 10%. 5,000 samples of 25 excess returns and 3 factors 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.

Figure B1: One asset

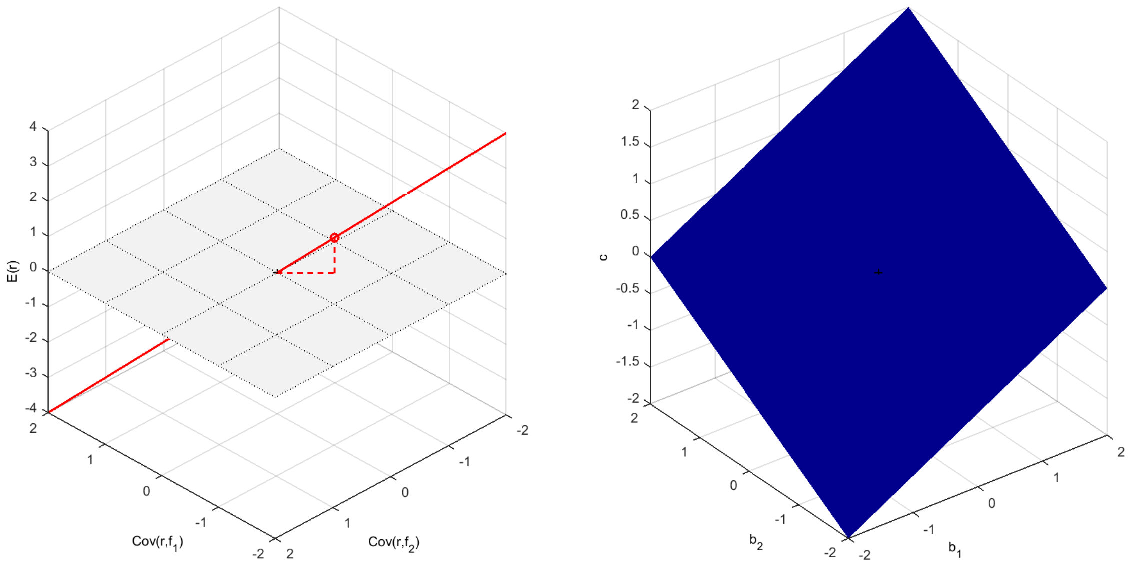

Figure B2: Two assets

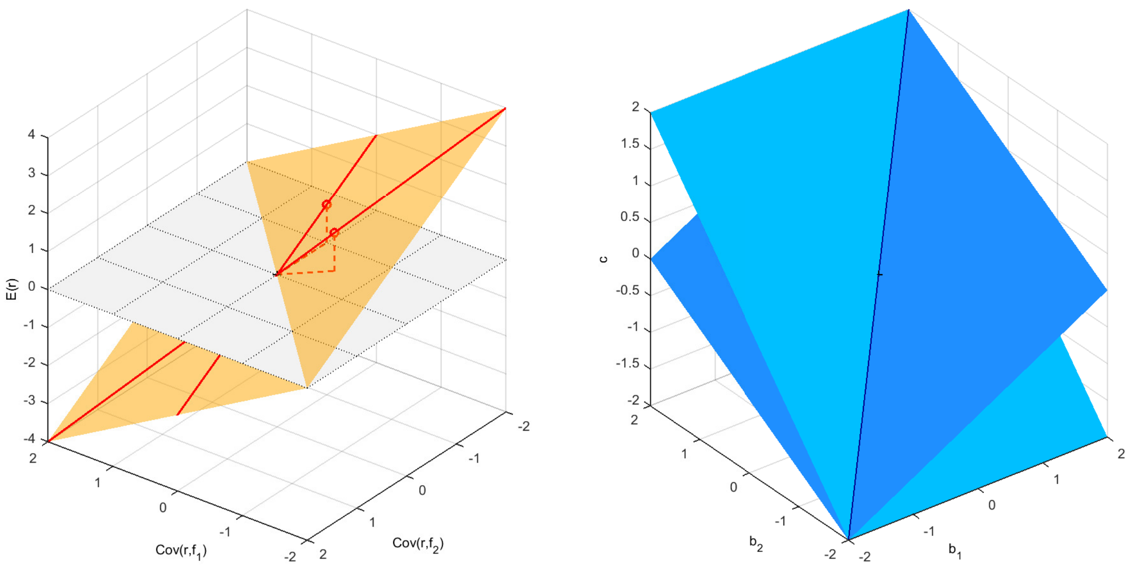

Figure B3: Three segmented asset markets

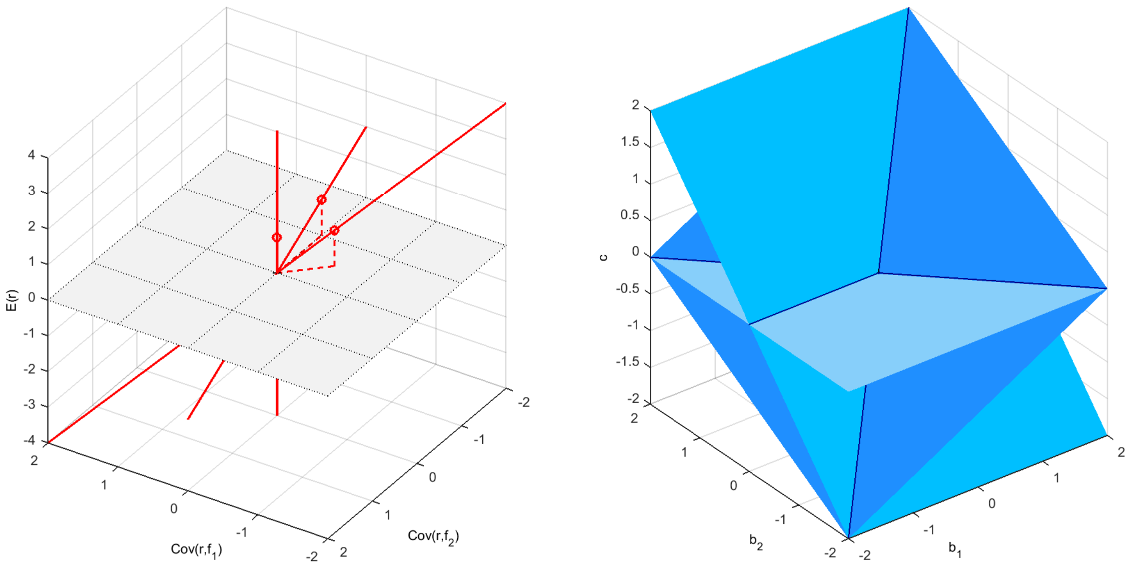

Figure B4: Three integrated asset markets

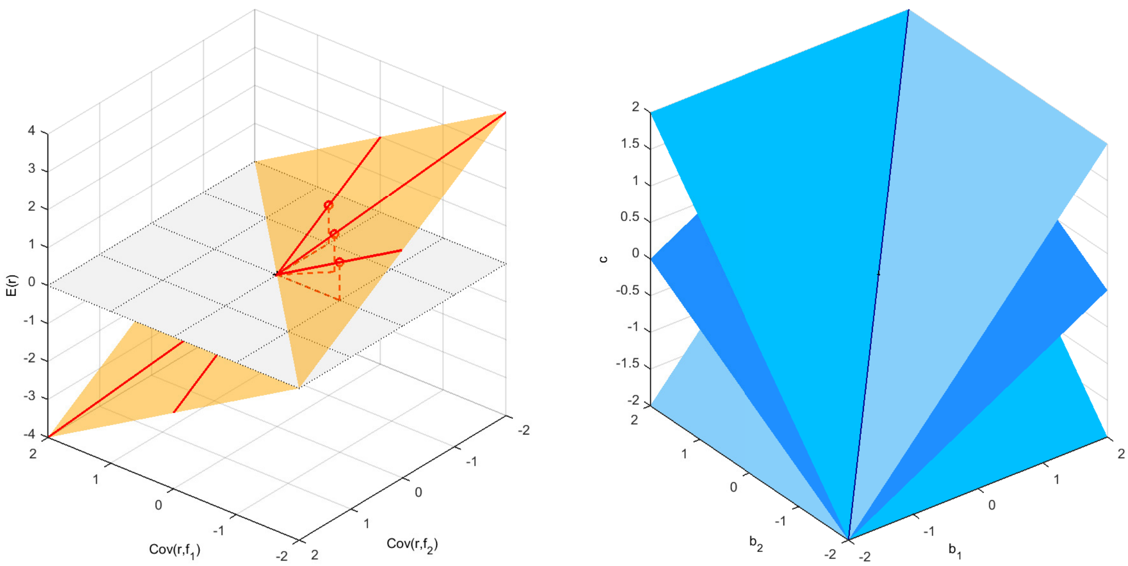

Figure B5: An unpriced second factor

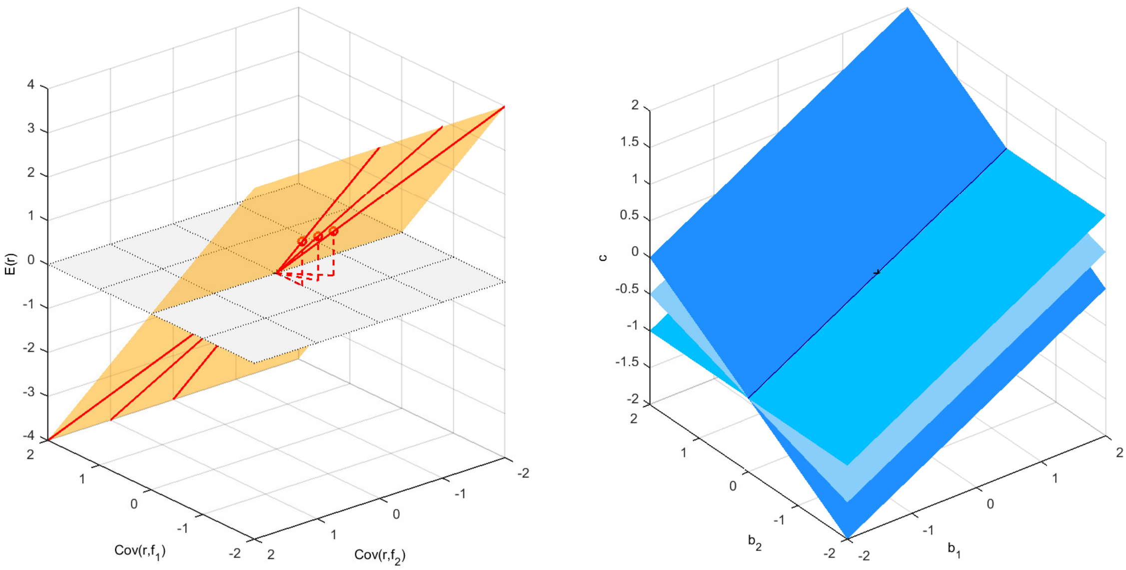

Figure B6: Two single factor models

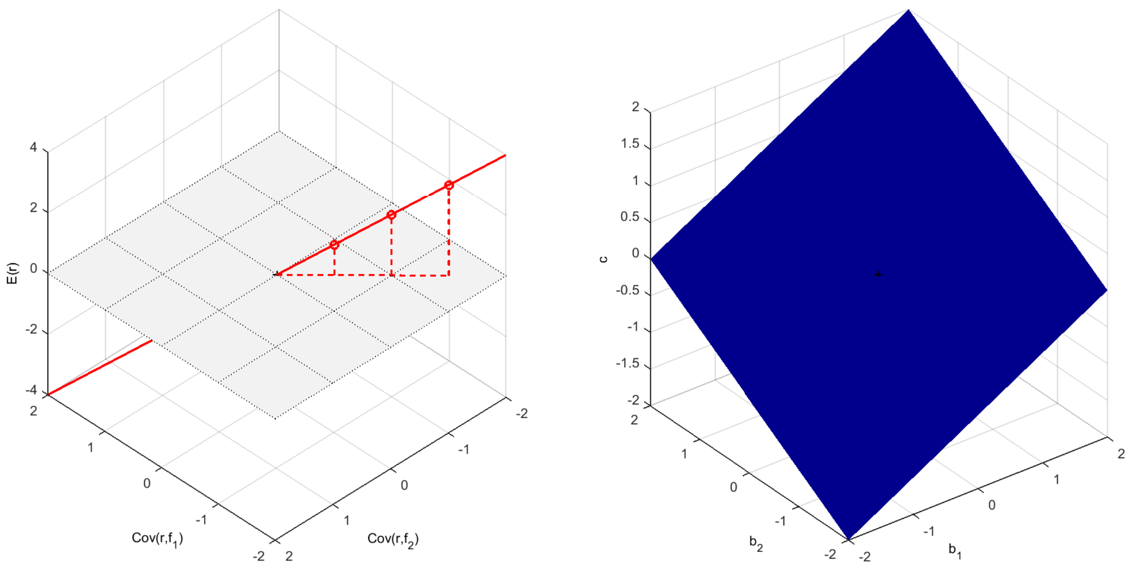

Figure B7: Admissible and attractive model with a useless factor

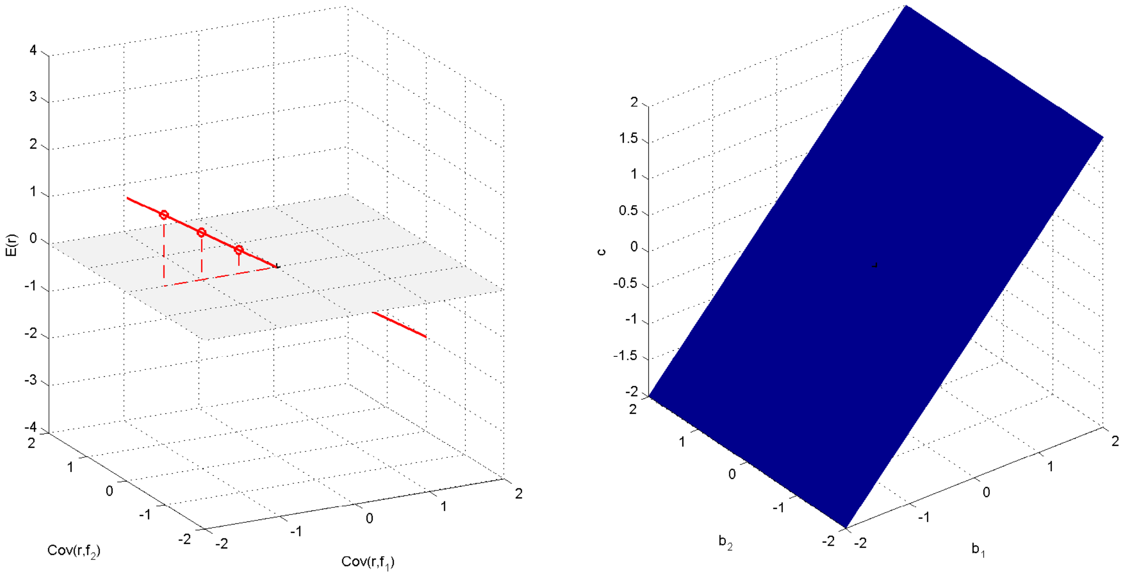

Figure B8: Admissible but unattractive model with a useless factor

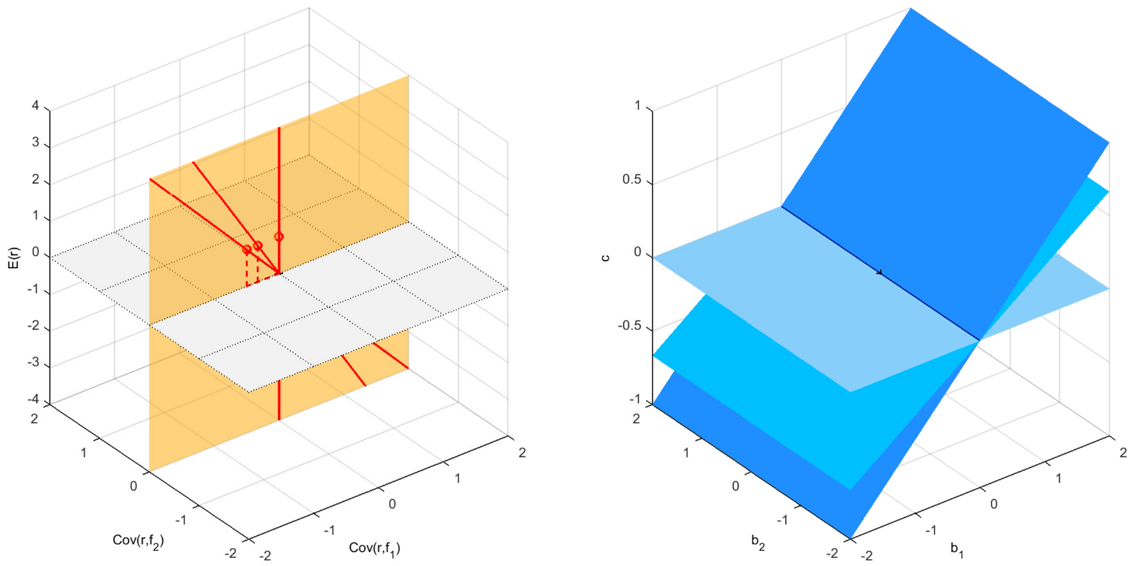

Figure B9: Two useless factors

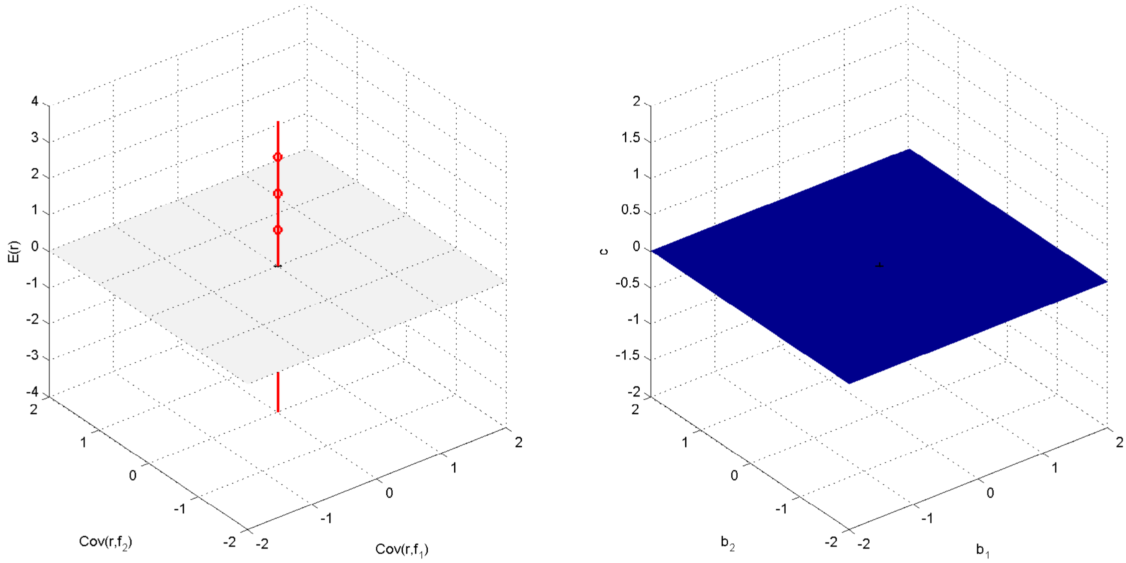

Figure C1: Normalizations

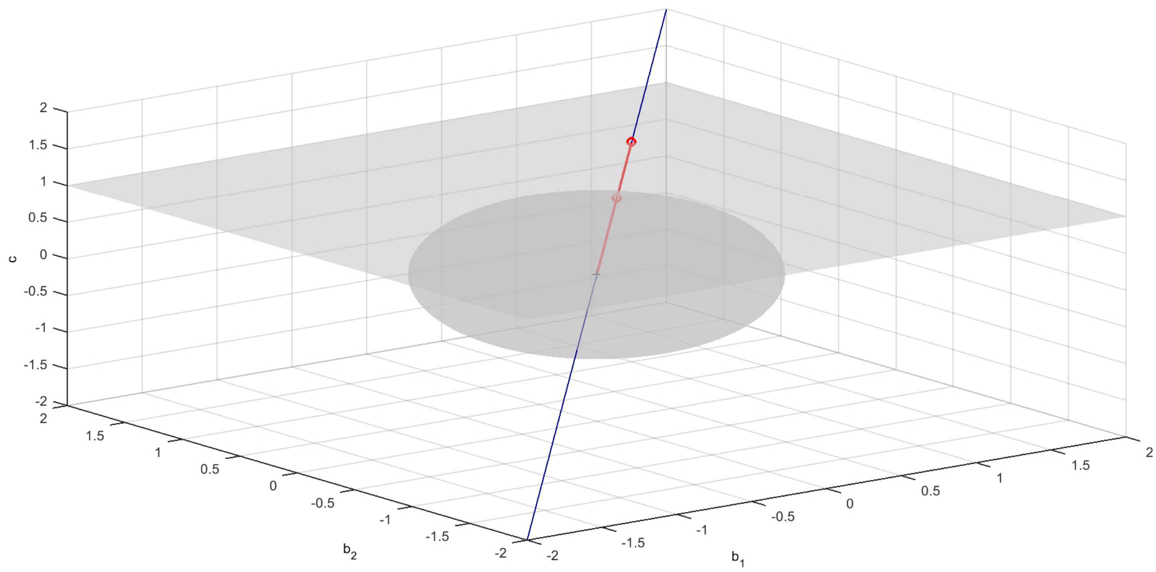

Figure E1: Risk premia from 2S-GMM

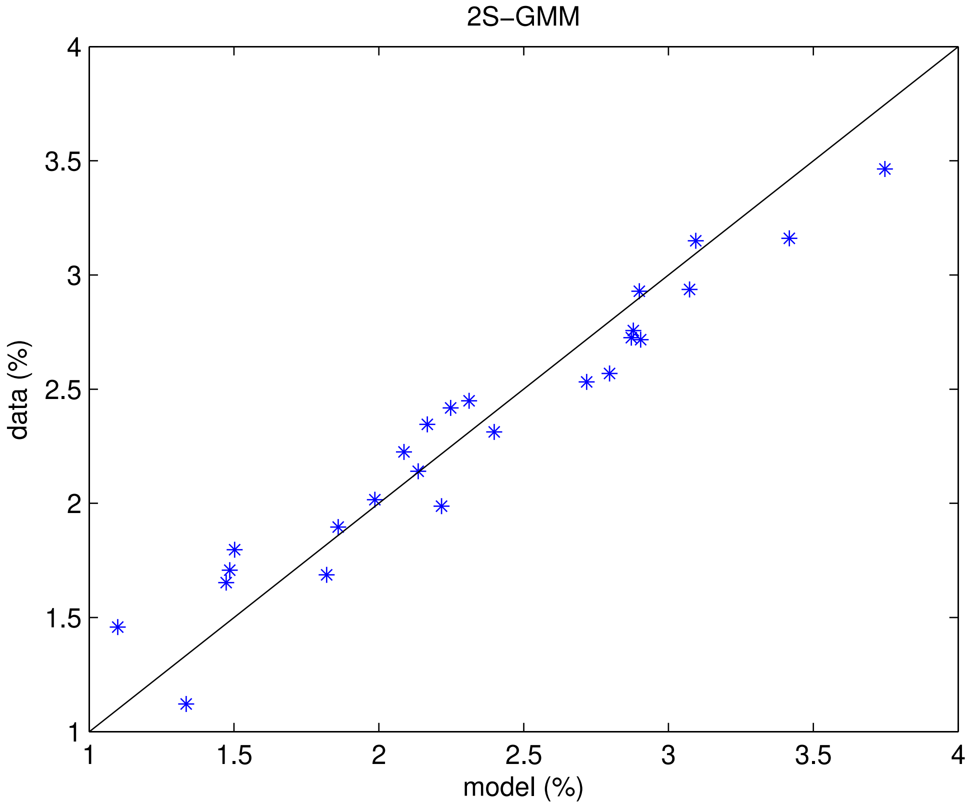

Figure E2: Risk premia from IT-GMM

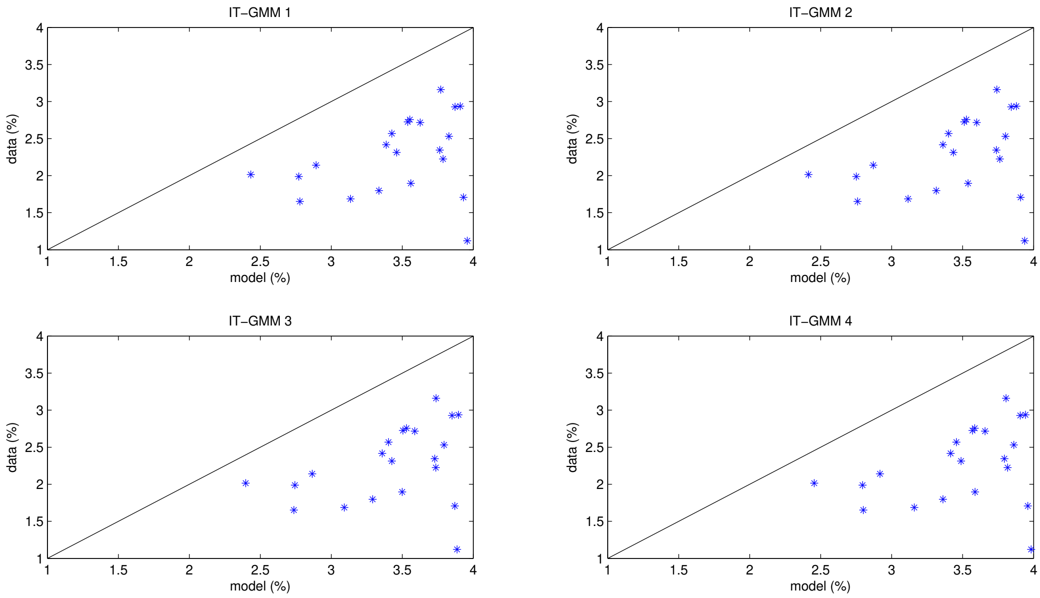

Figure E3: Risk premia from IT-GMM, free coefficients