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

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

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. □

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

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

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

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

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

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

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

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

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

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

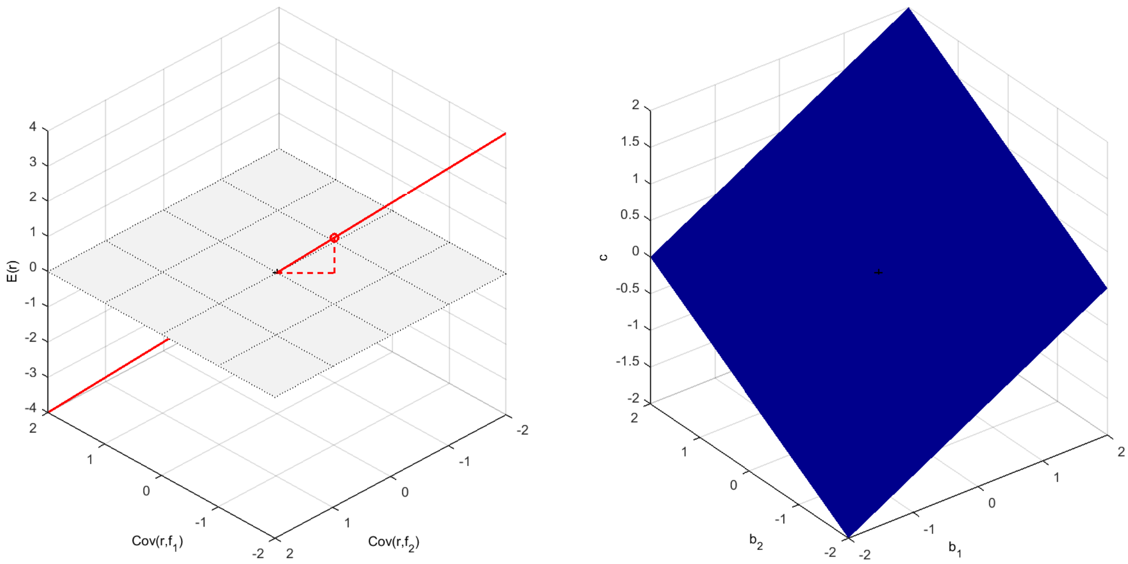

Figure B2: Two assets

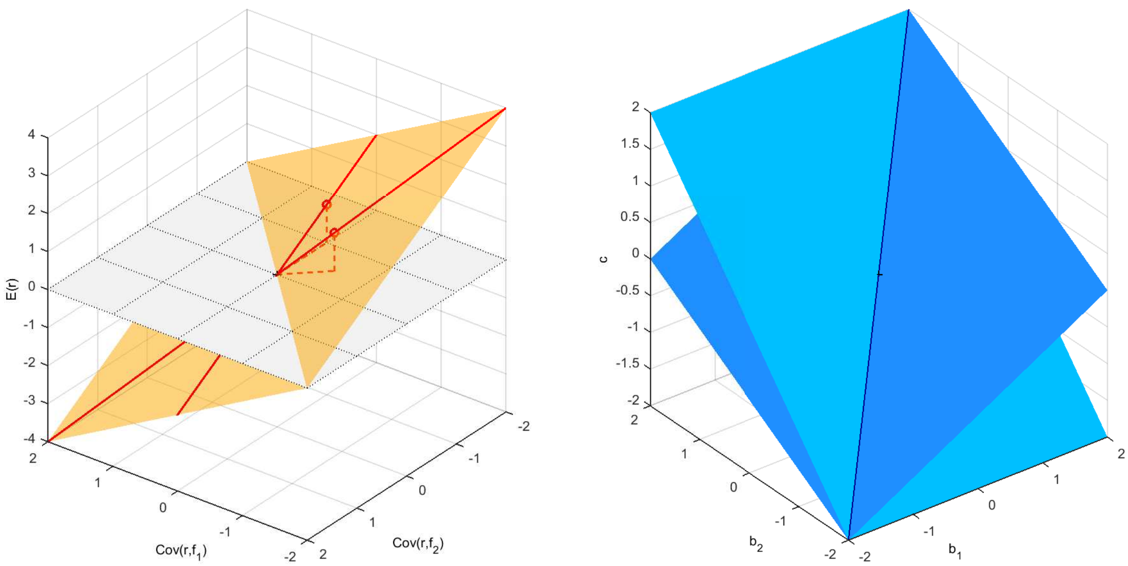

Figure B3: Three segmented asset markets

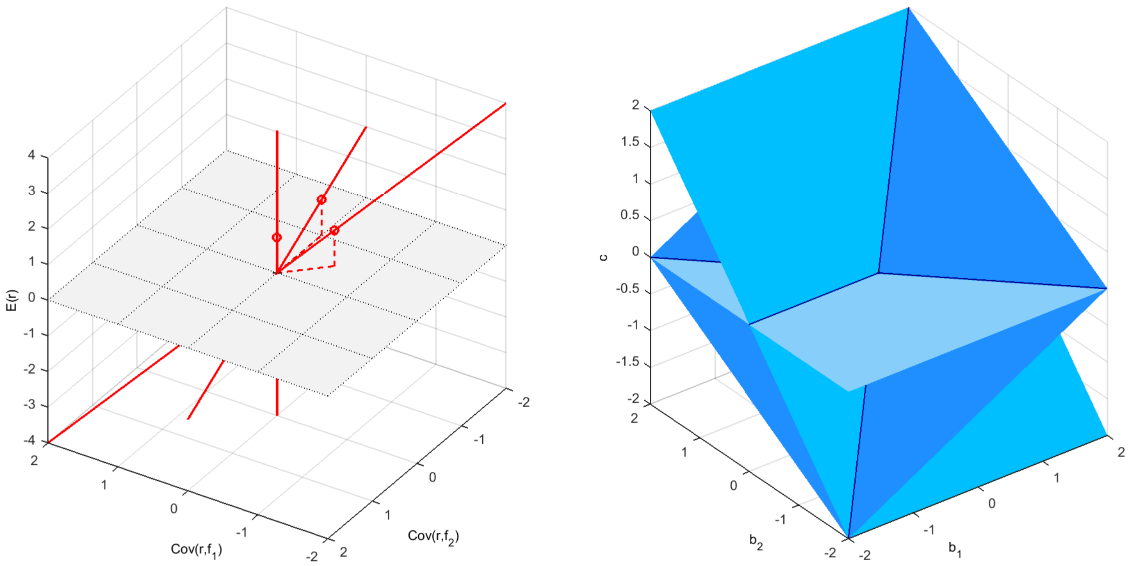

Figure B4: Three integrated asset markets

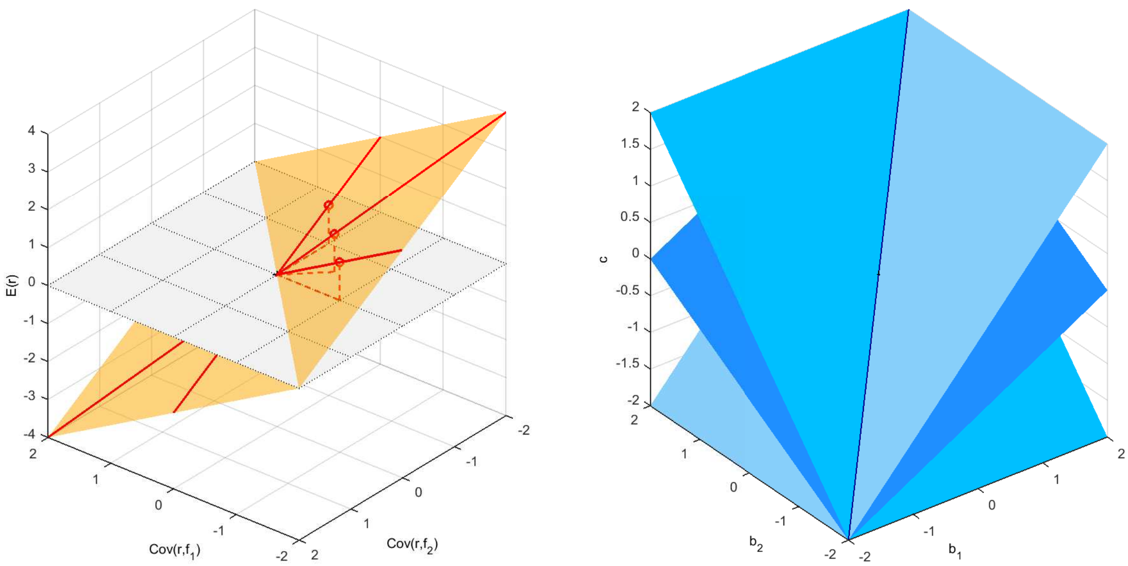

Figure B5: An unpriced second factor

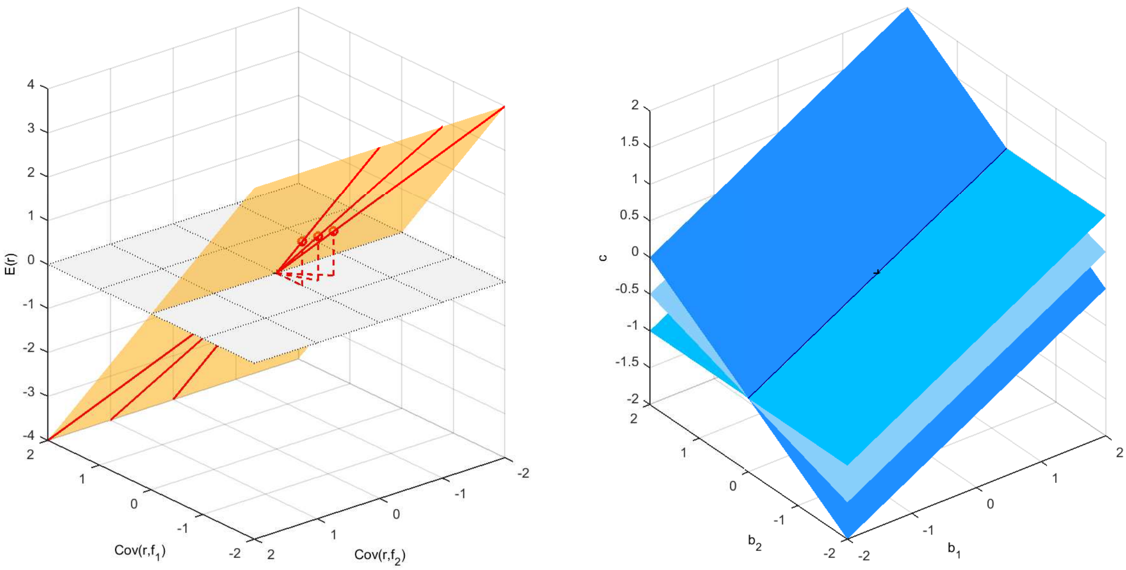

Figure B6: Two single factor models

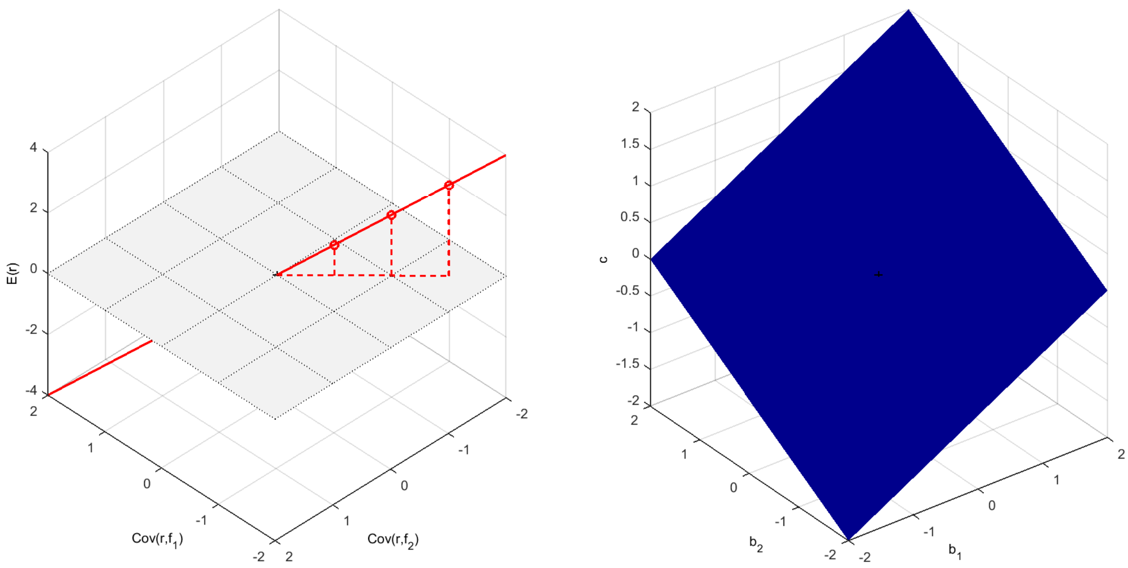

Figure B7: Valid and attractive model with a useless factor

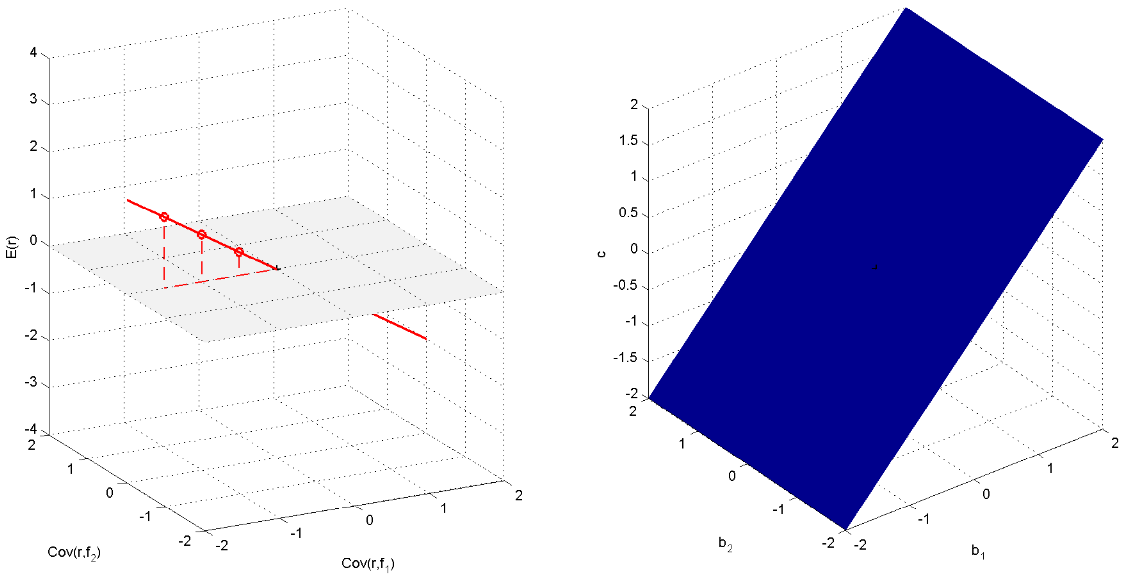

Figure B8: Valid but unattractive model with a useless factor

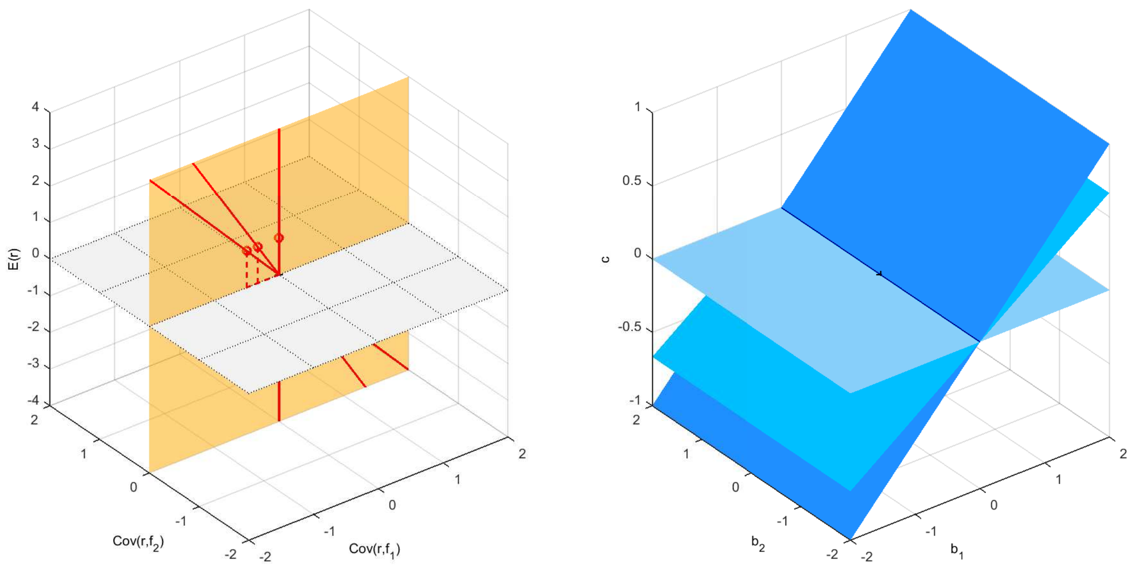

Figure B9: Two useless factors

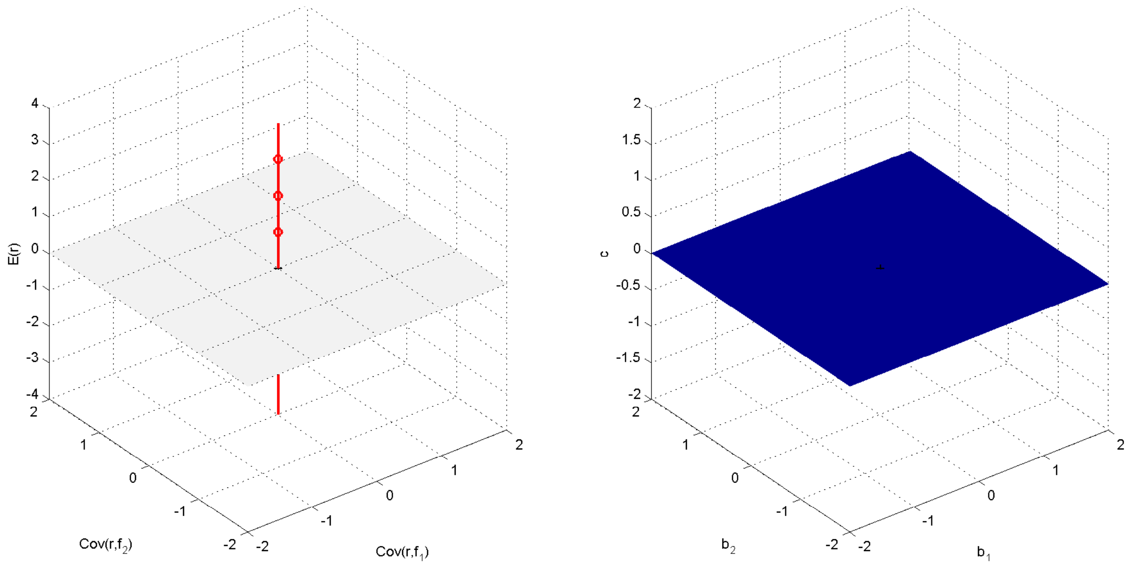

Figure C1: Normalizations