

Spanning Tests in Return and Stochastic Discount Factor Mean-Variance Frontiers: A Unifying Approach*

Francisco Peñaranda

UPF, Ramon Trias Fargas 25-27, E-08005 Barcelona, Spain.

<francisco.penaranda@upf.edu>

Enrique Sentana

CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain.

<sentana@cemfi.es>

Revised: September 2010

Abstract

We propose new spanning tests that assess if the initial and additional assets share the economically meaningful cost and mean representing portfolios. We prove their asymptotic equivalence to existing tests under local alternatives. We also show that unlike two-step or iterated procedures, single-step methods such as continuously updated GMM yield numerically identical overidentifying restrictions tests, so there is arguably a single spanning test. To prove these results, we extend optimal GMM inference to deal with singularities in the long run second moment matrix of the influence functions. Finally, we test for spanning using size and book-to-market sorted US stock portfolios.

Keywords: Asset Pricing, Continuously Updated GMM, Generalised Empirical Likelihood, Generalised Inverse, Representing Portfolios, Singular Covariance Matrix.

JEL: G11, G12, C12, C13.

*We would like to thank Manuel Arellano, Lars Hansen, Gur Huberman, Jan Magnus, Esteban Rossi-Hansberg, Guofu Zhou, participants at the European Meeting of the Econometric Society (Lausanne, 2001), the XI Finance Forum (Alicante, 2003), the III Workshop on International Finance (Malaga, 2003), as well as audiences at ANU, CEMFI, Chicago, CIDE, Columbia, INSEAD, LSE, Montreal, Nuffield College, Pompeu Fabra, Sydney, Tor Vergata and Venice for very useful comments and suggestions. The comments of an associate editor and an anonymous referee have also greatly improved the paper. Of course, the usual caveat applies. Financial support from the Spanish Ministry of Science and Innovation through grants ECO 2008-03066 (Peñaranda) and ECO 2008-00280 (Sentana) is gratefully acknowledged. Peñaranda also acknowledges the financial support of the Barcelona GSE and the Government of Catalonia.

1 Introduction

The return mean-variance frontier (RMVF) originally proposed by Markowitz (1952) is widely regarded as the cornerstone of modern investment theory. Similarly, the stochastic discount factor (SDF) mean-variance frontier (SMVF) introduced by Hansen and Jagannathan (1991) represents a major breakthrough in the way financial economists look at data on asset returns to discern which asset pricing theories are not empirically falsified. Somewhat remarkably, it turns out that both frontiers are intimately related, as they effectively summarise the sample information about the first and second moments of asset payoffs.

In this context, tests for mean-variance spanning in the RMVF and SMVF try to answer a very simple question: does the relevant frontier remain unchanged after increasing the number of assets that we analyse? And although the answer has to be the same for both frontiers, the implications of spanning are different. When we consider the RMVF, we want to assess if the exclusion of some assets reduces the risk-return trade-offs faced by investors, while when we study the SMVF, we want to determine if the additional assets impose tighter restrictions on asset pricing models irrespective of whether investors have mean-variance preferences. It is perhaps not surprising that there is a strand of the literature that develops tests for spanning in the RMVF (see Huberman and Kandel (1987) and Ferson, Foerster and Keim (1993)), and another one that develops tests for spanning in the SMVF (see De Santis (1993, 1995) and Bekaert and Urias (1996)).

Despite their different motivation, both approaches are systematically used in numerous empirical studies of (i) mutual fund performance evaluation (see De Roon and Nijman (2001) for a survey); (ii) gains from portfolio diversification arising from separate asset classes (Hunter and Simon (2005)) or cross-border investments (Errunza, Hogan and Hung (1999)), but also accruing from non-financial assets such as real estate (Stevenson (2001)), or human capital (Palacios-Huerta (2003)); and (iii) risk premia restrictions imposed by linear factor pricing models (see e.g. the Campbell, Lo and MacKinlay (1996) or Cochrane (2001) textbooks).

The purpose of this paper is to provide a unified treatment of spanning tests. We do so at three different levels:

a) We exploit the duality of the two frontiers to derive novel spanning tests that are not tied down to the specific properties of either frontier. In particular, since both frontiers are spanned by the uncentred or centred versions of the cost and mean representing portfolios introduced

by Chamberlain and Rothschild (1983), we propose to test if those portfolios are shared by the initial and extended sets of assets. An important advantage of our approach is that it has a direct economic interpretation because mean representing portfolios are the ones that mimic the safe asset with the minimum tracking error, while cost representing portfolios do the same thing for stochastic discount factors.

Although our tests fit rather naturally in a GMM framework, the introduction of additional moment conditions that define mean returns renders the joint covariance matrix of the augmented set of moment conditions singular in the population, but not necessarily in the sample, which complicates inference. In fact, the same type of singularity also affects other spanning tests in the RMVF and SMVF, a fact that had been so far overlooked. For that reason, we extend the theory of optimal GMM estimation in Hansen (1982) to those non-trivial situations in which the long run second moment matrix of the estimating functions is singular along a manifold in the parameter space that contains the true value. This case covers not only the spanning tests that we consider but also other situations with dynamic stochastic singularities.

b) We compare our proposed tests to the extant spanning tests, and show that the parametric restrictions are equivalent, which was known of the existing procedures. More importantly, we also show that all the tests are asymptotically equivalent under the null and compatible sequences of local alternatives, despite the fact that the number of parameters and moment conditions can be different, although the number of degrees of freedom is the same. We would like to emphasise that we obtain our novel asymptotic equivalence results under fairly weak assumptions on the distribution of asset returns. In particular, we do not require that returns are independent or identically distributed (*i.i.d.*) as Gaussian random vectors. And although we focus our discussion on overidentifying restrictions tests of spanning, which coincide with Distance Metric tests in this context, our equivalence results also apply to Lagrange Multiplier and Wald tests, which share their optimality properties under the null and sequences of local alternatives.

c) We show that by using single-step GMM procedures such as the Continuously Updated (CU) version in Hansen, Heaton and Yaron (1996), we can make all the different overidentification tests numerically identical, which means that one could argue that effectively there is only one GMM-based spanning test. Intuitively, generalised empirical likelihood procedures like CU-GMM allow us to make our inferences invariant to both reparametrisations and transformations of the moment conditions, thereby reproducing the uniqueness of the likelihood ratio (LR) test

in a classical context. Formally, though, the optimal GMM procedure under singularity that we develop is crucial in reaching this conclusion. In contrast, Wald tests will not be numerically equivalent, which confirms the well-known fact that their finite sample distribution can differ substantially from its asymptotic distribution in non-linear contexts.

For illustrative purposes, we apply our testing procedures to study if the US stock market portfolio and the familiar two Fama-French portfolios that capture size and value effects span the same return and SDF mean-variance frontiers as six portfolios sorted by size and book-to-market. Since we reject the null hypothesis of spanning, our results suggest that a mean-variance investor who is fully invested in US stocks would be worse off if her choice were constrained to strategies that only combine the size and value mimicking portfolios with the market. Equivalently, our results imply that these three funds do not fully capture the information about SDF's in the six size and book-to-market sorted portfolios.

The rest of the paper is as follows. In section 2, we introduce the required mathematical structure, while in section 3 we present our solution for optimal GMM inference with dynamic stochastic singularities. This section is written so that readers who are not interested in spanning tests can apply it to other problems, while those who are not interested in GMM inference can go directly to the new spanning tests proposed in section 4. Then, we carry out our comparison of all the tests in section 5. Finally, we present our empirical application in section 6 and summarise our conclusions in section 7. Proofs and auxiliary results are relegated to appendices.

2 Theoretical background

In this section, we first describe the representing portfolios introduced by Chamberlain and Rothschild (1983), which we then use to characterise the RMVF and SMVF. For the sake of brevity, we do not discuss any special cases, which we study in Peñaranda and Sentana (2004).¹

2.1 Cost and Mean Representing Portfolios

Consider an economy with a finite number N of risky assets whose random payoffs are defined on an underlying probability space. Let $\mathbf{R} = (R_1, \dots, R_N)'$ denote the vector of gross returns on those assets, with first and second uncentred moments given by $\boldsymbol{\nu}$ and $\boldsymbol{\Gamma}$, respectively. We

¹Specifically, Peñaranda and Sentana (2004) develop spanning tests in the presence of a safe asset, in the case of arbitrage (i.e. zero-cost) portfolios, and when all expected returns are equal.

assume that these moments are bounded, which implies that $R_i \in L^2$ ($i = 1, \dots, N$), where L^2 is the collection of all random variables defined on the underlying probability space with bounded second moments. We can then obtain the covariance matrix of the N asset returns, $\mathbf{\Sigma}$ say, as $\mathbf{\Gamma} - \boldsymbol{\nu}\boldsymbol{\nu}'$, which we assume has full rank. This implies that none of the original assets is either riskless or redundant, and consequently, that it is not possible to generate a riskless portfolio from \mathbf{R} , other than the trivial one. We also assume that not all expected returns are equal.

Let P be the set of the payoffs from all possible portfolios of the N original assets, which is given by the linear span of \mathbf{R} , $\langle \mathbf{R} \rangle$. Therefore, the elements of P will be of the form $p = \sum_{i=1}^N w_i R_i = \mathbf{w}'\mathbf{R}$, where $\mathbf{w} = (w_1, \dots, w_N)' \in \mathbb{R}^N$ is a vector of portfolio weights. There are at least three characteristics of portfolios in which investors are interested: their cost, the expected value of their payoffs, and their variance, which will be given by $C(p) = \mathbf{w}'\boldsymbol{\ell}_N$, $E(p) = \mathbf{w}'\boldsymbol{\nu}$ and $V(p) = \mathbf{w}'\mathbf{\Sigma}\mathbf{w}$ respectively, where $\boldsymbol{\ell}_N$ is a vector of N ones, which reflects the fact that we have normalised the price of all the original assets to 1. Since P is a closed linear subspace of L^2 , it is also a Hilbert space under the mean square inner product, $E(xy)$, and the associated mean square norm $\sqrt{E(x^2)}$, where $x, y \in L^2$. Such a topology allows us to define the least squares projection of any $q \in L^2$ onto P as:

$$E(q\mathbf{R})E^{-1}(\mathbf{R}\mathbf{R}')\mathbf{R}, \quad (1)$$

which is the element of P that is closest to q in the mean square norm.

In this context, we can formally understand $C(\cdot)$ and $E(\cdot)$ as linear functionals that map the elements of P onto the real line. The expected value functional is always continuous on L^2 while our full rank assumption on $\mathbf{\Sigma}$ implies that $\mathbf{\Gamma}$ has full rank too, and consequently that the cost functional is also continuous on P . The Riesz representation theorem then implies that there exist two unique elements of P that represent these functionals over P (see Chamberlain and Rothschild (1983)). In particular, the uncentred cost and mean representing portfolios, p^* and p^+ , respectively, will be such that:

$$C(p) = E(p^*p) \quad \text{and} \quad E(p) = E(p^+p) \quad \forall p \in P.$$

It is then straightforward to show that

$$\begin{aligned} p^* &= \boldsymbol{\phi}^{*'}\mathbf{R} = \boldsymbol{\ell}'_N\mathbf{\Gamma}^{-1}\mathbf{R}, \\ p^+ &= \boldsymbol{\phi}^{+'}\mathbf{R} = \boldsymbol{\nu}'\mathbf{\Gamma}^{-1}\mathbf{R}, \end{aligned} \quad (2)$$

where $(\boldsymbol{\phi}^*, \boldsymbol{\phi}^+)$ will be parameters of interest in spanning tests.

If P included a unit payoff, then p^+ would coincide with it. But even though it does not, it follows from (1) that p^+ is the projection of 1 onto P , which in financial markets parlance simply means that the mean representing portfolio is the portfolio that “mimics” the safe asset with the minimum “tracking error”. To give a similar economic interpretation to p^* , it is convenient to recall that a stochastic discount factor, m say, is any scalar random variable defined on the same underlying probability space which prices assets in terms of their expected cross product with it. We can again use (1) to interpret p^* as the projection of any m onto P , i.e. as the portfolio that best mimics stochastic discount factors. In addition, since $C(1) = E(1 \cdot m) = c$ say, the expected value of m defines the shadow price of a unit payoff.

Since $C(p^*) = E(p^{*2}) > 0$, we can always define an associated return R^* as $p^*/C(p^*)$. Similarly, we can usually define R^+ as $p^+/C(p^+)$, except when p^* and p^+ are orthogonal, which in view of our assumptions happens if and only if $\text{cov}(p^{**}, p^{++}) = \boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}\ell_N = 0$.

Finally, Chamberlain and Rothschild (1983) show that an alternative valid topology on P can be defined with covariance as inner product and standard deviation as norm when there is not a safe asset in P . Hence, we could also represent the two functionals by means of two alternative centred representing portfolios, p^{**} and p^{++} in P , such that

$$C(p) = \text{Cov}(p^{**}, p) \quad \text{and} \quad E(p) = \text{Cov}(p^{++}, p) \quad \forall p \in P.$$

Not surprisingly,

$$\begin{aligned} p^{**} &= \boldsymbol{\varphi}^{*\prime}\mathbf{R} = \ell'_N\boldsymbol{\Sigma}^{-1}\mathbf{R} = p^* + (\boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}\ell_N)p^+, \\ p^{++} &= \boldsymbol{\varphi}^{+\prime}\mathbf{R} = \boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}\mathbf{R} = (1 + \boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu})p^+, \end{aligned} \tag{3}$$

where again $(\boldsymbol{\varphi}^*, \boldsymbol{\varphi}^+)$ will be parameters of interest in spanning tests. We can then define the return associated with p^{**} as $R^{**} = p^{**}/C(p^{**}) = (\ell'_N\boldsymbol{\Sigma}^{-1}\mathbf{R}) / (\ell'_N\boldsymbol{\Sigma}^{-1}\ell_N)$, which coincides with the minimum variance return. Similarly, we can also define R^{++} as $p^{++}/C(p^{++}) = p^+/C(p^+) = R^+$ if (and only if) $\boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}\ell_N \neq 0$.²

2.2 SDF and Return Mean-Variance Frontiers

The SMVF, or Hansen and Jagannathan (1991) frontier, is the set of admissible SDF's with the lowest variance for a given mean. Therefore, its elements solve the programme

$$\min_{m \in L^2} V(m) \quad \text{s.t.} \quad E(m) = c \in \mathbb{R}, \quad E(m\mathbf{R}) = \ell_N.$$

²When $\boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}\ell_N = 0$, both p^+ and p^{++} are arbitrage portfolios, which means that neither R^+ nor R^{++} can be defined. In addition, $p^{**} = p^*$, so that $R^{**} = R^*$.

If there were a safe asset then its gross return would pin down a unique c^{-1} . But even though no safe asset exists, we can trace the SMVF by solving the above programme for any safe return $c^{-1} \geq 0$. As shown by Gallant, Hansen and Tauchen (1990), its solution $m^{MV}(c)$ can be expressed as

$$m^{MV}(c) = p^* + \alpha(c)(1 - p^+), \quad \alpha(c) = \frac{c - E(p^*)}{E(1 - p^+)},$$

which shows that all the elements of the SMVF are portfolios spanned by p^* and $1 - p^+$ alone. Note, however, that $m^{MV}(c) \notin P$ except for p^* . Graphically, p^* is the element on the SMVF that is closest to the origin because it has the lowest second moment (see Hansen and Jagannathan (1991)). On the other hand, p^+ is not located on this frontier but it is related to its asymptotes since the term $1 - p^+$ becomes increasingly important as $c \rightarrow \pm\infty$.

The RMVF, or Markowitz (1952) frontier, is the set of feasible unit-cost portfolios that have the lowest variance for a given mean. Therefore, its elements solve the programme

$$\min_{p \in P} V(p) \quad s.t. \quad E(p) = \nu \in \mathbb{R}, \quad C(p) = 1.$$

As shown by Hansen and Richard (1987), the RMVF portfolios will be:

$$R^{MV}(\nu) = R^* + \omega(\nu)A^+, \quad \omega(\nu) = \frac{\nu - E(R^*)}{E(A^+)},$$

where $A^+ = p^+ - C(p^+)R^*$, as long as not all ν_i are equal, which we are assuming throughout. Thus, the RMVF will also be spanned by p^* and p^+ . Graphically, R^* is the element of the RMVF that is closest to the origin because it has the minimum second moment, while R^+ is the point of tangency of the frontier with a ray from the origin.

It is easy to show that if we subtract from $m^{MV}(c)$ its position on the unit payoff, and compute the corresponding return, then we will generally find an element on the RMVF (see Peñaranda and Sentana (2008) for details).

3 Optimal GMM in a singular set-up

We begin by briefly reviewing the inference methods proposed by Hansen (1982). This allows us to introduce notation and assumptions for the extension required to deal with the spanning tests in section 4, in which the second moment matrix of the estimating functions is singular along an implicit manifold in the parameter space. Those readers who are not interested in GMM inference can go directly to section 4.

Let $\{\mathbf{x}_t\}_{t=1}^T$ denote a strictly stationary and ergodic stochastic process, and define $\mathbf{h}(\mathbf{x}_t; \boldsymbol{\theta})$ as an $n \times 1$ vector of known functions of \mathbf{x}_t , where $\boldsymbol{\theta}$ is a $k \times 1$ vector of unknown parameters. The true parameter value, $\boldsymbol{\theta}^0$, which we assume belongs to the interior of the compact set $\Theta \subseteq \mathbb{R}^k$, is implicitly defined by the (population) moment conditions:

$$E[\mathbf{h}(\mathbf{x}_t; \boldsymbol{\theta}^0)] = \mathbf{0},$$

where the expectation is taken with respect to the stationary distribution of \mathbf{x}_t .

GMM estimators minimize a specific norm $\bar{\mathbf{h}}_T'(\boldsymbol{\theta}) \boldsymbol{\Upsilon}_T \bar{\mathbf{h}}_T(\boldsymbol{\theta})$ of the sample moments $\bar{\mathbf{h}}_T(\boldsymbol{\theta}) = T^{-1} \sum_{t=1}^T \mathbf{h}(\mathbf{x}_t; \boldsymbol{\theta})$ defined by some weighting matrix $\boldsymbol{\Upsilon}_T$. A necessary condition for the identification of $\boldsymbol{\theta}$ is the usual order condition $n \geq k$. If the inequality is strict, then we say that $\boldsymbol{\theta}$ is (seemingly) overidentified. On the other hand, we say that $\boldsymbol{\theta}$ is (seemingly) exactly identified when both dimensions coincide, in which case the weighting matrix $\boldsymbol{\Upsilon}_T$ will become irrelevant for large enough T if its probability limit $\boldsymbol{\Upsilon}$ is a positive definite matrix. In the overidentified case, in contrast, Hansen (1982) showed that if the long-run covariance matrix of the moment conditions $\mathbf{S}(\boldsymbol{\theta}^0) = \text{avar}[\sqrt{T} \bar{\mathbf{h}}_T(\boldsymbol{\theta}^0)]$ has full rank, then $\mathbf{S}^{-1}(\boldsymbol{\theta}^0)$ will be the ‘‘optimal’’ weighting matrix, in the sense that the difference between the asymptotic covariance matrix of the resulting GMM estimator and a GMM estimator based on any other norm of the same moment conditions is positive semidefinite. Therefore, the optimal GMM estimator of $\boldsymbol{\theta}$ will be

$$\hat{\boldsymbol{\theta}}_T = \arg \min_{\boldsymbol{\theta} \in \Theta} J_T(\boldsymbol{\theta}),$$

where

$$J_T(\boldsymbol{\theta}) = \bar{\mathbf{h}}_T'(\boldsymbol{\theta}) \mathbf{S}^{-1}(\boldsymbol{\theta}^0) \bar{\mathbf{h}}_T(\boldsymbol{\theta}).$$

This optimal estimator is infeasible unless we know $\mathbf{S}(\boldsymbol{\theta}^0)$, but under additional regularity conditions, we can define an asymptotically equivalent but feasible two-step optimal GMM estimator by replacing $\mathbf{S}(\boldsymbol{\theta}^0)$ with an estimator $\mathbb{S}_T(\boldsymbol{\theta})$ evaluated at some initial consistent estimator of $\boldsymbol{\theta}^0$, $\hat{\boldsymbol{\theta}}_T$ say. There is an extensive literature on heteroskedasticity and autocorrelation consistent (HAC) estimators of long-run covariance matrices, which typically adopt the form

$$\mathbb{S}_T(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T w_{ts} \mathbf{h}(\mathbf{x}_t; \boldsymbol{\theta}) \mathbf{h}'(\mathbf{x}_s; \boldsymbol{\theta}) \quad (4)$$

for some weights w_{ts} , which are the kernel function of the implicit nonparametric spectral density estimator (see for example DeJong and Davidson (2000) and the references therein). Although

(4) is prevalent in empirical work, being the default choice in the most popular econometric packages, it implicitly assumes that the moments are correctly specified. To avoid misleading inferences in case of misspecification, Hall (2000) advocates to use instead a HAC estimator that first centres the sample moments, as in Hansen, Heaton and Yaron (1996):

$$\mathfrak{S}_T(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T w_{ts} \mathbf{u}(\mathbf{x}_t; \boldsymbol{\theta}) \mathbf{u}'(\mathbf{x}_s; \boldsymbol{\theta}), \quad \mathbf{u}(\mathbf{x}_t; \boldsymbol{\theta}) = \mathbf{h}(\mathbf{x}_t; \boldsymbol{\theta}) - \bar{\mathbf{h}}_T(\boldsymbol{\theta}). \quad (5)$$

But even if the moments are correctly specified, it will also be important to distinguish between $\mathfrak{S}_T(\boldsymbol{\theta})$ and $\mathfrak{S}_T(\boldsymbol{\theta})$ in our singular context.

An alternative way to make the optimal GMM estimator feasible is by explicitly taking into account in the criterion function the dependence of the long-run variance on the parameter values, as in the single-step CU-GMM estimator of Hansen, Heaton and Yaron (1996), which is defined as

$$\tilde{\boldsymbol{\theta}}_T = \arg \min_{\boldsymbol{\theta} \in \Theta} \tilde{J}_T(\boldsymbol{\theta}),$$

where

$$\tilde{J}_T(\boldsymbol{\theta}) = \bar{\mathbf{h}}_T'(\boldsymbol{\theta}) \mathbf{S}_T^{-1}(\boldsymbol{\theta}) \bar{\mathbf{h}}_T(\boldsymbol{\theta}).$$

Although this estimator is often more difficult to compute than a two-step estimator, particularly in linear models, an important advantage is that it is numerically invariant to both bijective reparametrisations and parameter-dependent linear transformations of the moment conditions, which will prove useful in our context. Newey and Smith (2004) highlight other important advantages of CU- over two-step GMM by going beyond the usual first-order asymptotic equivalence results. They also discuss alternative single-step estimators, such as empirical likelihood or exponentially-tilted methods, which are also numerically invariant to transformations.³ In contrast, these properties do not necessarily hold for two-step or iterated GMM.

Assuming that $\mathbf{h}(\mathbf{x}_t; \boldsymbol{\theta})$ is continuously differentiable in $\boldsymbol{\theta}$, with a Jacobian matrix $\mathbf{D}_t(\boldsymbol{\theta}) = \partial \mathbf{h}(\mathbf{x}_t; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}'$ whose sample and population means, $\bar{\mathbf{D}}_T(\boldsymbol{\theta})$ and $\mathbf{D}(\boldsymbol{\theta})$ respectively, are also continuous in $\boldsymbol{\theta}$, the condition $\text{rank}[\mathbf{D}'(\boldsymbol{\theta}^0) \mathbf{S}^{-1}(\boldsymbol{\theta}^0) \mathbf{D}(\boldsymbol{\theta}^0)] = k$, which requires that $\text{rank}[\mathbf{D}(\boldsymbol{\theta}^0)] = k$, is sufficient for the local identifiability of $\boldsymbol{\theta}$ at $\boldsymbol{\theta}^0$. Under the additional assumptions that $E(\sup_{\boldsymbol{\theta} \in \Theta} \|\mathbf{h}(\mathbf{x}_t; \boldsymbol{\theta})\|) < \infty$, $\bar{\mathbf{D}}_T(\boldsymbol{\theta}^i) \xrightarrow{p} \mathbf{D}(\boldsymbol{\theta}^0)$ if $\boldsymbol{\theta}^i \xrightarrow{p} \boldsymbol{\theta}^0$, and $\sqrt{T} \bar{\mathbf{h}}_T(\boldsymbol{\theta}^0) \xrightarrow{d} N[\mathbf{0}, \mathbf{S}(\boldsymbol{\theta}^0)]$ we will have that

$$\sqrt{T}[\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^0] \xrightarrow{d} N\{\mathbf{0}, [\mathbf{D}'(\boldsymbol{\theta}^0) \mathbf{S}^{-1}(\boldsymbol{\theta}^0) \mathbf{D}(\boldsymbol{\theta}^0)]^{-1}\}$$

³In fact, Antoine, Bonnal and Renault (2007) study the Euclidean empirical likelihood estimator, which is numerically equivalent to CU-GMM for $\boldsymbol{\theta}$.

(see Newey and MacFadden (1994) for more primitive regularity conditions and proofs). Moreover, $T \cdot J_T(\hat{\boldsymbol{\theta}}_T)$ will be asymptotically distributed as a chi-square with $n - k$ degrees of freedom if $E[\mathbf{h}(\mathbf{x}_t; \boldsymbol{\theta}^0)] = \mathbf{0}$ holds, so that we can perform an overidentifying restrictions (J) test when $n > k$.

Unfortunately, Hansen's (1982) definitions of optimal GMM estimators and J test break down when $\mathbf{S}(\boldsymbol{\theta}^0)$ is singular. In the remainder of this section we shall discuss optimal GMM estimators of $\boldsymbol{\theta}$ and the corresponding overidentification test when the asymptotic second moment matrix of $\bar{\mathbf{h}}_T(\mathbf{x}_t; \boldsymbol{\theta}^0)$ is singular in the population but not necessarily in the sample, which is a prevalent feature of the spanning tests in section 4. In particular, the following two assumptions cover the singularities that arise in the context of those tests:

Assumption 1 Let $\boldsymbol{\Pi}(\boldsymbol{\theta})$ denote a $n \times s$ matrix of continuously differentiable functions of $\boldsymbol{\theta}$, where $0 \leq s \leq k$. The subset of Θ for which

$$\boldsymbol{\Pi}'(\boldsymbol{\theta}) \left[\sqrt{T} \bar{\mathbf{h}}_T(\boldsymbol{\theta}) \right] \xrightarrow{L^2} \mathbf{0} \quad (6)$$

can be fully characterised by $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$, where $\mathbf{m}(\boldsymbol{\theta})$ is a $s \times 1$ known continuously differentiable transformation of $\boldsymbol{\theta}$.

Assumption 2 If $s > 0$, then $\mathbf{m}(\boldsymbol{\theta}^0) = \mathbf{0}$, $\text{rank}[\mathbf{S}(\boldsymbol{\theta}^0)] = n - s$, and $\text{rank}[\partial \mathbf{m}(\boldsymbol{\theta}^0) / \partial \boldsymbol{\theta}'] = s$.

For the non-standard case of $s > 0$, the first assumption simply defines $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$ as the implicit $k - s$ -dimensional manifold in Θ over which s linear combinations of $\sqrt{T} \bar{\mathbf{h}}_T(\boldsymbol{\theta})$ converge in mean square to zero. In turn, the second assumption says that the true values of the parameters belong to that manifold, and it also ensures that the singularity of $\mathbf{S}(\boldsymbol{\theta}^0)$, when it exists, is fully characterised by Assumption 1. Finally, the rank condition in Assumption 2, together with the convergence in mean square to zero in Assumption 1, will guarantee the local identifiability of $\boldsymbol{\theta}$.

Assumption 1 covers dynamic models with singularities in the spectral density matrix of the influence functions, such as

$$[\boldsymbol{\Pi}_0(\boldsymbol{\theta}) + \boldsymbol{\Pi}_1(\boldsymbol{\theta})L]' \mathbf{h}(\mathbf{x}_t, \boldsymbol{\theta}) = \mathbf{0} \quad \forall \mathbf{x}_t,$$

where L denotes the lag operator, so that $\boldsymbol{\Pi}(\boldsymbol{\theta}) = \boldsymbol{\Pi}_0(\boldsymbol{\theta}) + \boldsymbol{\Pi}_1(\boldsymbol{\theta})$ (see Arellano, Hansen and Sentana (2009) and Diez de los Rios and Sentana (2010) for examples). In fact, in the case of spanning tests we will have that

$$\boldsymbol{\Pi}'(\boldsymbol{\theta}) \mathbf{h}(\mathbf{x}_t, \boldsymbol{\theta}) = \mathbf{0} \quad \forall \mathbf{x}_t, \quad (7)$$

which implies a singularity not only in the long run but also in the contemporaneous second moment matrix of $\mathbf{h}(\mathbf{x}_t, \boldsymbol{\theta})$.

In that context, Assumption 1 goes far beyond trivial situations with “duplicated” moment conditions, in which some linear combinations of $\mathbf{h}(\mathbf{x}_t, \boldsymbol{\theta})$ with coefficients that do not depend on $\boldsymbol{\theta}$ are 0 for all $\boldsymbol{\theta}$. Given that in those cases any HAC estimator $\mathbf{S}_T(\boldsymbol{\theta})$ will be singular in finite samples irrespective of (i) the use of CU or two-step estimators, (ii) the choice of first-step estimator $\dot{\boldsymbol{\theta}}_T$ in the second case, and (iii) the use of centred or uncentred moments in the estimation of $\mathbf{S}(\boldsymbol{\theta}^0)$, the appropriate action is simply to eliminate the “duplicated” moment conditions, which can be mechanically achieved by using as weighting matrix any generalised inverse of the HAC estimators (4) or (5).

Similarly, Assumption 1 is also more general than

$$\boldsymbol{\Pi}'(\boldsymbol{\theta})\mathbf{h}(\mathbf{x}_t, \boldsymbol{\theta}) = \mathbf{m}(\boldsymbol{\theta}) \quad \forall \mathbf{x}_t \text{ and } \forall \boldsymbol{\theta} \in \Theta, \quad (8)$$

in which case there would be s linear combinations of $\mathbf{h}(\mathbf{x}_t, \boldsymbol{\theta})$ with coefficients that depend on $\boldsymbol{\theta}$ which would be constant regardless of the parameter values. While (8) implies that $\mathbf{m}(\boldsymbol{\theta}^0) = \mathbf{0}$, Assumption 1 does not require that the singularity of the covariance matrix is uniform over the parameter space. This distinction is important for our purposes because not all spanning tests fit within this more restrictive framework.

Our proposed solution for conducting optimal GMM estimation and inference under the type of singularity characterised by Assumptions 1 and 2 involves the following two steps:

a) impose the parametric restrictions $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$, which means that we implicitly estimate a subset of $k - s$ free parameters, the rest being pinned down by the manifold, and *simultaneously*

b) replace the ordinary inverse of $\mathbf{S}(\boldsymbol{\theta}^0)$, which cannot be defined when $s > 0$, by any of its generalised inverses, $\mathbf{S}^-(\boldsymbol{\theta}^0)$.

In this way, we effectively decrease both the number of parameters and the number of moment conditions to avoid the singularity, but their difference (the degrees of freedom of the J test) remains the same.⁴ Our solution is similar in spirit to the approach discussed in Judge et al. (1985, section 12.5.2) to deal with a classical multivariate regression model with a singular covariance matrix, which arises for example in complete linear expenditure systems. In that

⁴In the unlikely situation of $s = k$, the dimension of free parameters would be zero, which reflects the fact that the manifold $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$ collapses to the single point $\boldsymbol{\theta} = \boldsymbol{\theta}^0$. As a result, we should be able to recover the true value of the parameters without any sampling variability. In contrast, if $s = 0$, the dimension of fixed parameters would be zero, which reflects the fact that $\mathbf{S}^-(\boldsymbol{\theta}^0) = \mathbf{S}^{-1}(\boldsymbol{\theta}^0)$. As a result, we can estimate $\boldsymbol{\theta}$ by means of the regular GMM methods discussed before.

context, these authors suggest to reduce the number of parameters by exploiting the parametric restrictions that give rise to the singularity of the residual covariance matrix, as well as the number of equations.

The following Proposition formalises the optimality of our approach:

Proposition 1 *Let*

$$\hat{\boldsymbol{\theta}}_T = \arg \min_{\boldsymbol{\theta} \in \Theta} \bar{\mathbf{h}}_T'(\boldsymbol{\theta}) \mathbf{S}^-(\boldsymbol{\theta}^0) \bar{\mathbf{h}}_T(\boldsymbol{\theta}) \text{ s.t. } \mathbf{m}(\boldsymbol{\theta}) = \mathbf{0} \quad (9)$$

denote our proposed GMM estimator of the $k \times 1$ vector of unknown parameter $\boldsymbol{\theta}$ defined by the $n \geq k$ moment conditions $E[\mathbf{h}(\mathbf{x}_t; \boldsymbol{\theta})] = \mathbf{0}$, which satisfy all the usual regularity conditions, together with Assumptions 1 and 2. Then

a) $\boldsymbol{\theta}$ is first-order identified along the manifold $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$.

b) $\hat{\boldsymbol{\theta}}_T$ is asymptotically efficient, in the sense that the difference between its asymptotic covariance matrix and the asymptotic covariance matrix of another GMM estimator of $\boldsymbol{\theta}$ based on any other norm of the moment conditions is negative semidefinite regardless of the weighting matrix Υ_T and regardless of whether or not we impose the equality restrictions $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$.

c)

$$T \cdot \bar{\mathbf{h}}_T'(\hat{\boldsymbol{\theta}}_T) \mathbf{S}^-(\boldsymbol{\theta}^0) \bar{\mathbf{h}}_T(\hat{\boldsymbol{\theta}}_T) \xrightarrow{d} \chi^2(n - k),$$

where $\chi^2(0) = 0$ with probability 1.

d) If $\bar{\boldsymbol{\theta}}_T$ denotes a variant of (9) based on a different choice of generalised inverse $\mathbf{S}^-(\boldsymbol{\theta}^0)$ then

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \bar{\boldsymbol{\theta}}_T) = o_p(1).$$

The first-order identifiability of $\boldsymbol{\theta}$ along the manifold $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$ in part a) guarantees that $\boldsymbol{\theta}^0$ will be locally identified. This result is also necessary for the asymptotic normality of $\hat{\boldsymbol{\theta}}_T$, whose asymptotic covariance matrix will be of rank $k - s$, and the asymptotic χ^2 distribution of the associated J test (see Sargan (1983) and Dovonon and Renault (2009) for a discussion of those cases in which $\boldsymbol{\theta}^0$ remains locally identified but the Jacobian of the moment conditions is of reduced rank).

More importantly, part b) implies that if we simply weighted the original moment conditions by $\mathbf{S}^-(\boldsymbol{\theta}^0)$ without exploiting the equality restrictions implicit in $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$, the resulting estimators and testing procedures would generally be suboptimal because they would give no weight to precisely the s asymptotically degenerate linear combinations of $\sqrt{T} \bar{\mathbf{h}}_T(\boldsymbol{\theta})$.⁵

Part b) also implies that if we imposed the parametric restrictions $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$, but used a weighting matrix whose probability limit differed from $\mathbf{S}^-(\boldsymbol{\theta}^0)$, then the resulting estimators and testing procedures would also be generally suboptimal.

⁵In fact, it may well happen that $\boldsymbol{\theta}$ is not even identified from the reduced set of $n - s$ moment conditions implicitly defined by a generalised inverse, because, for instance, $n - s < k$. After imposing the restriction $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$, on the other hand, the reduced set of moment conditions will locally identify $\boldsymbol{\theta}$ at $\boldsymbol{\theta}^0$.

Although part d) implies that the choice of generalised inverse is asymptotically inconsequential, in practice the simplest solution would be to delete s of the moment conditions that interact with $\mathbf{\Pi}(\boldsymbol{\theta})$ in Assumption 1, which is equivalent to a Cholesky-based generalised inverse for a specific re-ordering of $\mathbf{h}(\mathbf{x}_t; \boldsymbol{\theta})$. However, since we might end up deleting moments that are not really affected by $\mathbf{\Pi}(\boldsymbol{\theta}^0)$, a safer symmetric alternative would be to use the Moore-Penrose inverse. Specifically, let

$$\mathbf{S}(\boldsymbol{\theta}^0) = \begin{bmatrix} \mathbf{P} & \mathbf{Q} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{P}' \\ \mathbf{Q}' \end{bmatrix} = \mathbf{P}\boldsymbol{\Delta}\mathbf{P}' \quad (10)$$

denote the spectral (eigenvalue-eigenvector) decomposition of $\mathbf{S}(\boldsymbol{\theta}^0)$, where $\boldsymbol{\Lambda}$ is a positive definite diagonal matrix of order $n - s$. Given that the Moore-Penrose inverse is simply

$$\mathbf{S}^+(\boldsymbol{\theta}^0) = \mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}',$$

optimal GMM under singularity effectively works with the $n - s$ sample moments $\mathbf{P}'\bar{\mathbf{h}}_T(\boldsymbol{\theta})$, whose asymptotic long-run variance is the non-singular matrix $\boldsymbol{\Lambda}$.

In fact, if we imposed the equality restrictions implicit in $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$ by reparametrising the model in terms of $k - s$ free parameters, as in the proof of Proposition 1, then the asymptotic covariance matrix of these parameters would be given by the usual GMM formulas as long as we substituted $\mathbf{S}^{-1}(\boldsymbol{\theta}^0)$ by $\mathbf{S}^-(\boldsymbol{\theta}^0)$.

Once again, the optimal GMM approach that we have just described is not feasible unless we know $\mathbf{S}^-(\boldsymbol{\theta}^0)$, but under standard regularity conditions, the asymptotics will not change if we replace this matrix by a consistent estimator. However, an estimator of $\mathbf{S}^-(\boldsymbol{\theta}^0)$ must be chosen with some care when $s > 0$ in order to avoid discontinuities in the limit. The reason is the following: as we saw before, if $\dot{\boldsymbol{\theta}}_T$ is an initial consistent estimator of $\boldsymbol{\theta}^0$, then we can easily compute a consistent estimator of $\mathbf{S}(\boldsymbol{\theta}^0)$, say $\mathbf{S}_T(\dot{\boldsymbol{\theta}}_T)$, by means of a HAC covariance matrix estimator based on $\mathbf{h}(\mathbf{x}_t; \dot{\boldsymbol{\theta}}_T)$. But in general, we will not consistently estimate $\mathbf{S}^-(\boldsymbol{\theta}^0)$ in singular cases if $\mathbf{S}_T(\dot{\boldsymbol{\theta}}_T)$ has full rank for finite T .⁶

This is an empirically relevant issue even if $\mathbf{\Pi}'(\boldsymbol{\theta})\mathbf{h}(\mathbf{x}_t, \boldsymbol{\theta}) = \mathbf{m}(\boldsymbol{\theta})$ for all \mathbf{x}_t and for all $\boldsymbol{\theta}$ as in (8) because a researcher who, unaware of the singularity of $\mathbf{S}(\boldsymbol{\theta}^0)$, combines a first step

⁶A trivial non-random example of discontinuities is the sequence $1/T$, which converges to 0 while $(1/T)^- = T$ diverges. Theorem 1 in Andrews (1987) derives the conditions under which a quadratic form based on a g-inverted weighting matrix converges to a chi-square distribution. See also Dufour and Valery (2009) for regularisation procedures that avoid generalised inverses.

estimator such that $\mathbf{m}(\dot{\boldsymbol{\theta}}_T) \neq \mathbf{0}$ with $\mathbb{S}_T(\boldsymbol{\theta})$ instead of $\mathbb{S}_T(\dot{\boldsymbol{\theta}}_T)$, will end up with seemingly optimal estimators and testing procedures whose asymptotic distribution is non-standard.

In the static case that is relevant for our spanning tests, which is characterised by condition (7), we can obtain HAC estimators $\mathbb{S}_T(\dot{\boldsymbol{\theta}}_T)$ or $\mathbb{S}_T(\dot{\boldsymbol{\theta}}_T)$ whose rank is $n - s$ in finite samples by simply using an estimator of $\boldsymbol{\theta}^0$ that belongs to the manifold $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$. For that reason, our proposed two-step optimal GMM estimator in that context would be

$$\ddot{\boldsymbol{\theta}}_T = \arg \min_{\boldsymbol{\theta} \in \Theta} \bar{\mathbf{h}}_T'(\boldsymbol{\theta}) \mathbf{S}_T^- \left(\dot{\boldsymbol{\theta}}_T \right) \bar{\mathbf{h}}_T(\boldsymbol{\theta}) \text{ s.t. } \mathbf{m}(\boldsymbol{\theta}) = \mathbf{0},$$

where $\dot{\boldsymbol{\theta}}_T$ is an initial consistent estimator of $\boldsymbol{\theta}^0$ that satisfies the known equality restrictions $\mathbf{m}(\dot{\boldsymbol{\theta}}_T) = \mathbf{0}$. A similar argument shows that the optimal CU-GMM estimator that exploits the singularity of the second moment matrix in finite samples will be

$$\tilde{\boldsymbol{\theta}}_T = \arg \min_{\boldsymbol{\theta} \in \Theta} \bar{\mathbf{h}}_T'(\boldsymbol{\theta}) \mathbf{S}_T^- (\boldsymbol{\theta}) \bar{\mathbf{h}}_T(\boldsymbol{\theta}) \text{ s.t. } \mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}.$$

In Appendix C.2 we show how to compute this CU-GMM estimator by means of certain OLS regressions which are robust to multicollinearity. In that way, we do not really need to specify ex-ante the number of zero eigenvalues in order to compute a generalised inverse. In addition, our results show that the CU criterion function is numerically invariant to the choice of generalised inverse.

In the more general dynamic case covered by Assumption 1, though, restricting $\boldsymbol{\theta}$ to lie on the manifold is not generally enough to ensure the singularity of $\mathbb{S}_T(\boldsymbol{\theta})$ or $\mathbb{S}_T(\dot{\boldsymbol{\theta}}_T)$ in finite samples. In those cases, we would additionally recommend the elimination of the s smallest eigenvalues and associated eigenvectors of the spectral decomposition of those matrices before computing their generalised inverse.

4 Application to spanning tests

Let \mathbf{R}_1 and \mathbf{R}_2 denote the gross returns to two subsets of N_1 and N_2 assets, respectively, so that the dimension of the expanded set of returns $\mathbf{R} = (\mathbf{R}'_1, \mathbf{R}'_2)'$ is $N = N_1 + N_2$, which we treat as fixed hereinafter in line with the existing literature. Importantly, these assets can be either primitive, like stocks and bonds, or mutual funds managed according to some specific active portfolio strategy. We want to compare the SMVF and RMVF frontiers generated by \mathbf{R}_1 with the ones generated by the whole of \mathbf{R} . In general, when we also consider \mathbf{R}_2 , the RMVF frontier will shift to the left because the available risk-return trade-offs improve, while the SMVF frontier

will rise because there is more information in the data about the underlying SDF. However, this is not always the case. In particular, we say that \mathbf{R}_1 spans the SMVF and/or RMVF generated from \mathbf{R} when the original and extended frontiers coincide.⁷ The purpose of this section is to study spanning tests given panel data on R over T periods. As we shall see, the optimal GMM theory under singularity developed in the previous section is particularly relevant.

4.1 Uncentred cost and mean representing portfolios

Given that the cost and mean representing portfolios defined in (2) span both the SMVF and RMVF, a rather natural way to test for spanning that is not tied down to the specific properties of either frontier consists in studying whether these portfolios are common to the linear spans $\langle \mathbf{R}_1 \rangle$ and $\langle \mathbf{R} \rangle$. In particular, if p_1^* and p_1^+ denote the cost and mean representing portfolios corresponding to $\langle \mathbf{R}_1 \rangle$ then mean-variance spanning of \mathbf{R} by \mathbf{R}_1 is equivalent to

$$\begin{aligned} E(\mathbf{R}p_1^+) &= \boldsymbol{\nu}, \\ E(\mathbf{R}p_1^*) &= \ell_N. \end{aligned} \tag{11}$$

Therefore, we can easily test for spanning by checking whether the portfolios of \mathbf{R}_1 that best mimic both the safe asset and the stochastic discount factor continue to do so when we also consider the assets in \mathbf{R}_2 .

If $\langle \mathbf{R}_1 \rangle$ and $\langle \mathbf{R} \rangle$ only share the same mean representing portfolio, and p^* and p^+ are not orthogonal, then the two RMVF's are tangent at the point that corresponds to R^+ . In contrast, the two SMVF's will have no common point, but they will share the asymptotes, and the location of the global minimum (see Figure 1). On the other hand, if $\langle \mathbf{R}_1 \rangle$ and $\langle \mathbf{R} \rangle$ only share the same cost representing portfolio, then R^* and p^* will be the common elements of the frontiers generated from \mathbf{R}_1 alone, and the ones generated from \mathbf{R} (see Figure 2). Thus, if we add both conditions, the old and new frontiers will be equal.

Given the moments in (11), we could perform a spanning test on the basis of the overidentified system of moment conditions

$$E \begin{pmatrix} \mathbf{R}\mathbf{R}'_1\phi_1^+ - \boldsymbol{\nu} \\ \mathbf{R}\mathbf{R}'_1\phi_1^* - \ell_N \\ \mathbf{R} - \boldsymbol{\nu} \end{pmatrix} = E \begin{bmatrix} \mathbf{h}_V(\mathbf{R}; \phi_1, \boldsymbol{\nu}) \\ \mathbf{h}_M(\mathbf{R}; \boldsymbol{\nu}) \end{bmatrix} = E[\mathbf{g}_U(\mathbf{R}; \phi_1, \boldsymbol{\nu})] = \mathbf{0},$$

⁷A third, and last, possibility is that the original and extended frontiers touch at a single point. Although it is common in the literature to refer to this situation as “intersection”, we prefer to use the word “tangency” because the frontiers are never secant to each other, as the word “intersection” may suggest. We discuss this case in detail in Peñaranda and Sentana (2004).

where we have defined ϕ_1 as $(\phi_1^{+'}, \phi_1^{*'})'$ and added N estimating functions for the generally unknown vector of expected returns ν . Unfortunately, there is a singularity that precludes the use of standard GMM methods:

Lemma 1 *Let $\Pi_U(\phi_1) = [(-\phi_1^{*'}, \mathbf{0}'_{N_2}), (\phi_1^{+'}, \mathbf{0}'_{N_2}), \mathbf{0}'_N]'$ and $m_U(\phi_1, \nu) = \phi_1^{*'}\nu - \phi_1^{+'}\ell_{N_1}$. Then,*

$$\Pi_U'(\phi_1)\mathbf{g}_U(\mathbf{R}; \phi_1, \nu) = m_U(\phi_1, \nu) \forall \mathbf{R} \text{ and } \forall \phi_1.$$

Hence $\Pi_U'(\phi_1)\mathbf{g}_U(\mathbf{R}; \phi_1, \nu) = \mathbf{0} \forall \mathbf{R}$ if and only if $m_U(\phi_1, \nu) = 0$.

Given that the true parameters will satisfy $m_U(\phi_1^0, \nu^0) = 0$ in view of (11), this lemma implies that the rank of the asymptotic covariance matrix of $\sqrt{T}\bar{\mathbf{g}}_{UT}(\phi_1^0, \nu^0)$ will be $3N - 1$ instead of $3N$. In this case, in fact, the estimating functions $\mathbf{g}_U(\mathbf{R}; \phi_1, \nu)$ satisfy (8), which means that $\mathfrak{S}_T(\theta)$ in (5) will be singular regardless of θ . As we mentioned in the previous section, though, most empirical researchers will instead use $\mathbb{S}_T(\theta)$ in (4), which will not be singular unless θ satisfies $m_U(\phi_1, \nu) = 0$. In this sense, note that all parameter values along this implicit manifold satisfy a well known property of the uncentred representing portfolios, namely

$$E(p_1^*) = E(p_1^*p_1^+) = C(p_1^+).$$

Our optimal GMM procedure will combine the equality restriction $\phi_1^{*'}\nu - \phi_1^{+'}\ell_{N_1} = 0$ with a generalised inverse to optimally deal with the singularity in Lemma 1, so that under the null of spanning the resulting J test will have $2N_2$ degrees of freedom. It turns out that in this particular example our optimal procedure implicitly transforms the estimating functions $\mathbf{g}_U(\mathbf{R}; \phi_1, \nu)$ into a smaller system of estimating functions in which one can apply standard (i.e. non-singular) optimal GMM procedures:

Proposition 2 *1. The J test based on the moment conditions $E[\mathbf{g}_U(\mathbf{R}; \phi_1, \nu)] = \mathbf{0}$ is asymptotically equivalent under the null of spanning and sequences of local alternatives to the J test based on the reduced set of moment conditions*

$$E \begin{pmatrix} \mathbf{R}\mathbf{R}'_1\phi_1^+ - \mathbf{R} \\ \mathbf{R}\mathbf{R}'_1\phi_1^* - \ell_N \end{pmatrix} = E[\mathbf{h}_U(\mathbf{R}; \phi_1)] = \mathbf{0}. \quad (12)$$

2. Both tests numerically coincide if we use CU-GMM for a specific choice of HAC estimator.

The main advantage of two-step GMM in this context is that the linearity of the estimating equations (12) in ϕ_1 gives rise to closed form expressions for a fixed weighting matrix $\mathbf{S}_T(\phi_{1T})$. And although the choice of initial consistent estimator $\dot{\phi}_{1T}$ does not affect the asymptotic distribution of two-step GMM estimators up to $O_p(T^{-1/2})$ terms, there is some Monte Carlo

evidence suggesting that their finite sample properties can be negatively affected by an arbitrary choice of initial weighting matrix such as the identity (see e.g. Kan and Zhou (2001)). In contrast, CU-GMM (12) induces a non-linearity in the GMM objective function, which highlights the need for good initial values.

For those reasons, we propose a computationally simple intuitive estimator of ϕ_1 that is always consistent, but which would become efficient for a data generating process that guarantees the compatibility of mean-variance preferences with expected utility maximisation regardless of investors' preferences:⁸

Lemma 2 *If \mathbf{R}_t is an i.i.d. elliptical random vector with mean $\boldsymbol{\nu}$, covariance matrix $\boldsymbol{\Sigma}$, and bounded fourth moments, then the linear combinations of the moment conditions in (12) that provide the most efficient estimators of ϕ_1^+ and ϕ_1^* under spanning will be given by*

$$E \begin{pmatrix} \mathbf{R}_1 \mathbf{R}'_1 \phi_1^+ - \mathbf{R}_1 \\ \mathbf{R}_1 \mathbf{R}'_1 \phi_1^* - \ell_{N_1} \end{pmatrix} = \mathbf{0},$$

so that $\dot{\phi}_{1T}^+ = \hat{\boldsymbol{\Gamma}}_{11T}^{-1} \hat{\boldsymbol{\nu}}_{1T}$ and $\dot{\phi}_{1T}^* = \hat{\boldsymbol{\Gamma}}_{11T}^{-1} \ell_{N_1}$, where $\hat{\boldsymbol{\nu}}_{1T} = T^{-1} \sum_{t=1}^T \mathbf{R}_{1t}$ is the sample mean and $\hat{\boldsymbol{\Gamma}}_{11T} = T^{-1} \sum_{t=1}^T \mathbf{R}_{1t} \mathbf{R}'_{1t}$ the sample second moment matrix of \mathbf{R}_1 .

Intuitively, this means that under those circumstances the blocks involving \mathbf{R}_1 exactly identify the parameters ϕ_1^* and ϕ_1^+ , while the blocks corresponding to \mathbf{R}_2 provide the $2N_2$ overidentification restrictions to test. Although the elliptical family is rather broad (see e.g. Fang, Kotz and Ng (1990)), and includes the multivariate normal and t distribution as special cases, it is important to stress that $\dot{\phi}_{1T}^+$ and $\dot{\phi}_{1T}^*$ will remain consistent under H_0 even if the assumptions of serial independence and ellipticity are not totally realistic in practice, unlike the semiparametric estimators used by Vorkink (2003), or the parametric estimators studied by Amengual and Sentana (2010).

4.2 Centred cost and mean representing portfolios

As we discussed in section 2.1, in the standard situation in which the SMVF and RMVF are non-degenerate hyperbolas, we can define an alternative pair of mean and cost representing portfolios (3) in terms of central moments, which also span both mean variance frontiers. Therefore, another rather natural way to test for spanning that is not tied down to the specific properties of either frontier consists in studying whether these portfolios are common to the linear spans $\langle \mathbf{R}_1 \rangle$ and $\langle \mathbf{R} \rangle$. In particular, if p_1^{**} and p_1^{++} denote the centred cost and mean representing

⁸See Chamberlain (1983) and Owen and Rabinovitch (1983).

portfolios corresponding to $\langle \mathbf{R}_1 \rangle$ then mean-variance spanning of \mathbf{R} by \mathbf{R}_1 is equivalent to

$$\begin{aligned} cov(\mathbf{R}, p_1^{++}) &= \boldsymbol{\nu}, \\ cov(\mathbf{R}, p_1^{**}) &= \ell_N. \end{aligned} \tag{13}$$

The graphical implication of sharing the centred mean representing portfolio has already been explained in section 4.1 in terms of R^+ when p^* and p^+ are not orthogonal, because p^{++} is proportional to p^+ (see Figure 1). In contrast, the reduced and expanded RMVF's will share the minimum variance return R^{**} if $p^{**} = p_1^{**}$, while the original and extended SMVF's will share $m^{MV}(0)$, which is the value at the origin (see Figure 3). Hence, if we add both conditions, it is once more clear that the original and expanded frontiers must be equal.

Given the moments in (13), we can perform a spanning test on the basis of the alternative system of overidentified moment conditions:

$$E \begin{bmatrix} (\mathbf{R} - \boldsymbol{\nu})(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^+ - \boldsymbol{\nu} \\ (\mathbf{R} - \boldsymbol{\nu})(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^* - \ell_N \\ \mathbf{R} - \boldsymbol{\nu} \end{bmatrix} = E[\mathbf{g}_C(\mathbf{R}; \boldsymbol{\varphi}_1, \boldsymbol{\nu})] = \mathbf{0}, \tag{14}$$

where we have again added N estimating functions for the vector of expected returns $\boldsymbol{\nu}$, and defined $\boldsymbol{\varphi}_1$ as $(\boldsymbol{\varphi}_1^+, \boldsymbol{\varphi}_1^*)'$. Unfortunately, this system is also singular:

Lemma 3 *Let $\boldsymbol{\Pi}_C(\boldsymbol{\varphi}_1) = [(-\boldsymbol{\varphi}_1^{*'}, \mathbf{0}'_{N_2}), (\boldsymbol{\varphi}_1^{+'}, \mathbf{0}'_{N_2}), \mathbf{0}'_N]'$ and $m_C(\boldsymbol{\varphi}_1, \boldsymbol{\nu}_1) = \boldsymbol{\varphi}_1^{*'} \boldsymbol{\nu}_1 - \boldsymbol{\varphi}_1^{+'} \ell_{N_1}$. Then,*

$$\boldsymbol{\Pi}'_C(\boldsymbol{\varphi}_1) \mathbf{g}_C(\mathbf{R}; \boldsymbol{\varphi}_1, \boldsymbol{\nu}) = m_C(\boldsymbol{\varphi}_1, \boldsymbol{\nu}_1) \quad \forall \mathbf{R} \text{ and } \forall \boldsymbol{\varphi}_1.$$

Hence $\boldsymbol{\Pi}'_C(\boldsymbol{\varphi}_1) \mathbf{g}_C(\mathbf{R}; \boldsymbol{\varphi}_1, \boldsymbol{\nu}) = \mathbf{0} \quad \forall \mathbf{R}$ if and only if $m_C(\boldsymbol{\varphi}_1, \boldsymbol{\nu}_1) = 0$.

Given that the true parameters will satisfy $m_C(\boldsymbol{\varphi}_1^0, \boldsymbol{\nu}_1^0) = 0$ in view of (13), this lemma implies that the rank of the asymptotic covariance matrix of $\sqrt{T} \bar{\mathbf{g}}_{CT}(\boldsymbol{\varphi}_1^0, \boldsymbol{\nu}_1^0)$ is also $3N - 1$ instead of $3N$. As in the case of the uncentred representing portfolios discussed in the previous subsection, the estimating functions $\mathbf{g}_C(\mathbf{R}; \boldsymbol{\varphi}_1, \boldsymbol{\nu})$ also satisfy (8), which means that $\mathfrak{S}_T(\boldsymbol{\theta})$ will be singular regardless of $\boldsymbol{\theta}$. But again, most empirical researchers will use $\mathbb{S}_T(\boldsymbol{\theta})$ instead, which will not be singular unless $\boldsymbol{\theta}$ satisfies $m_C(\boldsymbol{\varphi}_1, \boldsymbol{\nu}_1) = 0$. In this sense, note that all parameter values along this implicit manifold satisfy the well known property of the centred representing portfolios

$$E(p_1^{**}) = cov(p_1^{**}, p_1^{++}) = C(p_1^{++}).$$

Our optimal GMM procedure will again combine the equality restriction $\boldsymbol{\varphi}_1^{*'} \boldsymbol{\nu}_1 - \boldsymbol{\varphi}_1^{+'} \ell_{N_1} = 0$ with a generalised inverse to optimally deal with the singularity in Lemma 3, so that under

the null of spanning the resulting J test will also have $2N_2$ degrees of freedom. However, the non-linearity in parameters of the moment conditions (14) implies that in this case we cannot transform the problem in a smaller set of non-singular moment conditions. Nevertheless, we can still simplify the computations by concentrating $\boldsymbol{\nu}_2$ out of the criterion function:

Proposition 3 1. *The J test based on the singular moment conditions $E[\mathbf{g}_C(\mathbf{R}; \boldsymbol{\varphi}_1, \boldsymbol{\nu})] = \mathbf{0}$ is asymptotically equivalent under the null of spanning and local alternatives to the J test based on the singular moment conditions*

$$E \begin{bmatrix} \mathbf{R}(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^+ - \mathbf{R} \\ \mathbf{R}(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^* - \ell_N \\ \mathbf{R}_1 - \boldsymbol{\nu}_1 \end{bmatrix} = E \begin{bmatrix} \mathbf{h}_C(\mathbf{R}; \boldsymbol{\varphi}_1, \boldsymbol{\nu}_1) \\ \mathbf{h}_M(\mathbf{R}_1; \boldsymbol{\nu}_1) \end{bmatrix} = E[\mathbf{f}_C(\mathbf{R}; \boldsymbol{\varphi}_1, \boldsymbol{\nu}_1)] = \mathbf{0}. \quad (15)$$

2. *Both tests numerically coincide if we use CU-GMM for a specific choice of HAC estimator.*

Since the optimal GMM procedure in section 3 applied to (15) imposes the equality restriction $\boldsymbol{\varphi}_1^{*'} \boldsymbol{\nu}_1 - \boldsymbol{\varphi}_1^{+'} \ell_{N_1} = 0$, we will effectively pin down one of the parameters in $(\boldsymbol{\varphi}_1, \boldsymbol{\nu}_1)$ given the others. As for the remaining parameters, we will use a consistent estimator of a generalised inverse of the asymptotic covariance matrix of $\sqrt{T} \mathbf{f}_{CT}(\boldsymbol{\varphi}_1^0, \boldsymbol{\nu}_1^0)$, say $\mathbf{S}_C(\boldsymbol{\varphi}_1^0, \boldsymbol{\nu}_1^0)$, which effectively eliminates the singular linear combination. As discussed in that section, though, in order to obtain a consistent estimator of $\mathbf{S}_C^-(\boldsymbol{\varphi}_1^0, \boldsymbol{\nu}_1^0)$, we need a consistent estimator of $\mathbf{S}_C(\boldsymbol{\varphi}_1^0, \boldsymbol{\nu}_1^0)$ that is singular in finite samples. The following result justifies an obvious first-step estimator for two-step GMM, which can also be used as initial condition for CU-GMM:

Lemma 4 *If \mathbf{R}_t is an i.i.d. elliptical random vector with mean $\boldsymbol{\nu}$, covariance matrix $\boldsymbol{\Sigma}$, and bounded fourth moments, then the linear combinations of the moment conditions in (15) that provide the most efficient estimators of $\boldsymbol{\varphi}_1^+$ and $\boldsymbol{\varphi}_1^*$ under spanning will be given by*

$$E \begin{bmatrix} \mathbf{R}_1(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^+ - \mathbf{R}_1 \\ \mathbf{R}_1(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^* - \ell_{N_1} \\ \mathbf{R}_1 - \boldsymbol{\nu}_1 \end{bmatrix} = \mathbf{0},$$

so that $\bar{\boldsymbol{\nu}}_{1T} = \hat{\boldsymbol{\nu}}_{1T}$, $\bar{\boldsymbol{\varphi}}_{1T}^+ = \hat{\boldsymbol{\Sigma}}_{11T}^{-1} \hat{\boldsymbol{\nu}}_{1T}$ and $\bar{\boldsymbol{\varphi}}_{1T}^* = \hat{\boldsymbol{\Sigma}}_{11T}^{-1} \ell_{N_1}$, where $\hat{\boldsymbol{\Sigma}}_{11T} = \hat{\boldsymbol{\Gamma}}_{11T} - \hat{\boldsymbol{\nu}}_{1T} \hat{\boldsymbol{\nu}}_{1T}'$ is the sample covariance matrix of \mathbf{R}_1 .

Intuitively, this means that under those circumstances, the blocks involving \mathbf{R}_1 exactly identify $\boldsymbol{\nu}_1$, $\boldsymbol{\varphi}_1^+$ and $\boldsymbol{\varphi}_1^*$, while the blocks corresponding to \mathbf{R}_2 provide the $2N_2$ testable restrictions. But note again that $\bar{\boldsymbol{\varphi}}_{1T}^+$ and $\bar{\boldsymbol{\varphi}}_{1T}^*$ will remain consistent under H_0 even if the assumptions of serial independence and ellipticity are not totally realistic in practice. Note also that the first-step estimator defined in Lemma 4 does indeed guarantee that $\mathbf{S}_C(\bar{\boldsymbol{\varphi}}_{1T}, \bar{\boldsymbol{\nu}}_{1T})$ will be singular because the linear combination defined in Lemma 3 only involves those blocks under the null of spanning, and $\bar{\boldsymbol{\varphi}}_{1T}^{+'} \ell_{N_1} - \bar{\boldsymbol{\varphi}}_{1T}^{*'} \bar{\boldsymbol{\nu}}_{1T} = \mathbf{0}$.

4.3 Regression tests for spanning

The centred and uncentred representing portfolios constitute rather natural choices for testing for mean-variance spanning. However, there are infinitely many more pairs of portfolios that could be used for the same purposes, because the two fund spanning property of both frontiers does not depend on the particular funds used.

Building on Jobson and Korkie (1982), Gibbons, Ross and Shanken (1989) and Huberman and Kandel (1987) showed that in mean-standard deviation space, the RMVF generated by \mathbf{R}_1 and \mathbf{R} coincide at the point of tangency with a ray that starts from $(0, c_i^{-1})$ if and only if the intercepts in the multivariate regression of $(\mathbf{R}_2 - c_i^{-1}\ell_{N_2})$ on a constant and $(\mathbf{R}_1 - c_i^{-1}\ell_{N_1})$ are all 0.⁹ Consequently, we can implement the corresponding tangency test by means of the following set of overidentified moment conditions:

$$E \left[\begin{pmatrix} 1 \\ \mathbf{R}_1 - c_i^{-1}\ell_{N_1} \end{pmatrix} \otimes [(\mathbf{R}_2 - c_i^{-1}\ell_{N_2}) - \mathbf{B}_i(\mathbf{R}_1 - c_i^{-1}\ell_{N_1})] \right] = E[\mathbf{h}_J(\mathbf{R}; \mathbf{b}_i)] = \mathbf{0},$$

where $\mathbf{b}_i = \text{vec}(\mathbf{B}_i)$.

Therefore, a rather natural way to test for spanning in the RMVF is to test if there is simultaneous tangency at two points. Specifically, let c_i^{-1} and c_{ii}^{-1} , with $c_i \neq c_{ii}$, denote two arbitrary expected returns on the shadow safe asset. Then, a spanning test can be based on the $2N_2(N_1 + 1)$ moment conditions:

$$E \begin{bmatrix} \mathbf{h}_J(\mathbf{R}; \mathbf{b}_i) \\ \mathbf{h}_J(\mathbf{R}; \mathbf{b}_{ii}) \end{bmatrix} = E[\mathbf{h}_L(\mathbf{R}; \mathbf{b}_i, \mathbf{b}_{ii})] = \mathbf{0}. \quad (16)$$

However, as pointed out by Marín (1996), the asymptotic covariance matrix of the sample analogues of (16) is singular under the null of spanning. More explicitly:

Lemma 5 *Let*

$$\mathbf{\Pi}_L(\mathbf{b}_i, \mathbf{b}_{ii}) = \begin{pmatrix} \mathbf{\Phi}_i'^{-1} \otimes \mathbf{I}_{N_2} \\ -\mathbf{\Phi}_{ii}'^{-1} \otimes \mathbf{I}_{N_2} \end{pmatrix}$$

and

$$\mathbf{m}_L(\mathbf{b}_i, \mathbf{b}_{ii}) = \begin{bmatrix} -c_{ii}^{-1}(\mathbf{B}_{ii}\ell_{N_1} - \ell_{N_2}) + c_i^{-1}(\mathbf{B}_i\ell_{N_1} - \ell_{N_2}) \\ \mathbf{b}_{ii} - \mathbf{b}_i \end{bmatrix},$$

where

$$\mathbf{\Phi}_i = \begin{pmatrix} 1 & \mathbf{0}' \\ -c_i^{-1}\ell_{N_1} & \mathbf{I}_{N_1} \end{pmatrix}$$

⁹If we regard c_i^{-1} as the expected return of a zero-beta frontier portfolio orthogonal to the tangency portfolio made up of elements of \mathbf{R}_1 only, then we can interpret the regression intercepts as the so-called Jensen's alphas in the portfolio evaluation literature. These coefficients should all be 0 if the tangency portfolio of \mathbf{R}_1 is really mean-variance efficient with respect to \mathbf{R} (see De Roon and Nijman (2001)).

is a known full rank square matrix of order $N_1 + 1$ whose only role is to linearly transform the non-constant regressors from $(\mathbf{R}_1 - c_i^{-1}\ell_{N_1})$ to \mathbf{R}_1 , with Φ_{ii} similarly defined. Given that

$$\begin{aligned} & \Pi'_L(\mathbf{b}_i, \mathbf{b}_{ii})\mathbf{h}_L(\mathbf{R}; \mathbf{b}_i, \mathbf{b}_{ii}) = \\ & = \begin{pmatrix} 1 \\ \mathbf{R}_1 \end{pmatrix} \otimes [(\mathbf{B}_{ii} - \mathbf{B}_i)\mathbf{R}_1 + c_i^{-1}(\mathbf{B}_i\ell_{N_1} - \ell_{N_2}) - c_{ii}^{-1}(\mathbf{B}_{ii}\ell_{N_1} - \ell_{N_2})], \end{aligned}$$

then $\Pi'_L(\mathbf{b}_i, \mathbf{b}_{ii})\mathbf{h}_L(\mathbf{R}; \mathbf{b}_i, \mathbf{b}_{ii}) = \mathbf{0} \quad \forall \mathbf{R}$ if and only if $\mathbf{m}_L(\mathbf{b}_i, \mathbf{b}_{ii}) = \mathbf{0}$.

Since $\mathbf{m}_L(\mathbf{b}_i, \mathbf{b}_{ii})$ is 0 at the true values, the previous lemma implies that the rank of the asymptotic covariance matrix of $\bar{\mathbf{h}}_{LT}(\mathbf{b}_i^0, \mathbf{b}_{ii}^0)$ will be $N_2(N_1 + 1)$ instead of $2N_2(N_1 + 1)$. In this sense, the singularity in Lemma 5 shows important differences with respect to those in Lemmas 1 and 3. First, there are multiple singularities, in fact half of the moments. Second, the manifold imposes two sets of constraints on the matrices of slope coefficients of the two multivariate regressions under the null of spanning, namely $\mathbf{B}_{ii} = \mathbf{B}_i = \mathbf{B}$ and $\ell_{N_2} = \mathbf{B}\ell_{N_1}$. Third, and more importantly, $\Pi'_L(\mathbf{b}_i, \mathbf{b}_{ii})\mathbf{h}_L(\mathbf{R}; \mathbf{b}_i, \mathbf{b}_{ii})$ is generally a non-constant function of \mathbf{R}_1 over the parameter space, and hence even the centred estimator of the asymptotic covariance matrix $\mathfrak{S}_T(\boldsymbol{\theta})$ will usually have full rank. But even if we impose that $\mathbf{B}_{ii} = \mathbf{B}_i = \mathbf{B}$ to make $\Pi'_L(\mathbf{b}_i, \mathbf{b}_{ii})\mathbf{h}_L(\mathbf{R}; \mathbf{b}_i, \mathbf{b}_{ii})$ constant, which will guarantee a singular $\mathfrak{S}_T(\boldsymbol{\theta})$, $\mathfrak{S}_T(\boldsymbol{\theta})$ will not be singular unless we add the additional set of equality restrictions $\mathbf{B}\ell_{N_1} = \ell_{N_2}$. In addition, note that $\Pi'_L(\mathbf{b}_i, \mathbf{b}_{ii})\mathbf{h}_L(\mathbf{R}; \mathbf{b}_i, \mathbf{b}_{ii})$ will not coincide with $\mathbf{m}_L(\mathbf{b}_i, \mathbf{b}_{ii})$ even when both \mathbf{b} 's are equal.

Our optimal GMM procedure applied to (16) will combine the equality restrictions $\mathbf{b}_i = \mathbf{b}_{ii} = \mathbf{b}$ and $\mathbf{B}\ell_{N_1} = \ell_{N_2}$ with a generalised inverse to optimally deal with the singularity in Lemma 5, so that under the null of spanning the resulting J test will have $2N_2$ degrees of freedom. But as in the case of uncentred representing portfolios, it turns out that in this particular example our optimal procedure implicitly transforms the estimating functions $\mathbf{h}_L(\mathbf{R}; \mathbf{b}_i, \mathbf{b}_{ii})$ into a smaller system of estimating functions in which one can apply standard (i.e. non-singular) optimal GMM procedures:

Proposition 4 1. *The J test based on the moment conditions (16) is asymptotically equivalent to the J test based on the $N_2(N_1 + 1)$ moment conditions*

$$E \left\{ \begin{pmatrix} 1 \\ R_{10} \\ \mathbf{R}_{11} - R_{10}\ell_{N_1-1} \end{pmatrix} \otimes \begin{bmatrix} (\mathbf{R}_2 - R_{10}\ell_{N_2}) - \\ -\mathbf{B}_1(\mathbf{R}_{11} - R_{10}\ell_{N_1-1}) \end{bmatrix} \right\} = E[\mathbf{h}_H(\mathbf{R}; \mathbf{b}_1)] = \mathbf{0}, \quad (17)$$

under the null of spanning and local alternatives for any choice of reference portfolio R_{10} , where we have partitioned $\mathbf{B} = (\mathbf{b}_0, \mathbf{B}_1)$ and $\ell_{N_1} = (1, \ell'_{N_1-1})'$ conformably with $\mathbf{R}_1 = (R_{10}, \mathbf{R}'_{11})'$, and $\mathbf{b}_1 = \text{vec}(\mathbf{B}_1)$.

2. *Both tests numerically coincide if we use CU-GMM for a specific choice of HAC estimator.*

Therefore, the optimal combination of two GMM-based return mean variance efficiency tests at two separate points reduces to the GMM version of the spanning test in Huberman and Kandel (1987) discussed by Ferson, Foerster and Kim (1993). In fact, those authors derived the F version of the likelihood ratio test,¹⁰ whose finite sample distribution is exact under the assumption that the distribution of \mathbf{R}_{2t} given \mathbf{R}_{1s} ($s = 1, \dots, T$) is multivariate normal with linear mean and constant covariance matrix.

In practice, we need an initial consistent estimator of \mathbf{b}_1 either to calculate the optimal weighting matrix in a two-step GMM procedure, or to use it as a good initial condition for CU-GMM. Our next lemma suggests some sensible ways of doing so:

Lemma 6 *If \mathbf{R}_t is an i.i.d. elliptical random vector with mean $\boldsymbol{\nu}$, covariance matrix $\boldsymbol{\Sigma}$, bounded fourth moments, and coefficient of multivariate excess kurtosis $\kappa < \infty$, then the linear combinations of the moment conditions (17) that provide the most efficient estimators of \mathbf{b}_1 under spanning will be given by*

$$E \left\{ \left[\begin{array}{c} (\mathbf{R}_{11} - R_{10}\ell_{N_1-1}) \\ +\kappa(\boldsymbol{\nu}_{11} - \nu_{10}\ell_{N_1-1}) \end{array} \right] \otimes \left[\begin{array}{c} (\mathbf{R}_2 - R_{10}\ell_{N_2})- \\ -\mathbf{B}_1(\mathbf{R}_{11} - R_{10}\ell_{N_1-1}) \end{array} \right] \right\} = \mathbf{0}. \quad (18)$$

Since $\boldsymbol{\nu}$ and κ are unknown, we could set κ to 0, which is its value under Gaussianity, in which case the first-step estimator of \mathbf{B}_1 will come from the multivariate regression of $(\mathbf{R}_2 - R_{10}\ell_{N_2})$ on $(\mathbf{R}_{11} - R_{10}\ell_{N_1-1})$. Alternatively, we could use the sample analogues of $\boldsymbol{\nu}$ and κ to obtain an IV estimator of \mathbf{B}_1 from (18).¹¹ In either case, such first-step estimators will remain consistent under the null of spanning even if those assumptions are not totally realistic in practice.

5 Comparisons of spanning tests

So far, we have introduced two new separate families of spanning tests: centred and uncentred representing portfolios. We also encountered a third separate family of spanning tests based on regressions. Now we extensively compare the different tests.

5.1 Equivalence of the parametric restrictions

As we have already seen, the parametric restrictions involved in the novel testing procedures proposed in sections 4.1 and 4.2 simply mean that the centred or uncentred cost and mean representing portfolios of \mathbf{R} depend exclusively on \mathbf{R}_1 . Given that the SMVF is spanned by

¹⁰However, both Peñaranda (1999) and Kan and Zhou (2001) noticed a typo in their paper, whereby a square root is missing in the ratio of determinants of the residual variances.

¹¹It is trivial to compute the sample analogue of the coefficient of multivariate excess kurtosis of any random vector \mathbf{R} , which is defined as $\kappa = E[(\mathbf{R} - \boldsymbol{\nu})' \boldsymbol{\Sigma}^{-1} (\mathbf{R} - \boldsymbol{\nu})]^2 / [N(N+2)] - 1$ (see Mardia (1970)).

either pair of representing portfolios, it is straightforward to show that those restrictions are equivalent to the parametric restrictions tested by De Santis (1993, 1995), Bekaert and Urias (1996), and Kan and Zhou (2001), which amount to the hypothesis that the SMVF of \mathbf{R}_1 is also valid for \mathbf{R}_2 . In turn, Ferson (1995) and Bekaert and Urias (1996) showed that these SMVF parametric restrictions are equivalent to the restrictions tested by Huberman and Kandel (1987), which can be interpreted as saying that each element of \mathbf{R}_2 can be written as a unit cost portfolio of \mathbf{R}_1 , plus an orthogonal arbitrage portfolio with zero mean.¹²

These equivalences can be seen more formally if we write:

$$\begin{aligned}
p^+ &= p_1^+ + (1 + \boldsymbol{\nu}'_1 \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\nu}_1)^{-1} \mathbf{a}' \boldsymbol{\Lambda}^{-1} \mathbf{v}, \\
p^* &= p_1^* + (\ell_{N_2} - \mathbf{C} \ell_{N_1})' \boldsymbol{\Lambda}^{-1} \mathbf{v}, \\
p^{++} &= p_1^{++} + (1 + \boldsymbol{\nu}'_1 \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\nu}_1)^{-1} \mathbf{a}' \boldsymbol{\Omega}^{-1} \mathbf{w}, \\
p^{**} &= p_1^{**} + (\ell_{N_2} - \mathbf{B} \ell_{N_1}) \boldsymbol{\Omega}^{-1} \mathbf{w},
\end{aligned} \tag{19}$$

where $\mathbf{v} = \mathbf{R}_2 - \mathbf{C} \mathbf{R}_1$, $\mathbf{C} = \boldsymbol{\Gamma}_{21} \boldsymbol{\Gamma}_{11}^{-1}$ and $\boldsymbol{\Lambda} = \boldsymbol{\Gamma}_{22} - \boldsymbol{\Gamma}_{21} \boldsymbol{\Gamma}_{11}^{-1} \boldsymbol{\Gamma}'_{21}$ are the residual, slope and residual second moment, respectively, of the least squares projection of \mathbf{R}_2 on $\langle \mathbf{R}_1 \rangle$, while $\mathbf{w} = \mathbf{R}_2 - \mathbf{a} - \mathbf{B} \mathbf{R}_1$, $\mathbf{a} = \boldsymbol{\nu}_2 - \mathbf{B} \boldsymbol{\nu}_1$, $\mathbf{B} = \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1}$ and $\boldsymbol{\Omega} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}'_{21}$ are the residual, intercept, slope and residual variance, respectively, of the projection of \mathbf{R}_2 on $\langle 1, \mathbf{R}_1 \rangle$. From here, it immediately follows that

$$\begin{aligned}
p^+ &= p_1^+ \Leftrightarrow p^{++} = p_1^{++} \Leftrightarrow \mathbf{a} = \mathbf{0}, \\
p^* &= p_1^* \Leftrightarrow \ell_{N_2} = \mathbf{C} \ell_{N_1}, \\
p^{**} &= p_1^{**} \Leftrightarrow \ell_{N_2} = \mathbf{B} \ell_{N_1}.
\end{aligned}$$

Further, if two of these parametric restrictions are satisfied, so will be the third one, as

$$\ell_{N_2} - \mathbf{B} \ell_{N_1} = (\ell_{N_2} - \mathbf{C} \ell_{N_1}) + (\boldsymbol{\nu}'_1 \boldsymbol{\Sigma}_{11}^{-1} \ell_{N_1}) \mathbf{a} / (1 + \boldsymbol{\nu}'_1 \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\nu}_1).$$

5.2 Equivalence of the tests

The fact that the restrictions to test are equivalent does not necessarily imply that the corresponding GMM-based test statistics will be equivalent too. This is particularly true in the case of the regression versions of the tests, in which the number of moment and parameters involved is different, although the number of degrees of freedom of the overidentification tests is the same. The purpose of this subsection is to investigate the asymptotic and finite sample

¹²Kan and Zhou (2001) also linked the constraints in Huberman and Kandel (1987) to R^+ and R^{**} , although they did not relate those returns to representing portfolios.

relationships between the overidentification tests previously discussed under the null of spanning and sequences of local alternatives. The following proposition provides a very precise answer:

Proposition 5 1. *The J tests based on each of the following sets of moment conditions:*

$$\begin{aligned} E[\mathbf{h}_U(\mathbf{R}; \boldsymbol{\phi}_1)] &= \mathbf{0}, \\ E[\mathbf{f}_C(\mathbf{R}; \boldsymbol{\varphi}_1, \boldsymbol{\nu}_1)] &= \mathbf{0}, \\ E[\mathbf{h}_H(\mathbf{R}; \mathbf{b}_1)] &= \mathbf{0}, \end{aligned}$$

are asymptotically equivalent under the null of spanning and compatible sequences of local alternatives.

2. *All three tests numerically coincide if we use CU-GMM for a specific choice of HAC estimator.*

Therefore, there is no basis to prefer one test to another from this perspective because either all the statistics asymptotically converge to exactly the same random variable, or they are numerically equal to begin with if we use CU-GMM. In fact, as we mentioned before, such a numerical equivalence holds with other single-step estimators such as empirical likelihood or exponentially-tilted methods. Once again, though, it is important to emphasise the need to properly define overidentification tests in the presence of singularities to obtain our equivalence results.

Note that our asymptotic result is valid as long as the asymptotic distributions of the different tests are standard, which happens under fairly weak assumptions on the distribution of asset returns, as we saw in section 3. For obvious reasons, our numerical equivalence result does not depend either on the true return distribution.

In order to provide some intuition for the numerical equivalence between all the CU-GMM J tests that we have discussed, imagine that for estimation purposes we assumed that the joint distribution of asset returns is *i.i.d.* multivariate normal. In that context, we could test for spanning by means of the LR test, because the null of spanning would allow us to parametrise $\boldsymbol{\nu}_2$ and $\boldsymbol{\Sigma}_{21}$ as restricted non-linear functions of $\boldsymbol{\nu}_1$ and $\boldsymbol{\Sigma}_{11}$. We could then factorise the joint log-likelihood function of \mathbf{R}_1 and \mathbf{R}_2 as the marginal log-likelihood of \mathbf{R}_1 , whose parameters would be unrestricted, and the conditional log-likelihood of \mathbf{R}_2 given \mathbf{R}_1 . As a result, the LR version of the original Huberman and Kandel (1987) test would be numerically identical to the LR test in the joint system irrespective of whether we parametrise the joint covariance matrix as $\boldsymbol{\Sigma}$ or $\boldsymbol{\Gamma} - \boldsymbol{\nu}\boldsymbol{\nu}'$. The CU-GMM overidentification test, which implicitly uses the Gaussian pseudo-scores as influence functions, inherits the invariance of the LR test. The advantage, though, is

that we can make it robust to departures from normality, serial independence or conditional homoskedasticity.

5.3 Extension to SMVF spanning tests

Let us turn now to two-point GMM-based spanning tests in the SMVF, first developed by De Santis (1993, 1995) and Bekaert and Urias (1996). In this context, the null of spanning is simply $m^{MV}(c) = m_1^{MV}(c)$ for every c , where $m_1^{MV}(c)$ is the element of the SMVF for \mathbf{R}_1 for which $E[m_1^{MV}(c)] = c$. We can perform a spanning test by choosing two non-negative scalars $c_i \neq c_{ii}$ and adding moment conditions to estimate $\boldsymbol{\nu}_1$, so that

$$E \left\{ \begin{array}{c} \mathbf{R}[c_i + (\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\beta}_{1i}] - \ell_N \\ \mathbf{R}[c_{ii} + (\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\beta}_{1ii}] - \ell_N \\ \mathbf{R}_1 - \boldsymbol{\nu}_1 \end{array} \right\} = \mathbf{0}.$$

Such a system is a simple transformation and reparametrization of (15) since

$$\begin{aligned} & [\mathbf{R}(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^* - \ell_N] - c_i [\mathbf{R}(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^+ - \mathbf{R}] \\ &= \mathbf{R}(\mathbf{R}_1 - \boldsymbol{\nu}_1)' [\boldsymbol{\varphi}_1^* - c_i \boldsymbol{\varphi}_1^+] + \mathbf{R}c - \ell_N = \mathbf{R}[c + (\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\beta}_{1i}] - \ell_N. \end{aligned}$$

As a result, the CU-GMM overidentification test that optimally takes into account the singularity of the above system will be numerically identical in both systems regardless of the values of c_i and c_{ii} chosen by the researcher, while other GMM variants will lead to asymptotically equivalent tests under the null and sequences of local alternatives. Note, though, that there are no finite values of c_i and c_{ii} for which the system above reduces to $\mathbf{f}_C(\mathbf{R}; \boldsymbol{\varphi}_1, \boldsymbol{\nu}_1)$, which reflects that p_1^{++} does not belong to the SMVF. In fact, Figures 1 and 3 illustrate that we can link p_1^{**} to $m_1^{MV}(0)$ and p_1^{++} to an unbounded c in $m_1^{MV}(c)$, in the sense that

$$\lim_{c \rightarrow \pm\infty} E \left[\left(\frac{m_1^{MV}(c)}{c} - (1 + E(p_1^{++}) - p_1^{++}) \right)^2 \right] = 0,$$

so in effect (15) is choosing the limits of the admissible range of expected values of SDF as opposed to two arbitrary interior points.

More recently, Kan and Zhou (2001) discuss an alternative two-point spanning test for the SMVF frontier. Specifically, they suggest to reparametrise $m^{MV}(c)$ in terms of α (see section 2.2) instead of c . Consequently, they rely on the system

$$E \left[\begin{array}{c} \mathbf{R}(\alpha_i + \mathbf{R}'_1 \boldsymbol{\gamma}_{1i}) - \ell_N \\ \mathbf{R}(\alpha_{ii} + \mathbf{R}'_1 \boldsymbol{\gamma}_{1ii}) - \ell_N \end{array} \right] = \mathbf{0},$$

where $\alpha_i \neq \alpha_{ii}$ are two scalars chosen by the researcher.

Such a system is a simple transformation and reparametrization of (12) since

$$\begin{aligned} & [\mathbf{R}\mathbf{R}'_1\phi_1^* - \ell_N] - \alpha_i [\mathbf{R}\mathbf{R}'_1\phi_1^+ - \mathbf{R}] \\ &= \mathbf{R}\mathbf{R}'_1(\phi_1^* - \alpha_i\phi_1^+) + \mathbf{R}\alpha - \ell_N = \mathbf{R}(\alpha + \mathbf{R}'_1\gamma_{1i}) - \ell_N, \end{aligned}$$

and hence the CU-GMM overidentification test will also be numerically identical in both systems regardless of the values of α_i and α_{ii} chosen by the researcher. Once again, though, there are no finite values of α_i and α_{ii} for which the system above reduces to $\mathbf{h}_U(\mathbf{R}; \phi_1)$, which reflects that p_1^+ does not belong to the SMVF either. In fact, Figure 1 illustrates that p_1^+ corresponds to an unbounded α in m_1^{MV} , in the sense that

$$\lim_{\alpha \rightarrow \pm\infty} E \left[\left(\frac{m_1^{MV}}{\alpha} - (1 - p_1^+) \right)^2 \right] = 0.$$

6 Empirical Application

We consider six portfolios of US stocks sorted by size (small and large) and book-to-market (high, medium and low). We will refer to those six portfolios as SL, SM, SH, BL, BM and BH, which define the return vector \mathbf{R} . We want to study if the market portfolio, MK, and the two Fama-French portfolios that capture the size and value effects, SMB (long in small capitalisation stocks and short in big ones) and HML (long in high book-to-market stocks and short in low ones), respectively, span the same return and SDF mean-variance frontiers as the aforementioned six portfolios. Therefore, in our notation, \mathbf{R}_1 will be composed of MK, SMB and HML, while \mathbf{R}_2 is composed of SM, SH and BM.¹³ We use monthly data from 1952 to 2007 (670 observations) from Ken French's web page; see Fama and French (1993) for further details. Hence, our sample starts right after the 1951 Treasury-Fed accord whereby the latter stopped its wartime pegging of interest rates. We systematically work with net returns, (i.e. gross returns minus 1), to avoid numerical problems, and adjust the moment conditions accordingly.

Figure 4 shows the corresponding mean-variance frontiers. To see whether the observed differences are statistically significant, we compute CU-GMM J tests of spanning using 0, 5 and 10 lags in the Newey-West estimator of the long-run covariance matrix of the moment

¹³We make the joint span of \mathbf{R}_1 and \mathbf{R}_2 equivalent to \mathbf{R} by using the mimicking portfolio of MK, which has correlation 0.995 with the actual MK. Our results are not affected by this choice.

conditions.¹⁴ As Proposition 5 shows, for a given lag length the statistics are numerically identical across the different tests studied. Table 1 shows that spanning tests strongly reject the null hypothesis. In turn, Table 1 also shows spanning tests for the same \mathbf{R}_1 when we add the return of each of the six portfolios at a time to shed more light on the sources of the rejection. In this sense, we find strong evidence against the null for all portfolios except SH and BM.

Therefore, we can conclude that a mean-variance investor who is fully invested in US stocks cannot fully reproduce the mean-variance frontiers generated by SL, SM, SH, BL, BM and BH by combining only the two Fama-French portfolios with the market. Furthermore, we can also conclude that there is information about SDF's in those six portfolios that is not fully captured by the US aggregate stock market index and the Fama-French size and value factors.

7 Conclusions

We have provided a unifying approach to test for spanning in the return and stochastic discount factor mean-variance frontiers at three different levels: a) we propose novel testing procedures based on representing portfolios which are not tied down to the properties of either frontier and have a direct economic interpretation, b) we show the asymptotic equivalence of our proposed tests to the extant spanning tests under the null and local alternatives, and c) we prove that by using a single-step GMM procedure such as CU-GMM we can make all the different overidentification tests numerically identical, so one could argue that effectively there is a single spanning test.

The equivalence of the different tests is rather relevant in empirical finance in view of the controversy between regression- and SDF-based asset pricing tests. In a simplified context of *i.i.d* normal returns and factors, Jagannathan and Wang (2002) formally show that the usual beta (regression) and SDF methods lead to asymptotically equivalent risk premia estimators and specification tests. In fact, in a follow up paper (see Peñaranda and Sentana (2010)) we prove the more subtle result that the application to both the regression and SDF approaches of single-step GMM methods, including CU-GMM, gives rise to numerically identical estimates of prices of risk, pricing errors, Jensen's alphas and overidentifying restrictions tests irrespective of the validity of the asset pricing model. Therefore, one could again argue that there is a single method to empirically evaluate asset-pricing models too.

¹⁴ Although we use $\mathcal{S}_T(\boldsymbol{\theta})$ following standard empirical practice, we find qualitatively similar results with $\mathcal{S}_T(\boldsymbol{\theta})$.

Several versions of spanning tests give rise to a non-standard GMM set up in which the influence functions have a singular asymptotic second moment matrix along an implicit manifold in the parameter space that contains the true value. For that reason, we have extended the theory of optimal GMM inference to deal with those non-trivial situations.

In addition, for the benefit of practitioners, we also suggest sensible consistent first-step parameter estimators in spanning tests that can be used either as initial values for single-step GMM procedures, or to obtain consistent estimates of the optimal GMM weighting matrices with potentially better finite sample properties. The choice of first-step estimators is particularly important in our singular GMM set-up to avoid asymptotic discontinuities in the distributions of two-step estimators and tests.

We apply our testing procedures to study if the market portfolio and the two Fama-French portfolios that capture the size and value effect span the same return and SDF mean-variance frontier as six portfolios sorted by size and book-to-market. Given that spanning is clearly rejected, a mean-variance investor who is fully invested in US stocks would be worse off if her choice was constrained to strategies that only combine the size and value mimicking portfolios with the market. Equivalently, our results indicate that those three funds do not fully capture the information about SDF's in the six size and book-to-market sorted portfolios.

There are three situations in which the structure of the RMVF and SMVF implies that spanning will be achieved if the original and expanded frontiers share a single risky portfolio. This will happen when a safe asset is included in \mathbf{R}_1 , only arbitrage portfolios are available, and finally when all expected returns are equal. For the sake of brevity, these three special cases are separately discussed in a companion paper (see Peñaranda and Sentana (2004)). For the same reason, our analysis has not involved moments of order higher than the second, market frictions, or positivity restrictions on the discount factor. The first issue is studied in Snow (1991). Short-sales constraints and transaction costs are dealt with by De Roon, Nijman, and Werker (2000). De Roon, Nijman, and Werker (1997) also considered RMVF spanning under more general expected utility functions (see also Gouriéroux and Monfort (2005)), as well as non-traded assets. In this sense, it would be interesting to develop a utility-based measure of the importance of the additional assets that would complement our spanning tests.

Importantly, spanning tests based on representing portfolios generalise in a natural way to situations in which we want to consider not only a few selected managed portfolios but every

conceivable active (i.e. dynamic) portfolio strategy because the distribution of returns is predictable (see Hansen and Richard (1987), Gallant, Hansen and Tauchen (1990) and Peñaranda and Sentana (2008) for the analysis of mean-variance frontiers with active strategies). The extension of our analysis to the case of conditioning information that can be adequately summarised by means of a multinomial variable is straightforward. Moreover, such an approach can provide the basis for a formal semiparametric procedure that deals with the case of multiple continuous conditioning variables, which we leave for future research (see Chen (2005) for a survey of the recent literature on sieves methods). In such a context, though, we cannot generally apply standard GMM methods because both moments and parameters are infinite dimensional.

Spanning tests are partly related to mutual fund separation. As is well known, the vector \mathbf{R}_1 is separating relative to the investment universe \mathbf{R} if the optimal portfolio of each risk adverse agent can be described as a portfolio of \mathbf{R}_1 alone. In fact, the only additional restriction with respect to the spanning constraints in a regression context is that the residual of the theoretical regression of \mathbf{R}_2 on \mathbf{R}_1 must not only be orthogonal to \mathbf{R}_1 , but also mean independent (see Proposition 2 in Huberman and Kandel (1987), which adapts the original results by Ross (1978) to the spanning case). However, testing for mean independence also involves conditional moment restrictions, which is again qualitatively different from a standard parametric test. Again we leave this issue for further research.

Finally, it is worth mentioning that our singular GMM framework can be applied to other relevant situations, such as mean-variance efficiency tests when the reference portfolio is a linear combination of the available assets. In addition, it can also be applied to dynamic models in which the singularities appear in the long-run covariance matrix but not in the contemporaneous one. Those applications constitute other interesting avenues for further research.

Appendix

A Proofs of Propositions

Proposition 1:

Given the spectral (eigenvalue-eigenvector) decomposition of $\mathbf{S}(\boldsymbol{\theta}^0)$ in (10), all its generalised inverses, which must satisfy $\mathbf{S}\mathbf{S}^{-}\mathbf{S} = \mathbf{S}$, will be of the form

$$\mathbf{S}^{-}(\boldsymbol{\theta}^0) = [\mathbf{P} \quad \mathbf{Q}] \begin{pmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{\Delta}_{PQ} \\ \boldsymbol{\Delta}_{QP} & \boldsymbol{\Delta}_{QQ} \end{pmatrix} \begin{bmatrix} \mathbf{P}' \\ \mathbf{Q}' \end{bmatrix},$$

with $\boldsymbol{\Delta}_{PQ}$, $\boldsymbol{\Delta}_{QP}$ and $\boldsymbol{\Delta}_{QQ}$ arbitrary (see e.g. Rao and Mitra (1971)).

For notational simplicity, but without loss of generality, we shall work with the alternative k parameters $\boldsymbol{\alpha}$ ($s \times 1$) and $\boldsymbol{\beta}$ ($(k - s) \times 1$)

$$(\boldsymbol{\alpha}' \quad \boldsymbol{\beta}') = \mathbf{M}'(\boldsymbol{\theta}),$$

where the first s entries of $\mathbf{M}(\boldsymbol{\theta})$ are such that $\boldsymbol{\alpha} = \mathbf{m}(\boldsymbol{\theta})$. We can always choose $\mathbf{M}(\boldsymbol{\theta})$ to be a regular transformation (i.e. a C^1 -diffeomorphism) on an open neighbourhood of $\boldsymbol{\theta}^0$ in view of Assumptions 1 and 2 so that its inverse is continuously differentiable (see e.g. Fleming (1977, p. 143)).

Let us denote by $\mathbf{l}[\mathbf{M}(\boldsymbol{\theta})] = \boldsymbol{\theta}$ the corresponding inverse transformation, and its Jacobian by

$$\mathbf{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\partial \mathbf{l}(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial (\boldsymbol{\alpha}', \boldsymbol{\beta}')} = [\mathbf{L}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \quad \mathbf{L}_{\boldsymbol{\beta}}(\boldsymbol{\alpha}, \boldsymbol{\beta})].$$

We will impose the parametric restrictions $\mathbf{m}(\boldsymbol{\theta}) = \boldsymbol{\alpha} = \mathbf{0}$ by working with the smaller vector of parameters $\boldsymbol{\beta}$ and the influence functions $\mathbf{h}[\mathbf{x}_t, \mathbf{l}(\mathbf{0}, \boldsymbol{\beta})]$. The optimal GMM estimator is then defined as

$$\hat{\boldsymbol{\beta}}_T = \arg \min_{\boldsymbol{\beta} \in \mathbf{B}} \bar{\mathbf{h}}_T' [\mathbf{l}(\mathbf{0}, \boldsymbol{\beta})] \mathbf{S}^{-}(\boldsymbol{\theta}^0) \bar{\mathbf{h}}_T [\mathbf{l}(\mathbf{0}, \boldsymbol{\beta})],$$

where $\mathbf{B} \subseteq \mathbb{R}^{k-s}$, while the GMM estimator of $\boldsymbol{\theta}$ will be $\hat{\boldsymbol{\theta}}_T = \mathbf{l}(\mathbf{0}, \hat{\boldsymbol{\beta}}_T)$.

a) Given that $[\mathbf{P} \quad \mathbf{Q}]$ is an orthogonal matrix, and $\mathbf{M}(\boldsymbol{\theta})$ is regular in an open neighbourhood of $\boldsymbol{\theta}^0$, so that $\text{rank}[\mathbf{L}(\mathbf{0}, \boldsymbol{\beta})] = k$ by the inverse function theorem, it follows that

$$\text{rank}[\mathbf{D}(\boldsymbol{\theta}^0)] = \text{rank} \begin{bmatrix} \mathbf{P}'\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\alpha}}(\mathbf{0}, \boldsymbol{\beta}) & \mathbf{P}'\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\beta}}(\mathbf{0}, \boldsymbol{\beta}) \\ \mathbf{Q}'\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\alpha}}(\mathbf{0}, \boldsymbol{\beta}) & \mathbf{Q}'\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\beta}}(\mathbf{0}, \boldsymbol{\beta}) \end{bmatrix} = k. \quad (\text{A1})$$

Now, Assumption 1 implies that $\boldsymbol{\Pi} [\mathbf{l}(\mathbf{0}, \boldsymbol{\beta})]' \sqrt{T} \bar{\mathbf{h}}_T [\mathbf{l}(\mathbf{0}, \boldsymbol{\beta})]$ converges in probability to zero for all $\boldsymbol{\beta}$ in the neighbourhood of $\boldsymbol{\beta}^0$. Therefore, if we differentiate this random process with

respect to β , and evaluate the derivatives at β^0 , the continuously differentiable character of $\Pi(\theta)$, $\mathbf{h}_t(\theta)$ and $\mathbf{l}(\mathbf{0}, \beta)$ at the true value imply that

$$\{\sqrt{T}\bar{\mathbf{h}}_T[\mathbf{l}(\mathbf{0}, \beta^0)] \otimes \mathbf{I}_s\} \frac{\partial \text{vec}\{\Pi'[\mathbf{l}(\mathbf{0}, \beta^0)]\}}{\partial \beta} + \Pi'[\mathbf{l}(\mathbf{0}, \beta^0)]\sqrt{T} \frac{\partial \bar{\mathbf{h}}_T[\mathbf{l}(\mathbf{0}, \beta^0)]}{\partial \beta} \xrightarrow{p} \mathbf{0}$$

by the continuous mapping theorem, which a fortiori implies that

$$\{\bar{\mathbf{h}}_T[\mathbf{l}(\mathbf{0}, \beta^0)] \otimes \mathbf{I}_s\} \frac{\partial \text{vec}\{\Pi'[\mathbf{l}(\mathbf{0}, \beta^0)]\}}{\partial \beta} + \Pi'[\mathbf{l}(\mathbf{0}, \beta^0)] \frac{\partial \bar{\mathbf{h}}_T[\mathbf{l}(\mathbf{0}, \beta^0)]}{\partial \beta} \xrightarrow{p} \mathbf{0}$$

since $T^{-1/2}$ is trivially $o(1)$. But since the chain rule for first derivatives allows us to write the previous expression as

$$\{\bar{\mathbf{h}}_T[\mathbf{l}(\mathbf{0}, \beta^0)] \otimes \mathbf{I}_s\} \frac{\partial \text{vec}\{\Pi'[\mathbf{l}(\mathbf{0}, \beta^0)]\}}{\partial \theta'} \mathbf{L}_\beta(\mathbf{0}, \beta^0) + \Pi'[\mathbf{l}(\mathbf{0}, \beta^0)] \bar{\mathbf{D}}_T[\mathbf{l}(\mathbf{0}, \beta^0)] \mathbf{L}_\beta(\mathbf{0}, \beta^0),$$

where $\bar{\mathbf{D}}_T(\theta) = \partial \bar{\mathbf{h}}_T(\theta) / \partial \theta'$, our assumptions imply that

$$\Pi'[\mathbf{l}(\mathbf{0}, \beta^0)] \mathbf{D}[\mathbf{l}(\mathbf{0}, \beta^0)] \mathbf{L}_\beta(\mathbf{0}, \beta^0) = \mathbf{0}$$

in view of the fact that $\bar{\mathbf{D}}_T[\mathbf{l}(\mathbf{0}, \beta^0)] \xrightarrow{p} \mathbf{D}[\mathbf{l}(\mathbf{0}, \beta^0)]$ and $\bar{\mathbf{h}}_T[\mathbf{l}(\mathbf{0}, \beta^0)] \xrightarrow{p} E[\mathbf{h}(\mathbf{x}_t; \theta^0)] = \mathbf{0}$ by definition of θ^0 .

Finally, given that \mathbf{Q} must be a full-column rank linear transformation of $\Pi(\theta^0)$ because $\mathbf{Q}'\mathbf{S}(\theta^0) = \mathbf{0}$, we can also show that

$$\mathbf{Q}'\mathbf{D}(\theta^0)\mathbf{L}_\beta(\mathbf{0}, \beta^0) = \mathbf{0}. \quad (\text{A2})$$

As a result, $\text{rank}[\mathbf{P}'\mathbf{D}(\theta^0)\mathbf{L}_\beta(\mathbf{0}, \beta^0)]$ must indeed be $k - s$ for (A1) to be true. Therefore, we have shown that after imposing the restriction $\alpha = \mathbf{0}$, the reduced moment conditions $\mathbf{P}'\bar{\mathbf{h}}_T(\mathbf{l}(\mathbf{0}, \beta))$ will first-order identify β at β^0 .

b) If $\mathbf{h}(\mathbf{x}_t; \theta)$ satisfies the regularity conditions mentioned in the text, together with Assumptions 1 and 2, then we can easily prove that those regularity conditions will also be satisfied by $\mathbf{P}'\mathbf{h}[\mathbf{x}_t; \mathbf{l}(\mathbf{0}, \beta)]$ because the latter functions are a linear combination of the former, and the transformation from θ to (α, β) is regular over an open neighbourhood of θ^0 . This fact, together with property a), allows us to derive the asymptotic distribution of the infeasible optimal GMM estimator of the transformed parameters β . Specifically, in large samples

$$\begin{aligned} & \sqrt{T}(\hat{\beta}_T - \beta^0) \\ &= - [\mathbf{L}'_\beta(\mathbf{0}, \beta^0)\mathbf{D}'(\theta^0)\mathbf{S}^-(\theta^0)\mathbf{D}(\theta^0)\mathbf{L}_\beta(\mathbf{0}, \beta^0)]^{-1} \mathbf{L}'_\beta(\mathbf{0}, \beta^0)\mathbf{D}'(\theta^0)\mathbf{S}^-(\theta^0)\sqrt{T}\bar{\mathbf{h}}_T[\mathbf{l}(\mathbf{0}, \beta^0)] + o_p(1) \\ &= - [\mathbf{L}'_\beta(\mathbf{0}, \beta^0)\mathbf{D}'(\theta^0)\mathbf{S}^+(\theta^0)\mathbf{D}(\theta^0)\mathbf{L}_\beta(\mathbf{0}, \beta^0)]^{-1} \mathbf{L}'_\beta(\mathbf{0}, \beta^0)\mathbf{D}'(\theta^0)\mathbf{S}^+(\theta^0)\sqrt{T}\bar{\mathbf{h}}_T[\mathbf{l}(\mathbf{0}, \beta^0)] + o_p(1), \end{aligned}$$

for any $\mathbf{S}^-(\boldsymbol{\theta}^0)$, because

$$\begin{aligned} & \mathbf{L}'_{\beta}(\mathbf{0}, \beta^0) \mathbf{D}'(\boldsymbol{\theta}^0) \mathbf{S}^-(\boldsymbol{\theta}^0) \mathbf{D}(\boldsymbol{\theta}^0) \mathbf{L}_{\beta}(\mathbf{0}, \beta^0) \\ &= [\mathbf{L}'_{\beta}(\mathbf{0}, \beta^0) \mathbf{D}'(\boldsymbol{\theta}^0) \mathbf{P} \quad \mathbf{0}] \begin{pmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{\Delta}_{PQ} \\ \boldsymbol{\Delta}_{QP} & \boldsymbol{\Delta}_{QQ} \end{pmatrix} \begin{bmatrix} \mathbf{P}' \mathbf{D}(\boldsymbol{\theta}^0) \mathbf{L}_{\beta}(\mathbf{0}, \beta^0) \\ \mathbf{0} \end{bmatrix} \\ &= \mathbf{L}'_{\beta}(\mathbf{0}, \beta^0) \mathbf{D}'(\boldsymbol{\theta}^0) \mathbf{P} \boldsymbol{\Lambda}^{-1} \mathbf{P}' \mathbf{D}(\boldsymbol{\theta}^0) \mathbf{L}_{\beta}(\mathbf{0}, \beta^0) = \mathbf{L}'_{\beta}(\mathbf{0}, \beta^0) \mathbf{D}'(\boldsymbol{\theta}^0) \mathbf{S}^+(\boldsymbol{\theta}^0) \mathbf{D}(\boldsymbol{\theta}^0) \mathbf{L}_{\beta}(\mathbf{0}, \beta^0), \end{aligned}$$

in view of Assumptions 1 and 2, and similarly

$$\begin{aligned} & \mathbf{L}'_{\beta}(\mathbf{0}, \beta^0) \mathbf{D}'(\boldsymbol{\theta}^0) \mathbf{S}^-(\boldsymbol{\theta}^0) \bar{\mathbf{h}}_T [\mathbf{1}(\mathbf{0}, \beta^0)] \\ &= [\mathbf{L}'_{\beta}(\mathbf{0}, \beta^0) \mathbf{D}'(\boldsymbol{\theta}^0) \mathbf{P} \quad \mathbf{0}] \begin{pmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{\Delta}_{PQ} \\ \boldsymbol{\Delta}_{QP} & \boldsymbol{\Delta}_{QQ} \end{pmatrix} \begin{bmatrix} \mathbf{P}' \bar{\mathbf{h}}_T [\mathbf{1}(\mathbf{0}, \beta^0)] \\ \mathbf{0} \end{bmatrix} \\ &= \mathbf{L}'_{\beta}(\mathbf{0}, \beta^0) \mathbf{D}'(\boldsymbol{\theta}^0) \mathbf{P} \boldsymbol{\Lambda}^{-1} \mathbf{P}' \bar{\mathbf{h}}_T [\mathbf{1}(\mathbf{0}, \beta^0)] = \mathbf{L}'_{\beta}(\mathbf{0}, \beta^0) \mathbf{D}'(\boldsymbol{\theta}^0) \mathbf{S}^+(\boldsymbol{\theta}^0) \bar{\mathbf{h}}_T [\mathbf{1}(\mathbf{0}, \beta^0)]. \end{aligned}$$

Therefore,

$$\begin{aligned} & \sqrt{T}(\hat{\boldsymbol{\beta}}_T - \beta^0) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}), \\ & \mathbf{V} = [\mathbf{L}'_{\beta}(\mathbf{0}, \beta^0) \mathbf{D}(\boldsymbol{\theta}^0)' \mathbf{S}^+(\boldsymbol{\theta}^0) \mathbf{D}(\boldsymbol{\theta}^0) \mathbf{L}_{\beta}(\mathbf{0}, \beta^0)]^{-1}. \end{aligned}$$

We can use the standard delta method to show that the optimal GMM estimators of the parameters of interest, $\boldsymbol{\theta}$, which will be given by $\hat{\boldsymbol{\theta}}_T = \mathbf{1}(\mathbf{0}, \hat{\boldsymbol{\beta}}_T)$, will have the asymptotically normal distribution

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^0) \xrightarrow{d} N[\mathbf{0}, \mathbf{L}_{\beta}(\mathbf{0}, \beta^0) \mathbf{V} \mathbf{L}'_{\beta}(\mathbf{0}, \beta^0)]$$

whose covariance matrix is of rank $k - s$.

Similarly, the asymptotic covariance matrix of a GMM estimator that uses $\boldsymbol{\Upsilon}$ as weighting matrix but does not impose $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$ is

$$[\mathbf{D}'(\boldsymbol{\theta}^0) \boldsymbol{\Upsilon} \mathbf{D}(\boldsymbol{\theta}^0)]^{-1} \cdot [\mathbf{D}'(\boldsymbol{\theta}^0) \boldsymbol{\Upsilon} \mathbf{S}(\boldsymbol{\theta}^0) \boldsymbol{\Upsilon} \mathbf{D}(\boldsymbol{\theta}^0)] \cdot [\mathbf{D}'(\boldsymbol{\theta}^0) \boldsymbol{\Upsilon} \mathbf{D}(\boldsymbol{\theta}^0)]^{-1}.$$

Our claimed optimality of $\hat{\boldsymbol{\theta}}_T = \mathbf{1}(\mathbf{0}, \hat{\boldsymbol{\beta}}_T)$ depends on

$$[\mathbf{D}'(\boldsymbol{\theta}^0) \boldsymbol{\Upsilon} \mathbf{D}(\boldsymbol{\theta}^0)]^{-1} \cdot [\mathbf{D}'(\boldsymbol{\theta}^0) \boldsymbol{\Upsilon} \mathbf{S}(\boldsymbol{\theta}^0) \boldsymbol{\Upsilon} \mathbf{D}(\boldsymbol{\theta}^0)] \cdot [\mathbf{D}'(\boldsymbol{\theta}^0) \boldsymbol{\Upsilon} \mathbf{D}(\boldsymbol{\theta}^0)]^{-1} - \mathbf{L}_{\beta}(\mathbf{0}, \beta^0) \mathbf{V} \mathbf{L}'_{\beta}(\mathbf{0}, \beta^0)$$

being positive semidefinite, which in turn requires that

$$\mathbf{D}'(\boldsymbol{\theta}^0) \boldsymbol{\Upsilon} \mathbf{S}(\boldsymbol{\theta}^0) \boldsymbol{\Upsilon} \mathbf{D}(\boldsymbol{\theta}^0) - [\mathbf{D}'(\boldsymbol{\theta}^0) \boldsymbol{\Upsilon} \mathbf{D}(\boldsymbol{\theta}^0)] \mathbf{L}_{\beta}(\mathbf{0}, \beta^0) \mathbf{V} \mathbf{L}'_{\beta}(\mathbf{0}, \beta^0) [\mathbf{D}'(\boldsymbol{\theta}^0) \boldsymbol{\Upsilon} \mathbf{D}(\boldsymbol{\theta}^0)]$$

is positive semidefinite.

The result follows since this matrix is the asymptotic residual variance in the limiting least squares projection of $\sqrt{T}\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\bar{\mathbf{h}}_T(\boldsymbol{\theta}^0)$ on $\sqrt{T}\mathbf{L}'_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{D}'(\boldsymbol{\theta}^0)\mathbf{S}^{-}(\boldsymbol{\theta}^0)\bar{\mathbf{h}}_T(\boldsymbol{\theta}^0)$. Regardless of our generalised inverse choice,

$$\begin{aligned} & \lim_{T \rightarrow \infty} Var \left[\begin{array}{c} \sqrt{T}\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\bar{\mathbf{h}}_T(\boldsymbol{\theta}^0) \\ \sqrt{T}\mathbf{L}'_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{D}'(\boldsymbol{\theta}^0)\mathbf{S}^{-}(\boldsymbol{\theta}^0)\bar{\mathbf{h}}_T(\boldsymbol{\theta}^0) \end{array} \right] = \\ & = \left[\begin{array}{cc} \mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{S}(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0) & \mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0) \\ \mathbf{L}'_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0) & \mathbf{V}^{-1} \end{array} \right], \end{aligned}$$

where the only new object is the covariance term. The upper right covariance term follows from

$$\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{S}(\boldsymbol{\theta}^0)\mathbf{S}^{-}(\boldsymbol{\theta}^0)\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0) = \mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{P}\mathbf{P}'\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)$$

because $\mathbf{S}(\boldsymbol{\theta}^0)\mathbf{S}^{-}(\boldsymbol{\theta}^0)\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0) = \mathbf{P}\mathbf{P}'\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)$ in view of Assumptions 1 and 2, and

$$\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{P}\mathbf{P}'\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0) = \mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}[\mathbf{I}_n - \mathbf{Q}\mathbf{Q}']\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0) = \mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)$$

since the eigenvector matrix is orthogonal.

Finally, we can develop similar arguments to prove the efficiency of $\hat{\boldsymbol{\theta}}_T = \mathbf{1}(\mathbf{0},\hat{\boldsymbol{\beta}}_T)$ with respect to a third estimator that uses $\boldsymbol{\Upsilon}$ as weighting matrix but imposes $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$, and a fourth estimator that uses a generalised inverse $\mathbf{S}^{-}(\boldsymbol{\theta}^0)$ as weighting matrix but does not impose $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$. The asymptotic variance of the third estimator is

$$\begin{aligned} & [\mathbf{L}'_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)]^{-1} \cdot [\mathbf{L}'_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{S}(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)] \cdot \\ & [\mathbf{L}'_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)]^{-1}, \end{aligned}$$

while the asymptotic variance of the fourth estimator is

$$[\mathbf{D}'(\boldsymbol{\theta}^0)\mathbf{S}^{-}(\boldsymbol{\theta}^0)\mathbf{D}(\boldsymbol{\theta}^0)]^{-1} \mathbf{D}'(\boldsymbol{\theta}^0)\mathbf{S}^{-}(\boldsymbol{\theta}^0)\mathbf{S}(\boldsymbol{\theta}^0)\mathbf{S}^{-}(\boldsymbol{\theta}^0)\mathbf{D}(\boldsymbol{\theta}^0) [\mathbf{D}'(\boldsymbol{\theta}^0)\mathbf{S}^{-}(\boldsymbol{\theta}^0)\mathbf{D}(\boldsymbol{\theta}^0)]^{-1},$$

assuming $\mathbf{D}'(\boldsymbol{\theta}^0)\mathbf{S}^{-}(\boldsymbol{\theta}^0)\mathbf{D}(\boldsymbol{\theta}^0)$ has full rank (otherwise it will not even identify $\boldsymbol{\theta}$).

Straightforward algebra shows that any of those matrices minus $\mathbf{L}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{V}\mathbf{L}'_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)$ yields a positive semidefinite matrix.

c) Our previous results imply that in large samples

$$\begin{aligned} & \sqrt{T}\bar{\mathbf{h}}_T \left[\mathbf{1}(\mathbf{0},\hat{\boldsymbol{\beta}}_T) \right] = \sqrt{T}\bar{\mathbf{h}}_T(\boldsymbol{\theta}^0) + \mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)\sqrt{T} \left(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}^0 \right) + o_p(1) \\ & = \left\{ \mathbf{I}_n - \mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{V}\mathbf{L}'_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{D}'(\boldsymbol{\theta}^0)\mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}' \right\} \sqrt{T}\bar{\mathbf{h}}_T(\boldsymbol{\theta}^0) + o_p(1). \end{aligned}$$

If we define $\mathcal{D} = \mathbf{\Lambda}^{-1/2} \mathbf{P}' \mathbf{D}(\boldsymbol{\theta}^0) \mathbf{L}_\beta(\mathbf{0}, \boldsymbol{\beta}^0)$ then we can write the previous expression as

$$\begin{aligned} \sqrt{T} \bar{\mathbf{h}}_T \left[\mathbf{1}(\mathbf{0}, \hat{\boldsymbol{\beta}}_T) \right] &= \left\{ \mathbf{I}_n - \mathbf{D}(\boldsymbol{\theta}^0) \mathbf{L}_\beta(\mathbf{0}, \boldsymbol{\beta}^0) [\mathcal{D}' \mathcal{D}]^{-1} \mathcal{D}' \mathbf{\Lambda}^{-1/2} \mathbf{P}' \right\} \sqrt{T} \bar{\mathbf{h}}_T(\boldsymbol{\theta}^0) + o_p(1) \\ &= \left\{ \mathbf{P} \mathbf{\Lambda}^{1/2} - \mathbf{D}(\boldsymbol{\theta}^0) \mathbf{L}_\beta(\mathbf{0}, \boldsymbol{\beta}^0) [\mathcal{D}' \mathcal{D}]^{-1} \mathcal{D}' \right\} \sqrt{T} \mathbf{\Lambda}^{-1/2} \mathbf{P}' \bar{\mathbf{h}}_T(\boldsymbol{\theta}^0) + o_p(1) \\ &= \mathbf{P} \mathbf{\Lambda}^{1/2} \left\{ \mathbf{I}_{n-s} - \mathcal{D} [\mathcal{D}' \mathcal{D}]^{-1} \mathcal{D}' \right\} \sqrt{T} \mathbf{\Lambda}^{-1/2} \mathbf{P}' \bar{\mathbf{h}}_T(\boldsymbol{\theta}^0) + o_p(1). \end{aligned}$$

Therefore, in large samples, the criterion function evaluated at the estimator and scaled by T is

$$T \bar{\mathbf{h}}_T \left[\mathbf{1}(\mathbf{0}, \hat{\boldsymbol{\beta}}_T) \right] \mathbf{S}^-(\boldsymbol{\theta}^0) \bar{\mathbf{h}}_T \left[\mathbf{1}(\mathbf{0}, \hat{\boldsymbol{\beta}}_T) \right] = \mathbf{z}'_T \left\{ \mathbf{I}_{n-s} - \mathcal{D} [\mathcal{D}' \mathcal{D}]^{-1} \mathcal{D}' \right\} \mathbf{z}_T + o_p(1),$$

where $\mathbf{z}_T = \sqrt{T} \mathbf{\Lambda}^{-1/2} \mathbf{P}' \bar{\mathbf{h}}_T(\boldsymbol{\theta}^0)$ is asymptotically distributed as a standard multivariate normal. Given that the matrix $\left\{ \mathbf{I}_n - \mathcal{D} [\mathcal{D}' \mathcal{D}]^{-1} \mathcal{D}' \right\}$ is idempotent with rank $(n-s) - (k-s) = n-k$, we finally obtain an asymptotic chi-square limiting distribution with $n-k$ degrees of freedom.

d) Given that we impose the equality restriction $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$, we will find that

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \bar{\boldsymbol{\theta}}_T) = o_p(1)$$

if and only if the same relationship holds for the underlying non-singular estimator, i.e. if

$$\sqrt{T}\{\hat{\boldsymbol{\beta}}_T - \bar{\boldsymbol{\beta}}_T\} = o_p(1).$$

But since we have previously found that

$$\begin{aligned} &\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}^0) \\ &= -[\mathbf{L}'_\beta(\mathbf{0}, \boldsymbol{\beta}^0) \mathbf{D}'(\boldsymbol{\theta}^0) \mathbf{S}^+(\boldsymbol{\theta}^0) \mathbf{D}(\boldsymbol{\theta}^0) \mathbf{L}_\beta(\mathbf{0}, \boldsymbol{\beta}^0)]^{-1} \mathbf{L}'_\beta(\mathbf{0}, \boldsymbol{\beta}^0) \mathbf{D}'(\boldsymbol{\theta}^0) \mathbf{S}^+(\boldsymbol{\theta}^0) \sqrt{T} \bar{\mathbf{h}}_T[\mathbf{1}(\mathbf{0}, \boldsymbol{\beta}^0)] + o_p(1), \end{aligned}$$

regardless of the choice of generalised inverse $\mathbf{S}^-(\boldsymbol{\theta}^0)$, the result follows. \square

Proposition 2:

For the sake of brevity, we only prove the CU-GMM result, which in turn implies the GMM result using standard arguments. The system

$$E \begin{pmatrix} \mathbf{R} \mathbf{R}'_1 \phi_1^+ - \boldsymbol{\nu} \\ \mathbf{R} \mathbf{R}'_1 \phi_1^* - \ell_N \\ \mathbf{R} - \boldsymbol{\nu} \end{pmatrix}$$

is numerically equivalent to the system

$$E \begin{pmatrix} \mathbf{R} \mathbf{R}'_1 \phi_1^+ - \mathbf{R} \\ \mathbf{R} \mathbf{R}'_1 \phi_1^* - \ell_N \\ \mathbf{R} - \boldsymbol{\nu} \end{pmatrix} = \begin{pmatrix} \mathbf{h}_U(\mathbf{R}; \phi_1) \\ \mathbf{h}_M(\mathbf{R}; \boldsymbol{\nu}) \end{pmatrix},$$

because we are simply adding $\mathbf{h}_M(\mathbf{R}; \boldsymbol{\nu})$ to the mean representing portfolio block. If we define $\boldsymbol{\theta} = (\boldsymbol{\phi}_1, \boldsymbol{\nu})$, and drop the T subscripts from the HAC estimator $\mathbf{S}_T(\boldsymbol{\theta})$ and its components, we can decompose the joint long-run variance variance of the larger system as

$$\mathbf{S}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{S}_{UU}(\boldsymbol{\phi}_1) & \mathbf{S}'_{MU}(\boldsymbol{\theta}) \\ \mathbf{S}_{MU}(\boldsymbol{\theta}) & \mathbf{S}_{MM}(\boldsymbol{\nu}) \end{bmatrix} = \begin{pmatrix} \mathbf{I}_{2N} & \mathbf{0} \\ \mathbf{B}_{MU}(\boldsymbol{\theta}) & \mathbf{I}_N \end{pmatrix} \begin{bmatrix} \mathbf{S}_{UU}(\boldsymbol{\phi}_1) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_{MM}(\boldsymbol{\theta}) \end{bmatrix} \begin{pmatrix} \mathbf{I}_{2N} & \mathbf{B}'_{MU}(\boldsymbol{\theta}) \\ \mathbf{0} & \mathbf{I}_N \end{pmatrix},$$

where

$$\mathbf{B}_{MU}(\boldsymbol{\theta}) = \mathbf{S}_{MU}(\boldsymbol{\theta}) \mathbf{S}_{UU}^{-1}(\boldsymbol{\phi}_1), \quad \boldsymbol{\Omega}_{MM}(\boldsymbol{\theta}) = \mathbf{S}_{MM}(\boldsymbol{\nu}) - \mathbf{S}_{MU}(\boldsymbol{\theta}) \mathbf{S}_{UU}^{-1}(\boldsymbol{\phi}_1) \mathbf{S}'_{MU}(\boldsymbol{\theta}).$$

On this basis, we can transform $\mathbf{h}_M(\mathbf{R}; \boldsymbol{\nu})$ into

$$\mathbf{g}_M(\mathbf{R}; \boldsymbol{\theta}) = \mathbf{h}_M(\mathbf{R}; \boldsymbol{\nu}) - \mathbf{B}_{MU}(\boldsymbol{\theta}) \mathbf{h}_U(\mathbf{R}; \boldsymbol{\phi}_1),$$

which is such that the long-run covariance matrix of $(\mathbf{h}'_U(\mathbf{R}; \boldsymbol{\phi}_1), \mathbf{g}'_M(\mathbf{R}; \boldsymbol{\theta}))$

$$\begin{bmatrix} \mathbf{S}_{UU}(\boldsymbol{\phi}_1) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_{MM}(\boldsymbol{\theta}) \end{bmatrix}$$

is block diagonal, with the singularity confined to the first N_1 components of $\boldsymbol{\Omega}_{MM}(\boldsymbol{\theta})$. The advantage of a CU-GMM set-up is that the moment conditions $E[\mathbf{h}'_U(\mathbf{R}; \boldsymbol{\phi}_1), \mathbf{h}'_M(\mathbf{R}; \boldsymbol{\nu})] = \mathbf{0}$ and $E[\mathbf{h}'_U(\mathbf{R}; \boldsymbol{\phi}_1), \mathbf{g}'_M(\mathbf{R}; \boldsymbol{\theta})] = \mathbf{0}$ are numerically equivalent.

Focusing on the Moore-Penrose inverse without loss of generality, the CU-GMM criterion with the transformed moment conditions can be additively decomposed into two components, namely

$$\begin{aligned} & \begin{pmatrix} \bar{\mathbf{h}}_{UT}(\boldsymbol{\phi}_1) \\ \bar{\mathbf{g}}_{MT}(\boldsymbol{\theta}) \end{pmatrix}' \begin{bmatrix} \mathbf{S}_{UU}(\boldsymbol{\phi}_1) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_{MM}(\boldsymbol{\theta}) \end{bmatrix}^+ \begin{pmatrix} \bar{\mathbf{h}}_{UT}(\boldsymbol{\phi}_1) \\ \bar{\mathbf{g}}_{MT}(\boldsymbol{\theta}) \end{pmatrix} \\ &= \bar{\mathbf{h}}_{UT}(\boldsymbol{\phi}_1)' \mathbf{S}_{UU}^{-1}(\boldsymbol{\phi}_1) \bar{\mathbf{h}}_{UT}(\boldsymbol{\phi}_1) + \bar{\mathbf{g}}_{MT}(\boldsymbol{\theta})' \boldsymbol{\Omega}_{MM}^+(\boldsymbol{\theta}) \bar{\mathbf{g}}_{MT}(\boldsymbol{\theta}). \end{aligned}$$

The first component only depends on $\boldsymbol{\phi}_1$, and coincides with the optimal criterion that simply relies on the moment conditions $E[\mathbf{h}_U(\mathbf{R}; \boldsymbol{\phi}_1)] = \mathbf{0}$. Hence both the joint criterion and the simple criterion $\bar{\mathbf{h}}_{UT}(\boldsymbol{\phi}_1)' \mathbf{S}_{UU}^{-1}(\boldsymbol{\phi}_1) \bar{\mathbf{h}}_{UT}(\boldsymbol{\phi}_1)$ will provide the same estimator of $\boldsymbol{\phi}_1$ if we can make the second component $\bar{\mathbf{g}}_{MT}(\boldsymbol{\theta})' \boldsymbol{\Omega}_{MM}^+(\boldsymbol{\theta}) \bar{\mathbf{g}}_{MT}(\boldsymbol{\theta})$ equal to 0 for any $\boldsymbol{\phi}_1$ by a suitable choice of $\boldsymbol{\nu}$.

Let us show that this is indeed the case. For any given value of ϕ_1 , the singularity of $\bar{\mathbf{g}}_{MT}(\boldsymbol{\nu}, \phi_1)$ can be handled by computing the Moore-Penrose inverse of $\boldsymbol{\Omega}_{MM}(\boldsymbol{\theta})$, which effectively deletes one of the moments in $\bar{\mathbf{g}}_{MT}(\boldsymbol{\nu}, \phi_1)$, and at the same time imposing the manifold $\phi_1^{*\prime} \boldsymbol{\nu}_1 - \phi_1^{+\prime} \ell_{N_1} = 0$ on $\boldsymbol{\nu}_1$. Therefore the second component is exactly identified for $\boldsymbol{\nu}$ given ϕ_1 and can thus be set to zero for any ϕ_1 .

In fact, it is not difficult to fully characterise the optimal estimator $\boldsymbol{\nu}(\phi_1)$, which is not generally the value of $\boldsymbol{\nu}$ that sets $\bar{\mathbf{g}}_{MT}(\boldsymbol{\nu}, \phi_1)$ equal to zero because such a value does not necessarily satisfy the manifold. In fact, the optimal estimator will be the value of $\boldsymbol{\nu}$ that makes $\bar{\mathbf{g}}_{MT}(\boldsymbol{\nu}, \phi_1)$ proportional to the eigenvector of $\boldsymbol{\Omega}_{MM}(\boldsymbol{\theta})$ associated to its 0 eigenvalue, with a factor of proportionality such that the manifold is satisfied.

Let $\mathbf{q}(\boldsymbol{\theta})$ denotes the aforementioned eigenvector, which is equal to a vector whose first N_1 entries are proportional to the value of ϕ_1^* , and the rest N_2 entries are equal to zero. Then, the optimal choice of $\boldsymbol{\nu}(\phi_1)$ is such that $\bar{\mathbf{g}}_{MT}[\boldsymbol{\nu}(\phi_1), \phi_1] = k(\boldsymbol{\theta}) \mathbf{q}(\boldsymbol{\theta})$. In large samples, we can express the implicit solution as

$$\begin{bmatrix} \bar{\mathbf{R}}_{1T} - \boldsymbol{\nu}_1(\phi_1) \\ \bar{\mathbf{R}}_{2T} - \boldsymbol{\nu}_2(\phi_1) \end{bmatrix} - \begin{bmatrix} \mathbf{B}_{MU1}(\boldsymbol{\theta}) \\ \mathbf{B}_{MU2}(\boldsymbol{\theta}) \end{bmatrix} \bar{\mathbf{h}}_{UT}(\phi_1, \mathbf{0}) = k(\boldsymbol{\theta}) \begin{pmatrix} \phi_1^* \\ \mathbf{0} \end{pmatrix},$$

or

$$\boldsymbol{\nu}_1(\phi_1) = \bar{\mathbf{R}}_{1T} - \mathbf{B}_{MU1}(\boldsymbol{\theta}) \bar{\mathbf{h}}_{UT}(\phi_1) - k(\boldsymbol{\theta}) \phi_1^*, \quad \boldsymbol{\nu}_2(\phi_1) = \bar{\mathbf{R}}_{2T} - \mathbf{B}_{MU2}(\boldsymbol{\theta}) \bar{\mathbf{h}}_{UT}(\phi_1),$$

where $\boldsymbol{\theta} = [\phi_1, \boldsymbol{\nu}(\phi_1)]$ and $k(\boldsymbol{\theta})$ is implicitly defined by the manifold $\phi_1^{*\prime} \boldsymbol{\nu}_1 - \phi_1^{+\prime} \ell_{N_1} = 0$. \square

Proposition 3:

Once again, we prove the CU-GMM numerical result, which in turn implies the GMM asymptotic result. Let us study the original system

$$\begin{bmatrix} (\mathbf{R} - \boldsymbol{\nu})(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^+ - \boldsymbol{\nu} \\ (\mathbf{R} - \boldsymbol{\nu})(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^* - \ell_N \\ \mathbf{R} - \boldsymbol{\nu} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_N & \mathbf{0} & -\boldsymbol{\nu}(\boldsymbol{\varphi}_1^{+\prime}, \mathbf{0}'_{N_2}) - \mathbf{I}_N \\ \mathbf{0} & \mathbf{I}_N & -\boldsymbol{\nu}(\boldsymbol{\varphi}_1^{*\prime}, \mathbf{0}'_{N_2}) \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_N \end{bmatrix} \begin{bmatrix} \mathbf{R}(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^+ - \mathbf{R} \\ \mathbf{R}(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^* - \ell_N \\ \mathbf{R} - \boldsymbol{\nu} \end{bmatrix}.$$

In a CU-GMM set-up, we know that both systems are equivalent for the estimation of $\boldsymbol{\varphi}_1$ and $\boldsymbol{\nu}$. In addition, we can partition the latter system as

$$\begin{bmatrix} \mathbf{R}(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^+ - \mathbf{R} \\ \mathbf{R}(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^* - \ell_N \\ \mathbf{R}_1 - \boldsymbol{\nu}_1 \\ \mathbf{R}_2 - \boldsymbol{\nu}_2 \end{bmatrix},$$

which leaves $\boldsymbol{\nu}_2$ as an additional nuisance parameter that is exactly identified.

If we orthogonalise $\mathbf{R}_2 - \boldsymbol{\nu}_2$ with respect to the rest of influence functions then we can additively decompose the criterion function into two components. One component is the criterion function based on the sample average of

$$\mathbf{f}_C(\mathbf{R}; \boldsymbol{\varphi}_1, \boldsymbol{\nu}_1) = \begin{bmatrix} \mathbf{R}(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^+ - \mathbf{R} \\ \mathbf{R}(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^* - \ell_N \\ \mathbf{R}_1 - \boldsymbol{\nu}_1 \end{bmatrix},$$

where we impose the manifold $\bar{m}_C(\boldsymbol{\varphi}_1, \boldsymbol{\nu}_1) = 0$ and use a generalised inverse, while the other component is a criterion function based on the sample moments

$$(\bar{\mathbf{R}}_{2T} - \boldsymbol{\nu}_2) - \mathbf{B}_{MC}(\boldsymbol{\varphi}_1, \boldsymbol{\nu}_1, \boldsymbol{\nu}_2) \bar{\mathbf{f}}_{C,T}(\boldsymbol{\varphi}_1, \boldsymbol{\nu}_1),$$

where $\mathbf{B}_{MC}(\cdot)$ denotes the projection slopes from the corresponding HAC estimator $\mathbf{S}_T(\boldsymbol{\theta})$, and where we have again dropped the T subscripts to simplify the notation. These moments exactly identify $\boldsymbol{\nu}_2$ given $(\boldsymbol{\varphi}_1, \boldsymbol{\nu}_1)$, and hence for any value of $(\boldsymbol{\varphi}_1, \boldsymbol{\nu}_1)$ we can make the second component of the criterion function equal to zero by solving the implicit equation

$$\boldsymbol{\nu}_2(\boldsymbol{\varphi}_1, \boldsymbol{\nu}_1) = \bar{\mathbf{R}}_{2T} - \mathbf{B}_{MC}(\boldsymbol{\varphi}_1, \boldsymbol{\nu}_1, \boldsymbol{\nu}_2(\boldsymbol{\varphi}_1, \boldsymbol{\nu}_1)) \bar{\mathbf{f}}_{C,T}(\boldsymbol{\varphi}_1, \boldsymbol{\nu}_1).$$

Therefore this second component is irrelevant in large enough samples for the estimation of $(\boldsymbol{\varphi}_1, \boldsymbol{\nu}_1)$ and the corresponding J test, so effectively they are both relying on $\mathbf{f}_C(\mathbf{R}; \boldsymbol{\varphi}_1, \boldsymbol{\nu}_1)$ only. Finally, note that the optimal estimator of $\boldsymbol{\nu}_2$ will not be simply $\bar{\mathbf{R}}_{2T}$ in general, but rather the value of $\boldsymbol{\nu}_2$ that sets to 0 the orthogonalised condition above. \square

Proposition 4:

As before, we prove the numerical CU-GMM result, which in turn implies the asymptotic GMM result. Once we impose the manifold of Lemma 5, we can write $\mathbf{h}_L(\mathbf{R}; \mathbf{b}_i, \mathbf{b}_{ii})$ as

$$\begin{pmatrix} \boldsymbol{\Phi}_i \otimes \mathbf{I}_{N_2} \\ \boldsymbol{\Phi}_{ii} \otimes \mathbf{I}_{N_2} \end{pmatrix} \mathbf{h}_H(\mathbf{R}; \mathbf{b}_1) = \left[\begin{pmatrix} \boldsymbol{\Phi}_i & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_{ii} \end{pmatrix} \otimes \mathbf{I}_{N_2} \right] [\ell_2 \otimes \mathbf{h}_H(\mathbf{R}; \mathbf{b}_1)].$$

It is then straightforward to show that the long-run covariance matrix of $\sqrt{T} \bar{\mathbf{h}}_{LT}(\mathbf{b}_i, \mathbf{b}_{ii})$ will be

$$\left\{ \begin{pmatrix} \boldsymbol{\Phi}_i & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_{ii} \end{pmatrix} \otimes \mathbf{I}_{N_2} \right\} [\ell_2 \ell_2' \otimes \mathbf{S}_{H,T}(\mathbf{b}_1)] \left[\begin{pmatrix} \boldsymbol{\Phi}'_i & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}'_{ii} \end{pmatrix} \otimes \mathbf{I}_{N_2} \right],$$

where $\mathbf{S}_{H,T}(\mathbf{b}_1)$ denotes a HAC estimator of the long-run covariance matrix of $\sqrt{T}\bar{\mathbf{h}}_{HT}(\mathbf{b}_1)$, a generalised inverse of which is

$$\left[\begin{pmatrix} \Phi_i^{-1'} & \mathbf{0} \\ \mathbf{0} & \Phi_{ii}^{-1'} \end{pmatrix} \otimes \mathbf{I}_{N_2} \right] \left[\frac{\ell_2 \ell_2'}{4} \otimes \mathbf{S}_{H,T}^{-1}(\mathbf{b}_1) \right] \left[\begin{pmatrix} \Phi_i^{-1} & \mathbf{0} \\ \mathbf{0} & \Phi_{ii}^{-1} \end{pmatrix} \otimes \mathbf{I}_{N_2} \right].$$

Hence, it is clear that applying the optimal singular CU-GMM approach developed in section 3 to $E[\mathbf{h}_L(\mathbf{R}; \mathbf{b}_i, \mathbf{b}_{ii})] = \mathbf{0}$ is numerically equivalent to applying the standard optimal CU-GMM approach to $E[\mathbf{h}_H(\mathbf{R}; \mathbf{b}_1)] = \mathbf{0}$. \square

Proposition 5:

As usual, we prove the CU-GMM result, which in turn implies the GMM result. The proof starts with the connection between the uncentred representing portfolios test and the regression test. We can express the two sets of influence functions as transformations and reparametrisations of the influence functions that we can use to estimate $\boldsymbol{\nu}_1$, $\boldsymbol{\nu}_2$, $\boldsymbol{\Gamma}_{11}$ and $\boldsymbol{\Gamma}_{21}$

$$E \begin{bmatrix} \mathbf{R}_1 - \boldsymbol{\nu}_1 \\ \mathbf{R}_2 - \boldsymbol{\nu}_2 \\ \text{vec}(\mathbf{R}_1 \mathbf{R}_1' - \boldsymbol{\Gamma}_{11}) \\ \text{vec}(\mathbf{R}_2 \mathbf{R}_1' - \boldsymbol{\Gamma}_{21}) \end{bmatrix} = E \begin{bmatrix} \mathbf{f}_1(\mathbf{R}_1; \boldsymbol{\nu}_1) \\ \mathbf{f}_2(\mathbf{R}_2; \boldsymbol{\nu}_2) \\ \mathbf{f}_3(\mathbf{R}_1; \text{vec}(\boldsymbol{\Gamma}_{11})) \\ \mathbf{f}_4(\mathbf{R}; \text{vec}(\boldsymbol{\Gamma}_{21})) \end{bmatrix},$$

under the constraints that spanning imposes, namely $\boldsymbol{\nu}_2 = \boldsymbol{\Gamma}_{21} \boldsymbol{\Gamma}_{11}^{-1} \boldsymbol{\nu}_1$ and $\ell_2 = \boldsymbol{\Gamma}_{21} \boldsymbol{\Gamma}_{11}^{-1} \ell_1$.

Obviously, there are redundant terms in $\mathbf{f}_3(\cdot)$, so we should use $\text{vech}(\cdot)$ instead of $\text{vec}(\cdot)$, but we ignore this point for ease of exposition. In this context, the uncentred representing portfolios test is based on

$$\begin{pmatrix} \mathbf{R}_1 \mathbf{R}_1' \phi_1^+ - \mathbf{R}_1 \\ \mathbf{R}_2 \mathbf{R}_1' \phi_1^+ - \mathbf{R}_2 \\ \mathbf{R}_1 \mathbf{R}_1' \phi_1^* - \ell_1 \\ \mathbf{R}_2 \mathbf{R}_1' \phi_1^* - \ell_2 \end{pmatrix} = \begin{pmatrix} -\mathbf{I}_1 & \mathbf{0} & \phi_1^{+'} \otimes \mathbf{I}_1 & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_2 & \mathbf{0} & \phi_1^{+'} \otimes \mathbf{I}_2 \\ \mathbf{0} & \mathbf{0} & \phi_1^{*'} \otimes \mathbf{I}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \phi_1^{*'} \otimes \mathbf{I}_2 \end{pmatrix} \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \\ \mathbf{f}_4 \end{pmatrix},$$

$$\phi_1^+ = \boldsymbol{\Gamma}_{11}^{-1} \boldsymbol{\nu}_1, \quad \phi_1^* = \boldsymbol{\Gamma}_{11}^{-1} \ell_1,$$

under the null of spanning, while the HK test is based on

$$\begin{pmatrix} \mathbf{R}_2 - \mathbf{B} \mathbf{R}_1 \\ \mathbf{R}_1 \otimes (\mathbf{R}_2 - \mathbf{B} \mathbf{R}_1) \end{pmatrix} = \begin{pmatrix} -\mathbf{B} & \mathbf{I}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -(\mathbf{I}_1 \otimes \mathbf{B}) & \mathbf{I}_{21} \end{pmatrix} \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \\ \mathbf{f}_4 \end{pmatrix},$$

$$\mathbf{B} = \boldsymbol{\Gamma}_{21} \boldsymbol{\Gamma}_{11}^{-1}, \quad \mathbf{B} \ell_1 = \ell_2.$$

under the null again.

Now we saturate both systems until we get the same number of moments as in \mathbf{f} so as to achieve exactly identified systems. The representing portfolios system can be saturated with the estimation of all the elements of $\mathbf{\Gamma}_{11}$ and $\mathbf{\Gamma}_{21}$ excluding one of the columns of $\mathbf{\Gamma}_{21}$, while the regression system can be saturated with the estimation of $\boldsymbol{\nu}_1$ and $\mathbf{\Gamma}_{11}$. Both tests are numerically equal because CU-GMM is invariant to parameter-dependent transformations.

We can then reparametrise both systems back to their original parameters. Both tests continue to be numerically equal because CU-GMM is not affected by bijective reparametrisations. Now we can exclude the moments that we used to saturate both systems and still preserve a numerical equivalence, as in the proofs of Propositions 2 and 3.

Finally, we show the connection between the centred and uncentred representing portfolios test. Using (3), we can reparametrise the mean condition (12) as

$$\begin{aligned} \mathbf{R}\mathbf{R}'_1\phi_1^+ - \mathbf{R} &= \mathbf{R}\mathbf{R}'_1 \left[\frac{1}{1 + \boldsymbol{\nu}'_1\varphi_1^+} \varphi_1^+ \right] - \mathbf{R} \\ &= \left(\frac{1}{1 + \boldsymbol{\nu}'_1\varphi_1^+} \right) [\mathbf{R}\mathbf{R}'_1\varphi_1^+ - (1 + \boldsymbol{\nu}'_1\varphi_1^+) \mathbf{R}] = \left(\frac{1}{1 + \boldsymbol{\nu}'_1\varphi_1^+} \right) [\mathbf{R}(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \varphi_1^+ - \mathbf{R}]. \end{aligned}$$

The cost condition of (12) plus $\boldsymbol{\nu}'_1\varphi_1^*$ times the mean condition is

$$\begin{aligned} &[\mathbf{R}\mathbf{R}'_1\phi_1^* - \ell_N] + (\boldsymbol{\nu}'_1\varphi_1^*) [\mathbf{R}\mathbf{R}'_1\phi_1^+ - \mathbf{R}] \\ &= \left[\mathbf{R}\mathbf{R}'_1 \left[\varphi_1^* - \frac{\boldsymbol{\nu}'_1\varphi_1^*}{1 + \boldsymbol{\nu}'_1\varphi_1^+} \varphi_1^+ \right] - \ell_N \right] + (\boldsymbol{\nu}'_1\varphi_1^*) \left[\mathbf{R}\mathbf{R}'_1 \left[\frac{1}{1 + \boldsymbol{\nu}'_1\varphi_1^+} \varphi_1^+ \right] - \mathbf{R} \right] \\ &= \left[(\mathbf{R}\mathbf{R}'_1\varphi_1^* - \ell_N) - \left(\frac{\boldsymbol{\nu}'_1\varphi_1^*}{1 + \boldsymbol{\nu}'_1\varphi_1^+} \right) \mathbf{R}\mathbf{R}'_1\varphi_1^+ \right] + \left(\frac{\boldsymbol{\nu}'_1\varphi_1^*}{1 + \boldsymbol{\nu}'_1\varphi_1^+} \right) [\mathbf{R}\mathbf{R}'_1\varphi_1^+ - (1 + \boldsymbol{\nu}'_1\varphi_1^+) \mathbf{R}] \\ &= \mathbf{R}(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \varphi_1^* - \ell_N. \end{aligned}$$

Therefore, we can relate the two systems as

$$\begin{pmatrix} \mathbf{R}_1 - \boldsymbol{\nu}_1 \\ \mathbf{R}\mathbf{R}'_1\phi_1^+ - \mathbf{R} \\ \mathbf{R}\mathbf{R}'_1\phi_1^* - \ell_N \end{pmatrix} = \begin{pmatrix} \mathbf{I}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{1 + \boldsymbol{\nu}'_1\varphi_1^+} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & -\frac{\boldsymbol{\nu}'_1\varphi_1^*}{1 + \boldsymbol{\nu}'_1\varphi_1^+} \mathbf{I}_N & \mathbf{I}_N \end{pmatrix} \begin{pmatrix} \mathbf{R}_1 - \boldsymbol{\nu}_1 \\ \mathbf{R}(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \varphi_1^+ - \mathbf{R} \\ \mathbf{R}(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \varphi_1^* - \ell_N \end{pmatrix}.$$

Since the addition of $\mathbf{R}_1 - \boldsymbol{\nu}_1$ on the left does not affect the test, the CU-GMM J tests will be equal. \square

B Proofs of Lemmata

Lemmata 1, 3, and 5 are trivial to prove. Therefore, here we focus on the remaining lemmata.

Lemma 2:

Let us re-order the estimating functions in (12) as

$$\begin{pmatrix} \mathbf{R}_1 \mathbf{R}'_1 \phi_1^+ - \mathbf{R}_1 \\ \mathbf{R}_1 \mathbf{R}'_1 \phi_1^* - \ell_{N_1} \\ \mathbf{R}_2 \mathbf{R}'_1 \phi_1^+ - \mathbf{R}_2 \\ \mathbf{R}_2 \mathbf{R}'_1 \phi_1^* - \ell_{N_2} \end{pmatrix} = \begin{bmatrix} \mathbf{h}_{U1}(\mathbf{R}; \phi_1) \\ \mathbf{h}_{U2}(\mathbf{R}; \phi_1) \end{bmatrix}.$$

We just need to check that the condition (C1) in Lemma C1 (see Appendix C below) is satisfied for $s_1 = s_2 = 0$, with the additional peculiarity that since the second block of moment conditions contains no additional parameters in this case, we simply have to check that the left hand side of (C1) is equal to 0. But since $\mathbf{D}_{U1} = \mathbf{I}_2 \otimes \mathbf{\Gamma}_{11}$, and $\mathbf{D}_{U2} = \mathbf{I}_2 \otimes \mathbf{\Gamma}_{21}$, it is easy to see that $\mathbf{D}_{U2} \mathbf{D}_{U1}^{-1} \mathbf{S}_{U11}$ will be equal to

$$\begin{aligned} & \left\{ \begin{array}{l} (1 + B_1)^{-1} [1 + \kappa B_1 (1 + B_1)^{-1}] \\ A_1 (1 + B_1)^{-1} \kappa \\ A_1 (1 + B_1)^{-2} \kappa \end{array} \right\} \otimes \mathbf{\Gamma}_{21} \\ & [C_1 - A_1^2 (1 + B_1)^{-1}] + \kappa \left([C_1 - A_1^2 (1 + B_1)^{-1}] - A_1^2 (1 + B_1)^{-2} \right) \\ & + \left\{ \begin{array}{l} -2(1 + B_1)^{-2} + (3B_1^2 (1 + B_1)^{-2} - 5B_1 (1 + B_1)^{-1} + 2) \kappa \\ A_1 (1 + B_1)^{-2} (2 - 3\kappa) \end{array} \right\} \\ & -2A_1^2 (1 + B_1)^{-2} + \{3A_1^2 (1 + B_1)^{-2} - [C_1 - A_1^2 (1 + B_1)^{-1}]\} \kappa \left. \right\} \otimes \mathbf{\Gamma}_{21} \mathbf{\Gamma}_{11}^{-1} \boldsymbol{\nu}_1 \boldsymbol{\nu}'_1 \\ & + \begin{pmatrix} 0 & 0 \\ 0 & 1 + 2\kappa \end{pmatrix} \otimes \mathbf{\Gamma}_{21} \mathbf{\Gamma}_{11}^{-1} \ell_{N_1} \ell'_{N_1} \\ & + \begin{bmatrix} 0 & (1 + B_1)^{-1} \kappa \\ (1 + B_1)^{-1} \kappa & -2A_1 (1 + B_1)^{-1} \kappa \end{bmatrix} \otimes \mathbf{\Gamma}_{21} \mathbf{\Gamma}_{11}^{-1} (\boldsymbol{\nu}_1 \ell'_{N_1} + \ell_{N_1} \boldsymbol{\nu}'_1), \end{aligned}$$

where \mathbf{S}_{U11} has been obtained from Lemma D1 in Appendix D below, $A_1 = cov(p_1^{**}, p_1^{++}) = \boldsymbol{\nu}'_1 \boldsymbol{\Sigma}_{11}^{-1} \ell_{N_1}$, $B_1 = V(p_1^{++}) = \boldsymbol{\nu}'_1 \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\nu}_1$, and $C_1 = V(p_1^{**}) = \ell'_{N_1} \boldsymbol{\Sigma}_{11}^{-1} \ell_{N_1}$. Finally, since $\boldsymbol{\nu}_2 = \mathbf{\Gamma}_{21} \mathbf{\Gamma}_{11}^{-1} \boldsymbol{\nu}_1$ and $\ell_{N_2} = \mathbf{\Gamma}_{21} \mathbf{\Gamma}_{11}^{-1} \ell_{N_1}$ under the null of spanning, we will have that $\mathbf{D}_{U2} \mathbf{D}_{U1}^{-1} \mathbf{S}_{U11} = \mathbf{S}_{U21}$, where \mathbf{S}_{U21} can also be obtained from Lemma D1. \square

Lemma 4:

Let us begin again by re-ordering the estimating functions (15) as

$$\begin{bmatrix} \mathbf{R}_1 - \boldsymbol{\nu}_1 \\ \mathbf{R}_1(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^+ - \mathbf{R}_1 \\ \mathbf{R}_1(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^* - \ell_{N_1} \\ \mathbf{R}_2(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^+ - \mathbf{R}_2 \\ \mathbf{R}_2(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^* - \ell_{N_2} \end{bmatrix} = \begin{bmatrix} \mathbf{f}_1(\mathbf{R}_1; \boldsymbol{\varphi}_1, \boldsymbol{\nu}_1) \\ \mathbf{f}_2(\mathbf{R}; \boldsymbol{\varphi}_1, \boldsymbol{\nu}) \end{bmatrix}.$$

We have to check that condition (C1) in Lemma C1 applied to

$$\mathbf{h}_1(\mathbf{R}; \boldsymbol{\theta}) = \mathbf{f}_1(\mathbf{R}_1; \boldsymbol{\varphi}_1, \boldsymbol{\nu}_1), \quad \mathbf{h}_2(\mathbf{R}; \boldsymbol{\theta}) = \mathbf{f}_2(\mathbf{R}; \boldsymbol{\varphi}_1, \boldsymbol{\nu}_1),$$

is satisfied, where in this case $\boldsymbol{\theta} = (\boldsymbol{\nu}_1', \boldsymbol{\varphi}_1')'$, and again there are no elements in $\boldsymbol{\rho}$.

To do so, we must first orthogonalise the two blocks of moment conditions by regressing the second set of estimating functions evaluated at $\boldsymbol{\theta}^0$ onto the first one evaluated at $\boldsymbol{\theta}^0$. The regression coefficients, though, are not uniquely defined since the singularity in the system is confined to $\mathbf{h}_1(\mathbf{R}; \boldsymbol{\theta})$. Nevertheless, it is easy to see that $\mathbf{g}_2(\mathbf{R}; \boldsymbol{\theta}^0)$, where

$$\mathbf{g}_2(\mathbf{R}; \boldsymbol{\theta}) = \mathbf{h}_2(\mathbf{R}; \boldsymbol{\theta}) - \mathcal{B}\mathbf{h}_1(\mathbf{R}; \boldsymbol{\theta}),$$

$$\mathcal{B} = \begin{bmatrix} \mathbf{0}_{2N_2 \times N_1} & \mathbf{I}_2 \otimes \mathbf{B}^0 \end{bmatrix}, \quad \mathbf{B}^0 = \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1},$$

will be orthogonal to $\mathbf{h}_1(\mathbf{R}; \boldsymbol{\theta}^0)$ because $\mathbf{S}_{21} = \mathcal{B}\mathbf{S}_{11}$ given the result in Lemma D2 in Appendix D below. In this respect, note that $(\mathbf{I}_2 \otimes \mathbf{B}^0)$ can be interpreted as the coefficients in the multivariate regression of

$$\begin{bmatrix} \mathbf{R}_2(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^+ - \mathbf{R}_2 \\ \mathbf{R}_2(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^* - \ell_{N_2} \end{bmatrix}$$

on

$$\begin{bmatrix} \mathbf{R}_1(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^+ - \mathbf{R}_1 \\ \mathbf{R}_1(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\varphi}_1^* - \ell_{N_1} \end{bmatrix}.$$

The next step is to find the appropriate reparametrisation that exploits the singularity in $\mathbf{h}_1(\mathbf{R}; \boldsymbol{\theta})$, together with the Moore-Penrose inverses of the covariance matrices of $\mathbf{h}_1(\mathbf{R}; \boldsymbol{\theta}^0)$ and $\mathbf{g}_2(\mathbf{R}; \boldsymbol{\theta}^0)$. Given that the latter covariance matrix has full rank, we should concentrate on the first block. Specifically, we could work with the reparametrisation $\boldsymbol{\theta} = \mathbf{l}_1(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1) = (\boldsymbol{\nu}_1' \quad \alpha_1 \quad \boldsymbol{\varphi}_{-1}^{+'} \quad \boldsymbol{\varphi}_1^{*'})'$, where $\boldsymbol{\varphi}_{-1}^+$ contains the last $N_1 - 1$ elements of $\boldsymbol{\varphi}_1^+$, and $\alpha_1 = \boldsymbol{\varphi}_1^{*'} \boldsymbol{\nu}_1 - \boldsymbol{\varphi}_1^{+'} \ell_{N_1}$, together with the non-singular set of moment conditions

$$\mathbf{P}'_1 \mathbf{h}_1[\mathbf{R}; \mathbf{l}_1(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1)],$$

$$\mathbf{P}'_2 \mathbf{g}_2[\mathbf{R}; \mathbf{l}_1(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1)],$$

where \mathbf{P}_1 are the eigenvectors associated with the non-zero eigenvalues of the asymptotic covariance matrix of $\sqrt{T}\bar{\mathbf{h}}_{1T}(\boldsymbol{\theta}^0)$, while \mathbf{P}_2 contains all the eigenvectors of the asymptotic covariance matrix of $\sqrt{T}\bar{\mathbf{g}}_{2T}(\boldsymbol{\theta}^0)$.

In view of Lemma C1, the condition for asymptotic equivalence between the optimal GMMS estimators of $\boldsymbol{\theta}$ based on $E[\mathbf{h}_1(\mathbf{R}; \boldsymbol{\theta})] = \mathbf{0}$ alone, and the ones that also exploit the information in $E[\mathbf{h}_2(\mathbf{R}; \boldsymbol{\theta})] = \mathbf{0}$, is simply that

$$E \left[\frac{\partial (\mathbf{P}'_2 \mathbf{g}_2[\mathbf{R}; \mathbf{l}_1(0, \boldsymbol{\beta}_1^0)])}{\partial \boldsymbol{\beta}'_1} \right] = \mathbf{0}.$$

But since \mathbf{P}_2 has full rank, an equivalent condition in terms of the original moment conditions and parameters is

$$E \left[\frac{\partial \mathbf{h}_2(\mathbf{R}; \boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}'} - \mathcal{B} \frac{\partial \mathbf{h}_1(\mathbf{R}; \boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}'} \right] \mathbf{L}_{1\beta_1}(0, \boldsymbol{\beta}_1^0) = \mathbf{0},$$

where

$$\mathbf{L}_{1\beta_1}(0, \boldsymbol{\beta}_1) = \frac{\partial \mathbf{l}_1(0, \boldsymbol{\beta}_1)}{\partial \boldsymbol{\beta}'_1} = \begin{pmatrix} \mathbf{I}_{N_1} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\varphi}_1^{*'} & -\ell'_{N_1-1} & \boldsymbol{\nu}'_1 \\ \mathbf{0} & \mathbf{I}_{N_1-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{N_1} \end{pmatrix}.$$

We can finally show that

$$E \left[\frac{\partial \mathbf{h}_2(\mathbf{R}; \boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}'} \right] = \begin{pmatrix} -\boldsymbol{\nu}_2 \boldsymbol{\varphi}_1^{+'} & \boldsymbol{\Sigma}_{21} & \mathbf{0} \\ -\boldsymbol{\nu}_2 \boldsymbol{\varphi}_1^{*'} & \mathbf{0} & \boldsymbol{\Sigma}_{21} \end{pmatrix},$$

$$E \left[\frac{\partial \mathbf{h}_1(\mathbf{R}; \boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}'} \right] = \begin{pmatrix} -\mathbf{I}_{N_1} & \mathbf{0} & \mathbf{0} \\ -\boldsymbol{\nu}_1 \boldsymbol{\varphi}_1^{+'} & \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ -\boldsymbol{\nu}_1 \boldsymbol{\varphi}_1^{*'} & \mathbf{0} & \boldsymbol{\Sigma}_{11} \end{pmatrix},$$

so that under the null of spanning

$$E \left[\frac{\partial \mathbf{h}_2(\mathbf{R}; \boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}'} - \mathcal{B} \frac{\partial \mathbf{h}_1(\mathbf{R}; \boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}'} \right] = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix},$$

whence the result follows. \square

Lemma 6:

First of all, given the result of Lemma D3 below, the relevant long-run variance is

$$\mathbf{S}_L = (F \otimes \mathbf{I}_{N_2}) \left[\begin{pmatrix} 1 & \boldsymbol{\nu}'_1 \\ \boldsymbol{\nu}_1 & (\kappa + 1)\boldsymbol{\Sigma}_{11} + \boldsymbol{\nu}_1 \boldsymbol{\nu}'_1 \end{pmatrix} \otimes \boldsymbol{\Omega} \right] (F' \otimes \mathbf{I}_{N_2}),$$

where

$$F = \begin{pmatrix} 1 & 0 & \mathbf{0} \\ 0 & 1 & \mathbf{0} \\ \mathbf{0} & -\ell_{N_1-1} & \mathbf{I}_{N_1-1} \end{pmatrix}.$$

Similarly, the Jacobian of the moment conditions (17) will be

$$\mathbf{D}_L = - \left[F \begin{pmatrix} 1 & \boldsymbol{\nu}'_1 \\ \boldsymbol{\nu}_1 & \boldsymbol{\Gamma}_{11} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ -\ell'_{N_1-1} \\ \mathbf{I}_{N_1-1} \end{pmatrix} \otimes \mathbf{I}_{N_2} \right].$$

As a result, $\mathbf{D}'_L \mathbf{S}_L^{-1}$ will be equal to

$$\begin{aligned} & - \left\{ \begin{pmatrix} \mathbf{0} & -\ell_{N_1-1} & \mathbf{I}_{N_1-1} \end{pmatrix} \begin{pmatrix} 1 & \boldsymbol{\nu}'_1 \\ \boldsymbol{\nu}_1 & \boldsymbol{\Gamma}_{11} \end{pmatrix} \begin{bmatrix} 1 & \boldsymbol{\nu}'_1 \\ \boldsymbol{\nu}_1 & (\kappa+1)\boldsymbol{\Sigma}_{11} + \boldsymbol{\nu}_1\boldsymbol{\nu}'_1 \end{bmatrix}^{-1} F^{-1} \otimes \boldsymbol{\Omega}^{-1} \right\} \\ & = - \left\{ \begin{pmatrix} \mathbf{0} & -\ell_{N_1-1} & \mathbf{I} \end{pmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ \kappa(\kappa+1)^{-1}\boldsymbol{\nu}_1 & (\kappa+1)^{-1}\mathbf{I}_{N_1} \end{bmatrix} \begin{pmatrix} 1 & 0 & \mathbf{0} \\ 0 & 1 & \mathbf{0} \\ \mathbf{0} & \ell_{N_1-1} & \mathbf{I}_{N_1-1} \end{pmatrix} \otimes \boldsymbol{\Omega}^{-1} \right\} \\ & = - \left\{ (\kappa+1)^{-1} [\kappa(\boldsymbol{\nu}_{1b} - \boldsymbol{\nu}_{1a}\ell_{N_1-1}) \quad \mathbf{0} \quad \mathbf{I}_{N_1-1}] \otimes \boldsymbol{\Omega}^{-1} \right\}. \end{aligned}$$

which confirms that the optimal instrument is proportional to a constant translation of \mathbf{R}_{1b} - $R_{1a}\ell_{N_1-1}$. \square

C Some useful GMM results

C.1 Efficiency improving moments

We extend to the singular case earlier results in Gouriéroux, Monfort and Renault (1996) and Lezan and Renault (1999), which in turn nest Theorem 1 in Breusch et al. (1999).

Let $\mathbf{h}_1(\mathbf{x}_t; \boldsymbol{\theta})$ denote a set of n_1 estimating functions for $0 < k_1 \leq n_1$ unknown parameters $\boldsymbol{\theta}$, whose true values are implicitly defined by $E[\mathbf{h}_1(\mathbf{x}_t; \boldsymbol{\theta}^0)] = \mathbf{0}$, and let $\mathbf{h}_2(\mathbf{x}_t; \boldsymbol{\theta}, \boldsymbol{\rho})$ denote an additional set of n_2 estimating functions that depend not only on $\boldsymbol{\theta}$ but also on some additional $0 \leq k_2 \leq n_2$ unknown parameters $\boldsymbol{\rho}$, whose true values are implicitly defined by $E[\mathbf{h}_2(\mathbf{x}_t; \boldsymbol{\theta}^0, \boldsymbol{\rho}^0)] = \mathbf{0}$. Let

$$\mathbf{S}(\boldsymbol{\theta}^0, \boldsymbol{\rho}^0) = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}'_{21} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix}$$

denote the joint asymptotic covariance matrix of $[\sqrt{T}\bar{\mathbf{h}}'_{1T}(\boldsymbol{\theta}^0), \sqrt{T}\bar{\mathbf{h}}'_{2T}(\boldsymbol{\theta}^0, \boldsymbol{\rho}^0)]'$, and factorise it as

$$\begin{pmatrix} \mathbf{I}_{n_1} & \mathbf{0} \\ -\mathbf{S}_{21}\mathbf{S}_{11}^- & \mathbf{I}_{n_2} \end{pmatrix} \begin{pmatrix} \mathbf{S}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22|1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{n_1} & -\mathbf{S}_{11}^-\mathbf{S}'_{21} \\ \mathbf{0} & \mathbf{I}_{n_2} \end{pmatrix},$$

where \mathbf{S}_{11}^- is some generalised inverse of \mathbf{S}_{11} , and

$$\mathbf{S}_{22|1} = \mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^-\mathbf{S}'_{21}$$

can be regarded as the asymptotic covariance matrix of $\sqrt{T}\bar{\mathbf{g}}_{2T}(\boldsymbol{\theta}^0, \boldsymbol{\rho}^0)$, where

$$\mathbf{g}_2(\mathbf{x}_t; \boldsymbol{\theta}, \boldsymbol{\rho}) = \mathbf{h}_2(\mathbf{x}_t; \boldsymbol{\theta}, \boldsymbol{\rho}) - \mathbf{S}_{21}\mathbf{S}_{11}^-\mathbf{h}_1(\mathbf{x}_t; \boldsymbol{\theta}),$$

are some transformed estimating functions, which are invariant to the choice \mathbf{S}_{11}^- . In addition, let

$$\boldsymbol{\Pi}(\boldsymbol{\theta}, \boldsymbol{\rho}) = \begin{bmatrix} \boldsymbol{\Pi}_{11}(\boldsymbol{\theta}) & \boldsymbol{\Pi}_{12}(\boldsymbol{\theta}, \boldsymbol{\rho}) \\ \mathbf{0} & \boldsymbol{\Pi}_{22}(\boldsymbol{\theta}, \boldsymbol{\rho}) \end{bmatrix}$$

denote a $(n_1 + n_2) \times (s_1 + s_2)$ matrix of continuously differentiable known functions of $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$, with $0 \leq s_1 \leq k_1$ and $0 \leq s_2 \leq k_2$, such that

$$\begin{bmatrix} \boldsymbol{\Pi}'_{11}(\boldsymbol{\theta}) & \mathbf{0} \\ \boldsymbol{\Pi}'_{12}(\boldsymbol{\theta}, \boldsymbol{\rho}) & \boldsymbol{\Pi}'_{22}(\boldsymbol{\theta}, \boldsymbol{\rho}) \end{bmatrix} \sqrt{T} \begin{bmatrix} \bar{\mathbf{h}}_{1,T}(\boldsymbol{\theta}^0) \\ \bar{\mathbf{g}}_{2,T}(\boldsymbol{\theta}, \boldsymbol{\rho}) \end{bmatrix} \xrightarrow{L^2} \mathbf{0} \Leftrightarrow \begin{bmatrix} \mathbf{m}_1(\boldsymbol{\theta}) \\ \mathbf{m}_2(\boldsymbol{\theta}, \boldsymbol{\rho}) \end{bmatrix} = \mathbf{m}(\boldsymbol{\theta}, \boldsymbol{\rho}) = \mathbf{0},$$

where $\mathbf{m}(\boldsymbol{\theta}, \boldsymbol{\rho})$ is a $(s_1 + s_2)$ continuously differentiable function of $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$ with

$$\text{rank} \left[\frac{\partial \mathbf{m}(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0)}{\partial (\boldsymbol{\theta}', \boldsymbol{\rho}')} \right] = (s_1 + s_2).$$

Moreover, assume that $\mathbf{m}(\boldsymbol{\theta}^0, \boldsymbol{\rho}^0) = \mathbf{0}$ if $s_1 + s_2 > 0$, and that $\text{rank}[\mathbf{S}_{11}] = n_1 - s_1$, $\text{rank}[\mathbf{S}_{22|1}] = n_2 - s_2$, so that $\text{rank}[\mathbf{S}(\boldsymbol{\theta}^0, \boldsymbol{\rho}^0)] = (n_1 + n_2) - (s_1 + s_2)$.

Lemma C1 *Subject to the required regularity conditions, the optimal GMM estimator of $\boldsymbol{\theta}$ based on*

$$E[\mathbf{h}_1(\mathbf{x}_t; \boldsymbol{\theta})] = \mathbf{0}$$

alone is asymptotically as efficient as the optimal GMM estimator that additionally uses

$$E[\mathbf{h}_2(\mathbf{x}_t; \boldsymbol{\theta}, \boldsymbol{\rho})] = \mathbf{0}$$

if and only if

$$\mathbf{P}'_2 \mathbf{D}_{2\boldsymbol{\theta}}(\boldsymbol{\theta}^0, \boldsymbol{\rho}^0) \mathbf{L}_{1\beta_1}(\mathbf{0}, \boldsymbol{\beta}_1) \in \langle \mathbf{P}'_2 \mathbf{D}_{2\boldsymbol{\rho}}(\boldsymbol{\theta}^0, \boldsymbol{\rho}^0) \mathbf{L}_{2\beta_2}(\mathbf{0}, \boldsymbol{\beta}_1, \mathbf{0}, \boldsymbol{\beta}_2) \rangle, \quad (\text{C1})$$

where $\mathbf{P}_2 \boldsymbol{\Delta}_2 \mathbf{P}'_2$ provides the spectral decompositions of $\mathbf{S}_{22|1}$,

$$\begin{aligned} \mathbf{D}_{2\boldsymbol{\theta}}(\boldsymbol{\theta}, \boldsymbol{\rho}) &= E \left[\frac{\partial \mathbf{g}_2(\mathbf{x}_t; \boldsymbol{\theta}, \boldsymbol{\rho})}{\partial \boldsymbol{\theta}'} \right] = E \left[\frac{\partial \mathbf{h}_2(\mathbf{x}_t; \boldsymbol{\theta}, \boldsymbol{\rho})}{\partial \boldsymbol{\theta}'} \right] - \mathbf{S}_{21} \mathbf{S}_{11}^- E \left[\frac{\partial \mathbf{h}_1(\mathbf{x}_t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right], \\ \mathbf{D}_{2\boldsymbol{\rho}}(\boldsymbol{\theta}, \boldsymbol{\rho}) &= E \left[\frac{\partial \mathbf{g}_2(\mathbf{x}_t; \boldsymbol{\theta}, \boldsymbol{\rho})}{\partial \boldsymbol{\rho}'} \right] = E \left[\frac{\partial \mathbf{h}_2(\mathbf{x}_t; \boldsymbol{\theta}, \boldsymbol{\rho})}{\partial \boldsymbol{\rho}'} \right], \end{aligned}$$

$$\begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\beta}_1 \\ \boldsymbol{\alpha}_2 \\ \boldsymbol{\beta}_2 \end{pmatrix} = \begin{bmatrix} \mathbf{M}_1(\boldsymbol{\theta}) \\ \mathbf{M}_2(\boldsymbol{\theta}, \boldsymbol{\rho}) \end{bmatrix},$$

with the first s_1 entries of $\mathbf{M}_1(\boldsymbol{\theta})$ being $\boldsymbol{\alpha}_1 = \mathbf{m}_1(\boldsymbol{\theta})$ and the first s_2 entries of $\mathbf{M}_2(\boldsymbol{\theta}, \boldsymbol{\rho})$ being $\boldsymbol{\alpha}_2 = \mathbf{m}_2(\boldsymbol{\theta}, \boldsymbol{\rho})$,

$$\begin{aligned} \mathbf{L}_{1\beta_1}(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1) &= \frac{\partial \mathbf{l}_1(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1)}{\partial \boldsymbol{\beta}'_1}, \\ \mathbf{L}_{2\beta_2}(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1, \boldsymbol{\alpha}_2, \boldsymbol{\beta}_2) &= \frac{\partial \mathbf{l}_2(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1, \boldsymbol{\alpha}_2, \boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}'_2}, \end{aligned}$$

$$\begin{bmatrix} \mathbf{l}_1[\mathbf{M}_1(\boldsymbol{\theta})] \\ \mathbf{l}_2[\mathbf{M}_1(\boldsymbol{\theta}), \mathbf{M}_2(\boldsymbol{\theta}, \boldsymbol{\rho})] \end{bmatrix} = \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\rho} \end{pmatrix},$$

and $\langle \mathbf{A} \rangle$ denotes the column space of the matrix \mathbf{A} .

Proof. We know that the moment condition $E[\mathbf{g}_2(\mathbf{x}_t; \boldsymbol{\theta}, \boldsymbol{\rho})] = \mathbf{0}$ can replace $E[\mathbf{h}_2(\mathbf{x}_t; \boldsymbol{\theta}, \boldsymbol{\rho})] = \mathbf{0}$ without loss of asymptotic efficiency. Given that $\sqrt{T}\bar{\mathbf{g}}_{2T}(\boldsymbol{\theta}^0, \boldsymbol{\rho}^0)$ and $\sqrt{T}\bar{\mathbf{h}}_{1T}(\boldsymbol{\theta}^0)$ are asymptotically orthogonal by construction, the discussion in section 3 implies that the right way to exploit the potential singularities in both sets of moment conditions is to estimate the parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ from the transformed moment conditions:

$$\begin{aligned} &\mathbf{P}'_1 \mathbf{h}_1[\mathbf{x}_t; \mathbf{l}_1(\mathbf{0}, \boldsymbol{\beta}_1)], \\ &\mathbf{P}'_2 \mathbf{h}_2[\mathbf{x}_t; \mathbf{l}_1(\mathbf{0}, \boldsymbol{\beta}_1), \mathbf{l}_2(\mathbf{0}, \boldsymbol{\beta}_1, \mathbf{0}, \boldsymbol{\beta}_2)], \end{aligned}$$

where $\mathbf{P}_1 \boldsymbol{\Lambda}_1 \mathbf{P}'_1$ provides the spectral decomposition of \mathbf{S}_{11} . Let us denote by $\mathbf{D}_{11}(\boldsymbol{\beta}_1^0)$, $\mathbf{D}_{21}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)$, and $\mathbf{D}_{22}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)$ the corresponding Jacobians.

Let $\hat{\boldsymbol{\beta}}_{1T}$ and $\hat{\boldsymbol{\beta}}_{2T}$ denote the optimal GMM estimators of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ based on both subsets of moment conditions. Similarly, let $\bar{\boldsymbol{\beta}}_{1T}$ denote the optimal GMM estimator based on the first subset of moment conditions. Since we have transformed the potentially singular problem in a non-singular one, under standard regularity conditions the asymptotic variances of these estimators will be:

$$\begin{aligned} \lim_{T \rightarrow \infty} V \begin{bmatrix} \sqrt{T}(\hat{\boldsymbol{\beta}}_{1T} - \boldsymbol{\beta}_1^0) \\ \sqrt{T}(\hat{\boldsymbol{\beta}}_{2T} - \boldsymbol{\beta}_2^0) \end{bmatrix} &= \left\{ \begin{bmatrix} \mathbf{D}'_{11}(\boldsymbol{\beta}_1^0) & \mathbf{D}'_{21}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0) \\ \mathbf{0} & \mathbf{D}'_{22}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0) \end{bmatrix} \right. \\ &\times \left. \begin{bmatrix} \boldsymbol{\Lambda}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_2^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{D}_{11}(\boldsymbol{\beta}_1^0) & \mathbf{0} \\ \mathbf{D}_{21}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0) & \mathbf{D}_{22}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0) \end{bmatrix} \right\}^{-1}, \end{aligned}$$

and

$$\lim_{T \rightarrow \infty} V[\sqrt{T}(\bar{\boldsymbol{\beta}}_{1T} - \boldsymbol{\beta}_1^0)] = [\mathbf{D}'_{11}(\boldsymbol{\beta}_1^0) \boldsymbol{\Lambda}_1^{-1} \mathbf{D}_{11}(\boldsymbol{\beta}_1^0)]^{-1}.$$

Hence, we need to compare this last expression with $\lim_{T \rightarrow \infty} V[\sqrt{T}(\hat{\boldsymbol{\beta}}_{1T} - \boldsymbol{\beta}_1^0)]$, which is given by:

$$\begin{aligned} & \{\mathbf{D}'_{11}(\boldsymbol{\beta}_1^0)\boldsymbol{\Lambda}_1^{-1}\mathbf{D}_{11}(\boldsymbol{\beta}_1^0) + \mathbf{D}'_{21}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)\boldsymbol{\Lambda}_2^{-1}\mathbf{D}_{21}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0) \\ & - \mathbf{D}'_{21}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)\boldsymbol{\Lambda}_2^{-1}\mathbf{D}_{22}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)[\mathbf{D}'_{22}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)\boldsymbol{\Lambda}_2^{-1}\mathbf{D}_{22}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)]^{-1} \\ & \quad \times \mathbf{D}'_{22}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)\boldsymbol{\Lambda}_2^{-1}\mathbf{D}_{21}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)\}^{-1}. \end{aligned}$$

Since both asymptotic covariance matrices are positive definite, they will be equal if and only if the matrix

$$\begin{aligned} & \mathbf{D}'_{21}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)\boldsymbol{\Lambda}_2^{-1}\mathbf{D}_{21}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0) - \mathbf{D}'_{21}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)\boldsymbol{\Lambda}_2^{-1}\mathbf{D}_{22}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0) \\ & \times [\mathbf{D}'_{22}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)\boldsymbol{\Lambda}_2^{-1}\mathbf{D}_{22}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)]^{-1}\mathbf{D}'_{22}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)\boldsymbol{\Lambda}_2^{-1}\mathbf{D}_{21}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0) \end{aligned}$$

is 0. But since we can interpret this matrix as the residual variance in the asymptotic least squares projection of $\mathbf{D}'_{21}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)\boldsymbol{\Lambda}_2^{-1}\sqrt{T}\mathbf{P}'_2\bar{\mathbf{g}}_{2T}(\mathbf{0}, \boldsymbol{\beta}_1^0, \mathbf{0}, \boldsymbol{\beta}_2^0)$ onto $\mathbf{D}'_{22}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)\boldsymbol{\Lambda}_2^{-1}\sqrt{T}\mathbf{P}'_2\bar{\mathbf{g}}_{2T}(\mathbf{0}, \boldsymbol{\beta}_1^0, \mathbf{0}, \boldsymbol{\beta}_2^0)$, it will be zero if and only if we can write $\mathbf{D}_{21}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)$ as a linear combination of $\mathbf{D}_{22}(\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0)$. \square

C.2 Computation of CU-GMM

In this appendix we suggest an intuitive method that simplifies the computation of continuously updated GMM estimators in standard (i.e non-singular) situations, as well as in the singular case characterised by condition (7), which is the relevant one for spanning tests.

C.2.1 Uncentred CU-GMM and OLS output

We begin by generalising the minmax interpretation of the CU-GMM estimator in an IV context (see for example Arellano (2002), who traces it back to Sargan's work). Let $\mathbf{H}(\boldsymbol{\theta})$ define the $T \times n$ matrix that contains the sample values of the influence functions evaluated at some parameter values $\boldsymbol{\theta}$.

If we do not demean the influence functions, the HAC estimators of the long-run covariance matrix can be expressed as

$$\mathbb{S}_T(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^T w_{tj} \mathbf{h}(\mathbf{x}_t; \boldsymbol{\theta}) \mathbf{h}'(\mathbf{x}_j; \boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta})' \mathbf{W} \mathbf{H}(\boldsymbol{\theta}),$$

where \mathbf{W} is a $T \times T$ real matrix. For simplicity, we focus on those weighting schemes that guarantee the positive definiteness of \mathbf{W} . For instance, the usual Newey-West triangular scheme

with one lag yields

$$\mathbf{W} = \frac{1}{4} \begin{pmatrix} 1 & 0.5 & 0 & 0 \\ 0.5 & 1 & 0.5 & 0 \\ 0 & 0.5 & 1 & 0.5 \\ 0 & 0 & 0.5 & 1 \end{pmatrix}$$

when $T = 4$. If we do not want to consider lags in the HAC estimator then we simply work with $\mathbf{W} = T^{-1}\mathbf{I}_T$. Let us express the Cholesky decomposition of \mathbf{W} as

$$\mathbf{W} = \mathbf{W}_L \mathbf{W}'_L$$

where \mathbf{W}_L is lower triangular. The optimal CU-GMM estimator can be computed as a minmax criterion based on a certain R^2 :

Lemma C2 *Let*

$$\begin{aligned} \tilde{\boldsymbol{\theta}}_T &= \min_{\boldsymbol{\theta}} \tilde{\mathbb{J}}_T(\boldsymbol{\theta}), \\ \tilde{\mathbb{J}}_T(\boldsymbol{\theta}) &= \bar{\mathbf{h}}'_T(\boldsymbol{\theta}) \mathbb{S}_T^-(\boldsymbol{\theta}) \bar{\mathbf{h}}_T(\boldsymbol{\theta}). \end{aligned}$$

Then

$$\begin{aligned} \tilde{\boldsymbol{\theta}}_T &= \min_{\boldsymbol{\theta}} R^2(\boldsymbol{\theta}), \\ R^2(\boldsymbol{\theta}) &= \max_{\boldsymbol{\lambda}} \left[1 - \frac{(\mathbf{W}_L^{-1} \ell_T - \mathbf{W}'_L \mathbf{H}(\boldsymbol{\theta}) \boldsymbol{\lambda})' (\mathbf{W}_L^{-1} \ell_T - \mathbf{W}'_L \mathbf{H}(\boldsymbol{\theta}) \boldsymbol{\lambda})}{\ell'_T \mathbf{W}^{-1} \ell_T} \right] \end{aligned}$$

irrespective of the generalised inverse used.

Proof. We can understand the CU-GMM criterion as the explained variation in the OLS regression of $\mathbf{W}_L^{-1} \ell_T$ onto $\mathbf{W}'_L \mathbf{H}(\boldsymbol{\theta})$ because

$$\begin{aligned} \tilde{\mathbb{J}}_T(\boldsymbol{\theta}) &= \bar{\mathbf{h}}'_T(\boldsymbol{\theta}) [\mathbf{H}(\boldsymbol{\theta})' \mathbf{W} \mathbf{H}(\boldsymbol{\theta})]^{-1} \bar{\mathbf{h}}_T(\boldsymbol{\theta}) = \\ &= [T^{-1} \mathbf{H}(\boldsymbol{\theta})' \mathbf{W}_L \mathbf{W}_L^{-1} \ell_T]' [\mathbf{H}'(\boldsymbol{\theta}) \mathbf{W}_L \mathbf{W}'_L \mathbf{H}(\boldsymbol{\theta})]^{-1} [T^{-1} \mathbf{H}'(\boldsymbol{\theta}) \mathbf{W}_L \mathbf{W}_L^{-1} \ell_T]. \end{aligned}$$

The fact that the matrix $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is independent of the chosen inverse confirms that this CU-GMM criterion is independent of the particular generalised inverse chosen. \square

Note that if condition (7) in Section 3 holds, then $\mathbf{H}(\boldsymbol{\theta})$ will be of reduced column rank too. Specifically, $\mathbf{H}(\boldsymbol{\theta}) \boldsymbol{\Pi}(\boldsymbol{\theta}) = \mathbf{0}$ whenever $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$, where $\boldsymbol{\Pi}(\boldsymbol{\theta})$ and $\mathbf{m}(\boldsymbol{\theta})$ are defined in Assumption 1. The OLS interpretation of the CU-GMM criterion is very useful for the computation of the CU-GMM estimator in this context because the $R^2(\boldsymbol{\theta})$ is uniquely defined. Nevertheless, we should compute the necessary quantities with an OLS routine that is robust to multicollinearity.

For ease of exposition, we have not made explicit the equality restriction $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}$ in the previous expressions. Using the notation in the proof of Proposition 1, we can work with the alternative k parameters $\boldsymbol{\alpha}$ ($s \times 1$) and $\boldsymbol{\beta}$ ($(k - s) \times 1$)

$$\begin{pmatrix} \boldsymbol{\alpha}' & \boldsymbol{\beta}' \end{pmatrix} = \mathbf{M}'(\boldsymbol{\theta}),$$

where the first s entries are defined by $\boldsymbol{\alpha} = \mathbf{m}(\boldsymbol{\theta})$. We can impose the parametric restrictions $\mathbf{m}(\boldsymbol{\theta}) = \boldsymbol{\alpha} = \mathbf{0}$ by working with the smaller vector of parameters $\boldsymbol{\beta}$, so that the influence functions become $\mathbf{h}[\mathbf{x}_t, \mathbf{l}(\mathbf{0}, \boldsymbol{\beta})]$.

C.2.2 Centred CU-GMM and OLS output

Let us consider the same context of the previous section, but this time we demean the moment conditions before using the HAC estimator

$$\mathfrak{S}_T(\boldsymbol{\theta}) = \mathbf{U}'(\boldsymbol{\theta}) \mathbf{W} \mathbf{U}(\boldsymbol{\theta})$$

where $\mathbf{U}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta}) - \ell_T \bar{\mathbf{h}}_T'(\boldsymbol{\theta})$. If we define

$$\tilde{\mathbf{h}}_T(\boldsymbol{\theta}) = \mathbf{H}'(\boldsymbol{\theta}) \mathbf{W} \ell_T$$

then we can obtain $\mathfrak{S}_T(\boldsymbol{\theta})$ as follows

$$\mathbf{U}'(\boldsymbol{\theta}) \mathbf{W} \mathbf{U}(\boldsymbol{\theta}) = \mathbf{H}'(\boldsymbol{\theta}) \mathbf{W} \mathbf{H}(\boldsymbol{\theta}) - \tilde{\mathbf{h}}_T(\boldsymbol{\theta}) \bar{\mathbf{h}}_T'(\boldsymbol{\theta}) - \bar{\mathbf{h}}_T(\boldsymbol{\theta}) \tilde{\mathbf{h}}_T'(\boldsymbol{\theta}) + k \bar{\mathbf{h}}_T(\boldsymbol{\theta}) \bar{\mathbf{h}}_T'(\boldsymbol{\theta}),$$

where

$$k = \ell_T' \mathbf{W} \ell_T.$$

Let us define $\phi_1(\boldsymbol{\theta})$ as the solution to

$$[\mathbf{H}'(\boldsymbol{\theta}) \mathbf{W} \mathbf{H}(\boldsymbol{\theta})] \phi_1(\boldsymbol{\theta}) = \bar{\mathbf{h}}_T(\boldsymbol{\theta})$$

and $\phi_2(\boldsymbol{\theta})$ as the solution to

$$[\mathbf{H}'(\boldsymbol{\theta}) \mathbf{W} \mathbf{H}(\boldsymbol{\theta})] \phi_2(\boldsymbol{\theta}) = \tilde{\mathbf{h}}_T(\boldsymbol{\theta}).$$

Finally, let

$$K_T(\boldsymbol{\theta}) = \tilde{\mathbf{h}}_T'(\boldsymbol{\theta}) \phi_2(\boldsymbol{\theta}) = \tilde{\mathbf{h}}_T'(\boldsymbol{\theta}) [\mathbf{H}'(\boldsymbol{\theta}) \mathbf{W} \mathbf{H}(\boldsymbol{\theta})]^{-1} \tilde{\mathbf{h}}_T(\boldsymbol{\theta}),$$

$$\mathbb{K}_T(\boldsymbol{\theta}) = \tilde{\mathbf{h}}_T'(\boldsymbol{\theta}) \phi_1(\boldsymbol{\theta}) = \bar{\mathbf{h}}_T'(\boldsymbol{\theta}) \phi_2(\boldsymbol{\theta}) = \bar{\mathbf{h}}_T'(\boldsymbol{\theta}) [\mathbf{H}'(\boldsymbol{\theta}) \mathbf{W} \mathbf{H}(\boldsymbol{\theta})]^{-1} \tilde{\mathbf{h}}_T(\boldsymbol{\theta}),$$

and note that $\tilde{\mathbb{J}}_T(\boldsymbol{\theta}) = \bar{\mathbf{h}}_T'(\boldsymbol{\theta}) \phi_1(\boldsymbol{\theta}) = \bar{\mathbf{h}}_T'(\boldsymbol{\theta}) [\mathbf{H}'(\boldsymbol{\theta}) \mathbf{W} \mathbf{H}(\boldsymbol{\theta})]^{-1} \bar{\mathbf{h}}_T(\boldsymbol{\theta})$. Using this notation, we can once more relate a CU-GMM criterion function to certain OLS quantities:

Lemma C3 The CU-GMM criterion function based on $\mathfrak{S}_T(\boldsymbol{\theta})$ can be expressed as

$$\tilde{\mathfrak{J}}_T(\boldsymbol{\theta}) = \bar{\mathbf{h}}_T'(\boldsymbol{\theta}) \mathfrak{S}_T^-(\boldsymbol{\theta}) \bar{\mathbf{h}}_T(\boldsymbol{\theta}) = \frac{\tilde{\mathfrak{J}}_T(\boldsymbol{\theta})}{(k - K_T(\boldsymbol{\theta})) \tilde{\mathfrak{J}}_T(\boldsymbol{\theta}) + (1 - \mathbb{K}_T(\boldsymbol{\theta}))^2}$$

irrespective of the generalised inverse used.

Proof. It is easy to see that

$$\phi_3(\boldsymbol{\theta}) = \frac{\tilde{\mathfrak{J}}_T(\boldsymbol{\theta})}{(k - K_T(\boldsymbol{\theta})) \tilde{\mathfrak{J}}_T(\boldsymbol{\theta}) + (1 - \mathbb{K}_T(\boldsymbol{\theta}))^2} \left[\frac{1 - \mathbb{K}_T(\boldsymbol{\theta})}{\tilde{\mathfrak{J}}_T(\boldsymbol{\theta})} \phi_1(\boldsymbol{\theta}) + \phi_2(\boldsymbol{\theta}) \right]$$

solves

$$[\mathbf{U}'(\boldsymbol{\theta}) \mathbf{W} \mathbf{U}(\boldsymbol{\theta})] \phi_3(\boldsymbol{\theta}) = \bar{\mathbf{h}}_T(\boldsymbol{\theta}),$$

which means that although we cannot interpret $\phi_3(\boldsymbol{\theta})$ directly as a OLS slope, we can link it to the OLS slopes $\phi_1(\boldsymbol{\theta})$ and $\phi_2(\boldsymbol{\theta})$. On this basis, we can express the centred CU-GMM criterion as

$$\begin{aligned} \tilde{\mathfrak{J}}_T(\boldsymbol{\theta}) &= \bar{\mathbf{h}}_T'(\boldsymbol{\theta}) \mathfrak{S}_T^-(\boldsymbol{\theta}) \bar{\mathbf{h}}_T(\boldsymbol{\theta}) = \bar{\mathbf{h}}_T'(\boldsymbol{\theta}) [\mathbf{U}'(\boldsymbol{\theta}) \mathbf{W} \mathbf{U}(\boldsymbol{\theta})]^{-1} \bar{\mathbf{h}}_T(\boldsymbol{\theta}) \\ &= \bar{\mathbf{h}}_T'(\boldsymbol{\theta}) \phi_3(\boldsymbol{\theta}) = \frac{\tilde{\mathfrak{J}}_T(\boldsymbol{\theta})}{(k - K_T(\boldsymbol{\theta})) \tilde{\mathfrak{J}}_T(\boldsymbol{\theta}) + (1 - \mathbb{K}_T(\boldsymbol{\theta}))^2}. \end{aligned}$$

Again, $\tilde{\mathfrak{J}}_T(\boldsymbol{\theta})$, $K_T(\boldsymbol{\theta})$, and $\mathbb{K}_T(\boldsymbol{\theta})$ are constructed from a matrix that can be expressed as $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, which confirms the invariance of the CU-GMM criterion $\tilde{\mathfrak{J}}_T(\boldsymbol{\theta})$ to the choice of generalised inverse. \square

Unlike in the uncentred case, two different projections are required to interpret the centred CU-GMM criterion in terms of OLS. The first one regresses $\mathbf{W}_L^{-1}\ell_T$ onto $\mathbf{W}_L'\mathbf{H}(\boldsymbol{\theta})$, but now we also need to regress $\mathbf{W}_L'\ell_T$ onto $\mathbf{W}_L'\mathbf{H}(\boldsymbol{\theta})$. Once again, this OLS interpretation is very useful for the computation of the CU-GMM estimator, especially in the singular case.

C.2.3 Relationship between centred and uncentred criteria

In principle there is not a systematic ordering of the two criteria $\tilde{\mathfrak{J}}_T(\boldsymbol{\theta})$ and $\tilde{\mathfrak{J}}_T(\boldsymbol{\theta})$, which will generally lead to different estimators. However, in the *i.i.d.* case Newey and Smith (2004) show that $\tilde{\mathfrak{J}}_T(\boldsymbol{\theta})$ is a monotonic transformation of $\tilde{\mathfrak{J}}_T(\boldsymbol{\theta})$, so that both criteria will yield the same estimator of $\boldsymbol{\theta}$. Specifically,

$$\begin{aligned} \tilde{\mathfrak{J}}_T(\boldsymbol{\theta}) &= \bar{\mathbf{h}}'(\boldsymbol{\theta}) [T^{-1}\mathbf{H}(\boldsymbol{\theta})'\mathbf{H}(\boldsymbol{\theta})]^{-1} \bar{\mathbf{h}}(\boldsymbol{\theta}) \\ &\leq \tilde{\mathfrak{J}}_T(\boldsymbol{\theta}) = \bar{\mathbf{h}}'(\boldsymbol{\theta}) [T^{-1}\mathbf{U}'(\boldsymbol{\theta})\mathbf{U}(\boldsymbol{\theta})]^{-1} \bar{\mathbf{h}}(\boldsymbol{\theta}) = \frac{\tilde{\mathfrak{J}}_T(\boldsymbol{\theta})}{1 - \tilde{\mathfrak{J}}_T(\boldsymbol{\theta})}, \end{aligned}$$

with $\tilde{\mathbb{J}}_T(\boldsymbol{\theta}) \in [0, 1]$.

Finally, we can formally link our previous results to the Riesz representation theorem for Hilbert spaces that we used in Section 2. In particular, we can interpret

$$q^+(\boldsymbol{\theta}) = \mathbf{h}'(\boldsymbol{\theta}) [T^{-1} \mathbf{H}'(\boldsymbol{\theta}) \mathbf{H}(\boldsymbol{\theta})]^{-1} \bar{\mathbf{h}}_T(\boldsymbol{\theta})$$

as the “uncentred mean representing influence function” and

$$q^{++}(\boldsymbol{\theta}) = \mathbf{h}'(\boldsymbol{\theta}) [T^{-1} \mathbf{U}'(\boldsymbol{\theta}) \mathbf{U}(\boldsymbol{\theta})]^{-1} \bar{\mathbf{h}}_T(\boldsymbol{\theta})$$

as the “centred mean representing influence function”. The results for representing portfolios presented in Section 2 can then be used to link the two as

$$q^{++}(\boldsymbol{\theta}) = \{1 - E_T[q^+(\boldsymbol{\theta})]\}^{-1} q^+(\boldsymbol{\theta}),$$

and to understand the cases where these expressions are not well defined.

D Covariance matrices of the sample moment conditions under i.i.d. elliptical returns

Elliptical distributions are usually defined by means of the affine transformation $\mathbf{R}_t = \boldsymbol{\nu}^0 + (\boldsymbol{\Sigma}^0)^{1/2} \boldsymbol{\varepsilon}_t^\circ$, where $\boldsymbol{\varepsilon}_t^\circ$ is a spherically symmetric random vector of dimension N , which in turn is fully characterised in Theorem 2.5 (iii) of Fang, Kotz and Ng (1990) as $\boldsymbol{\varepsilon}_t^\circ = e_t \mathbf{u}_t$, where \mathbf{u}_t is uniformly distributed on the unit sphere surface in \mathbb{R}^N , and e_t is a non-negative random variable which is independent of \mathbf{u}_t . The variables e_t and \mathbf{u}_t are referred to as the generating variate and the uniform base of the spherical distribution. Assuming that $E(e_t^2) < \infty$, we can standardise $\boldsymbol{\varepsilon}_t^\circ$ by setting $E(e_t^2) = N$, so that $E(\boldsymbol{\varepsilon}_t^\circ) = \mathbf{0}$, $V(\boldsymbol{\varepsilon}_t^\circ) = \mathbf{I}_N$, $E(\mathbf{R}_t) = \boldsymbol{\nu}^0$ and $V(\mathbf{R}_t) = \boldsymbol{\Sigma}^0$. For instance, if $e_t = \sqrt{(v^0 - 2)\zeta_t/\xi_t}$, ζ_t is a chi-square random variable with N degrees of freedom, and ξ_t is an independent Gamma variate with mean $v^0 > 2$ and variance $2v^0$, then $\boldsymbol{\varepsilon}_t^\circ$ will be distributed as a standardised multivariate Student t random vector of dimension N with v^0 degrees of freedom, which converges to a standardised multivariate normal as $v^0 \rightarrow \infty$. If we further assume that $E(e_t^4) < \infty$, then the coefficient of multivariate excess kurtosis κ reduces to $E(e_t^4)/N(N+2) - 1$. For instance, $\kappa = 2/(v^0 - 4)$ in the Student t case, and $\kappa = 0$ under normality. In this respect, note that since $E(e_t^4) \geq E^2(e_t^2) = N^2$ by the Cauchy-Schwarz inequality, with equality if and only if $e_t = \sqrt{N}$ so that $\boldsymbol{\varepsilon}_t^\circ$ is proportional to \mathbf{u}_t , then $\kappa \geq -2/(N+2)$, the minimum value being achieved in the uniformly distributed case.

Then, it is easy to combine the representation of elliptical distributions above with the higher order moments of a multivariate normal vector in Balestra and Holly (1990) to prove that the third and fourth moments of the elliptical distribution are given by

$$E(\mathbf{R}\mathbf{R}' \otimes \mathbf{R}) = (\mathbf{I}_{N^2} + \mathbf{K}_{NN}) (\boldsymbol{\nu} \otimes \boldsymbol{\Sigma}) + \text{vec}(\boldsymbol{\Gamma}) \boldsymbol{\nu}', \quad (\text{D1})$$

and

$$\begin{aligned} E(\mathbf{R}\mathbf{R}' \otimes \mathbf{R}\mathbf{R}') &= (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) (\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + (\mathbf{I}_{N^2} + \mathbf{K}_{NN}) (\boldsymbol{\Sigma} \otimes \boldsymbol{\nu}\boldsymbol{\nu}') (\mathbf{I}_{N^2} + \mathbf{K}_{NN}) \\ &+ \text{vec}(\boldsymbol{\Gamma}) \text{vec}(\boldsymbol{\Gamma})' + \kappa [(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) (\mathbf{I}_{N^2} + \mathbf{K}_{NN}) + \text{vec}(\boldsymbol{\Sigma}) \text{vec}(\boldsymbol{\Sigma})'], \end{aligned} \quad (\text{D2})$$

respectively, where \mathbf{K}_{NN} is the commutation matrix studied in Magnus and Neudecker (1988). Similarly, it is possible to show that the mean vector and covariance matrix of the distribution of \mathbf{R}_2 conditional on \mathbf{R}_1 will be $E(\mathbf{R}_2|\mathbf{R}_1) = \boldsymbol{\nu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_1 - \boldsymbol{\nu}_1)$ and $V(\mathbf{R}_2|\mathbf{R}_1) = \varrho[(\mathbf{R}_1 - \boldsymbol{\nu}_1)'\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_1 - \boldsymbol{\nu}_1)] \cdot \boldsymbol{\Omega}$, where $\boldsymbol{\Omega} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}'_{21}$, and $\varrho(\cdot)$ is a scalar function whose form depends on the member of the elliptical class (see again Fang, Kotz and Ng (1990)). For instance, $\varrho(\cdot)$ is identically 1 in the multivariate normal case, and affine in its argument for the Student t (see Zellner (1971, pp. 383-389)).

The following three results exploit these properties to obtain closed form expressions for the asymptotic covariance matrices of the sample moment conditions that appear in the different testing procedures. For the sake of generality, we derive them without imposing the null of spanning, but this can be easily modified by using the relevant parametric restrictions. The lemmas use the following notation: $A = \text{cov}(p^{**}, p^{++}) = \boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}\ell_N$, $B = V(p^{++}) = \boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu}$, and $C = V(p^{**}) = \ell'_N\boldsymbol{\Sigma}^{-1}\ell_N$.

Lemma D1 *If \mathbf{R}_t is an i.i.d. elliptical random vector with mean $\boldsymbol{\nu}$, covariance matrix $\boldsymbol{\Sigma}$, and bounded fourth moments, then the asymptotic covariance matrix of*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \begin{pmatrix} \mathbf{R}_t\mathbf{R}'_t\phi^+ - \mathbf{R}_t \\ \mathbf{R}_t\mathbf{R}'_t\phi^* - \ell_N \end{pmatrix}$$

will be

$$\begin{aligned}
& \left\{ \begin{array}{c} (1+B)^{-1} [1 + \kappa B (1+B)^{-1}] \\ A (1+B)^{-1} \kappa \\ A (1+B)^{-2} \kappa \end{array} \right\} \otimes \mathbf{\Gamma} \\
& [C - A^2 (1+B)^{-1}] + \kappa \left([C - A^2 (1+B)^{-1}] - A^2 (1+B)^{-2} \right) \\
& + \left\{ \begin{array}{c} -2(1+B)^{-2} + \left(3B^2 (1+B)^{-2} - 5B (1+B)^{-1} + 2 \right) \kappa \\ A (1+B)^{-2} (2 - 3\kappa) \\ A (1+B)^{-2} (2 - 3\kappa) \\ -2A^2 (1+B)^{-2} + \{ 3A^2 (1+B)^{-2} - [C - A^2 (1+B)^{-1}] \} \kappa \end{array} \right\} \otimes \boldsymbol{\nu} \boldsymbol{\nu}' \\
& + \begin{pmatrix} 0 & 0 \\ 0 & 1 + 2\kappa \end{pmatrix} \otimes \ell_N \ell_N' + \begin{bmatrix} 0 & (1+B)^{-1} \kappa \\ (1+B)^{-1} \kappa & -2A (1+B)^{-1} \kappa \end{bmatrix} \otimes (\boldsymbol{\nu} \ell_N' + \ell_N \boldsymbol{\nu}').
\end{aligned}$$

Proof. Tedious but straightforward on the basis of (D1) and (D2). \square

Lemma D2 *If \mathbf{R}_t is an i.i.d. elliptical random vector with mean $\boldsymbol{\nu}$, covariance matrix $\boldsymbol{\Sigma}$, and bounded fourth moments, then the asymptotic covariance matrix of*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \begin{bmatrix} \mathbf{R}_t - \boldsymbol{\nu} \\ \mathbf{R}_t (\mathbf{R}_t - \boldsymbol{\nu})' \boldsymbol{\varphi}^+ - \mathbf{R}_t \\ \mathbf{R}_t (\mathbf{R}_t - \boldsymbol{\nu})' \boldsymbol{\varphi}^* - \ell_N \end{bmatrix}$$

will be

$$\begin{aligned}
& \begin{bmatrix} 1 & -1 & \mathbf{0} \\ -1 & 1 + (\kappa + 1)B & (\kappa + 1)A \\ \mathbf{0} & (\kappa + 1)A & (\kappa + 1)C \end{bmatrix} \otimes \boldsymbol{\Sigma} \\
& + \begin{bmatrix} \mathbf{0} & \boldsymbol{\nu} \boldsymbol{\nu}' & \ell_N \boldsymbol{\nu}' \\ \boldsymbol{\nu} \boldsymbol{\nu}' & (B + 2\kappa - 1) \boldsymbol{\nu} \boldsymbol{\nu}' & A \boldsymbol{\nu} \boldsymbol{\nu}' + \kappa (\ell_N \boldsymbol{\nu}' + \boldsymbol{\nu} \ell_N') \\ \boldsymbol{\nu} \ell_N' & A \boldsymbol{\nu} \boldsymbol{\nu}' + \kappa (\ell_N \boldsymbol{\nu}' + \boldsymbol{\nu} \ell_N') & C \boldsymbol{\nu} \boldsymbol{\nu}' + (2\kappa + 1) \ell_N \ell_N' \end{bmatrix}.
\end{aligned}$$

Proof. Tedious but straightforward on the basis of (D1) and (D2). \square

Lemma D3 *If \mathbf{R}_t is an i.i.d. elliptical random vector with mean $\boldsymbol{\nu}$, covariance matrix $\boldsymbol{\Sigma}$, and bounded fourth moments, then the asymptotic covariance matrix of*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \left[\begin{pmatrix} 1 \\ \mathbf{R}_{1t} \end{pmatrix} \otimes (\mathbf{R}_{2t} - \mathbf{a} - B \mathbf{R}_{1t}) \right]$$

will be

$$\begin{pmatrix} 1 & \boldsymbol{\nu}' \\ \boldsymbol{\nu}_1 & (\kappa + 1) \boldsymbol{\Sigma}_{11} + \boldsymbol{\nu}_1 \boldsymbol{\nu}' \end{pmatrix} \otimes \boldsymbol{\Omega}.$$

Proof. First of all, we can apply the law of iterated expectations to show that

$$\begin{aligned}
& E\{E[\mathbf{h} \mathbf{h}' | \mathbf{R}_1]\} \\
& = \begin{bmatrix} E\{\varrho[(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{R}_1 - \boldsymbol{\nu}_1)]\} & E\{\varrho[(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{R}_1 - \boldsymbol{\nu}_1)] \mathbf{R}_1'\} \\ E\{\varrho[(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{R}_1 - \boldsymbol{\nu}_1)] \mathbf{R}_1\} & E\{\varrho[(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{R}_1 - \boldsymbol{\nu}_1)] \mathbf{R}_1 \mathbf{R}_1'\} \end{bmatrix} \otimes \boldsymbol{\Omega},
\end{aligned}$$

for

$$\mathbf{h} = \begin{pmatrix} 1 \\ \mathbf{R}_1 \end{pmatrix} \otimes (\mathbf{R}_2 - \mathbf{a} - \mathbf{B}\mathbf{R}_1).$$

But since

$$\begin{aligned} V(\mathbf{R}_{2t}) &= \boldsymbol{\Sigma}_{22} = E[V(\mathbf{R}_2|\mathbf{R}_1)] + V[E(\mathbf{R}_2|\mathbf{R}_1)] \\ &= E\{\varrho[(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_1 - \boldsymbol{\nu}_1)]\} \cdot \boldsymbol{\Omega} + \mathbf{B}\boldsymbol{\Sigma}_{11}\mathbf{B}', \end{aligned}$$

then it must be the case that $E\{\varrho[(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_1 - \boldsymbol{\nu}_1)]\} = 1$. Similarly, since

$$\begin{aligned} E\{(\mathbf{R}_1 - \boldsymbol{\nu}_1) \cdot \text{vec}'[(\mathbf{R}_2 - \boldsymbol{\nu}_2)(\mathbf{R}_2 - \boldsymbol{\nu}_2)']\} &= E[(\mathbf{R}_1 - \boldsymbol{\nu}_1) \cdot E\{\text{vec}'[(\mathbf{R}_2 - \boldsymbol{\nu}_2)(\mathbf{R}_2 - \boldsymbol{\nu}_2)'] | \mathbf{R}_1\}] \\ &= E[(\mathbf{R}_1 - \boldsymbol{\nu}_1) \cdot \text{vec}\{\varrho[(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_1 - \boldsymbol{\nu}_1)] \cdot \boldsymbol{\Omega} + \mathbf{B}(\mathbf{R}_1 - \boldsymbol{\nu}_1)(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \mathbf{B}'\}] = \mathbf{0} \end{aligned}$$

by the symmetry of elliptical random vectors, it must also be the case that

$$E\{\varrho[(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_1 - \boldsymbol{\nu}_1)](\mathbf{R}_1 - \boldsymbol{\nu}_1)\} = \mathbf{0},$$

and consequently, that $E[\varrho[(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_1 - \boldsymbol{\nu}_1)]\mathbf{R}_1] = \boldsymbol{\nu}_1$.

Finally, since

$$\begin{aligned} &E[(\mathbf{R}_2 - \boldsymbol{\nu}_2)(\mathbf{R}_2 - \boldsymbol{\nu}_2)' \otimes (\mathbf{R}_1 - \boldsymbol{\nu}_1)(\mathbf{R}_1 - \boldsymbol{\nu}_1)'] \\ &= (\kappa + 1) [(\boldsymbol{\Sigma}_{22} \otimes \boldsymbol{\Sigma}_{11}) + (\boldsymbol{\Sigma}_{21} \otimes \boldsymbol{\Sigma}_{12}) \mathbf{K}_{N_1, N_2} + \text{vec}(\boldsymbol{\Sigma}_{12}) \text{vec}(\boldsymbol{\Sigma}_{12})'] \\ &= E\{E[(\mathbf{R}_2 - \boldsymbol{\nu}_2)(\mathbf{R}_2 - \boldsymbol{\nu}_2)' | \mathbf{R}_1] \otimes (\mathbf{R}_1 - \boldsymbol{\nu}_1)(\mathbf{R}_1 - \boldsymbol{\nu}_1)'\} \\ &= E\{\{\varrho[(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_1 - \boldsymbol{\nu}_1)] \cdot \boldsymbol{\Omega} \\ &\quad + \mathbf{B}(\mathbf{R}_1 - \boldsymbol{\nu}_1)(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \mathbf{B}'\} \otimes (\mathbf{R}_1 - \boldsymbol{\nu}_1)(\mathbf{R}_1 - \boldsymbol{\nu}_1)'\} \\ &= \boldsymbol{\Omega} \otimes E\{\varrho[(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_1 - \boldsymbol{\nu}_1)] \cdot (\mathbf{R}_1 - \boldsymbol{\nu}_1)(\mathbf{R}_1 - \boldsymbol{\nu}_1)'\} \\ &+ (\mathbf{B} \otimes \mathbf{I}_{N_1}) E[(\mathbf{R}_1 - \boldsymbol{\nu}_1)(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \otimes (\mathbf{R}_1 - \boldsymbol{\nu}_1)(\mathbf{R}_1 - \boldsymbol{\nu}_1)'] (\mathbf{B}' \otimes \mathbf{I}_{N_1}) \\ &= \boldsymbol{\Omega} \otimes E\{\varrho[(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_1 - \boldsymbol{\nu}_1)] \cdot (\mathbf{R}_1 - \boldsymbol{\nu}_1)(\mathbf{R}_1 - \boldsymbol{\nu}_1)'\} \\ &+ (\kappa + 1) [(\mathbf{B}\boldsymbol{\Sigma}_{12} \otimes \boldsymbol{\Sigma}_{11}) + (\boldsymbol{\Sigma}_{21} \otimes \boldsymbol{\Sigma}_{12}) \mathbf{K}_{N_1, N_2} + \text{vec}(\boldsymbol{\Sigma}_{12}) \text{vec}(\boldsymbol{\Sigma}_{12})'], \end{aligned}$$

where we have repeatedly used expression (D2) for the fourth moments of an elliptical vector,

and the fact that

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{I}_{N_1} & \mathbf{0} \\ \mathbf{B} & \mathbf{I}_{N_2} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{N_1} & \mathbf{B}' \\ \mathbf{0} & \mathbf{I}_{N_2} \end{pmatrix},$$

it must be the case that $E\{\varrho[(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_1 - \boldsymbol{\nu}_1)] \cdot (\mathbf{R}_1 - \boldsymbol{\nu}_1)(\mathbf{R}_1 - \boldsymbol{\nu}_1)'\} = (\kappa + 1)\boldsymbol{\Sigma}_{11}$, and

consequently, that $E\{\varrho[(\mathbf{R}_1 - \boldsymbol{\nu}_1)' \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_1 - \boldsymbol{\nu}_1)]\mathbf{R}_1\mathbf{R}_1'\} = (\kappa + 1)\boldsymbol{\Sigma}_{11} + \boldsymbol{\nu}_1\boldsymbol{\nu}_1'$. \square

References

- Amengual, D. and E. Sentana (2010): “A comparison of mean-variance efficiency tests”, *Journal of Econometrics* 154, 16-34.
- Andrews, D.W.K. (1987): “Asymptotic results for generalized Wald tests”, *Econometric Theory* 3, 348-358.
- Antoine, A., H. Bonnal and E. Renault (2007): “On the efficient use of the informational content of estimating equations: implied probabilities and Euclidean empirical likelihood”, *Journal of Econometrics* 138, 461-487.
- Arellano, M. (2002): “Sargan’s Instrumental Variable estimation and the Generalized Method of Moments”, *Journal of Business and Economic Statistics* 20, 450-459.
- Arellano, M., L-P. Hansen and E. Sentana (2009): “Underidentification?”, CEMFI Working Paper 0905.
- Balestra, P. and A. Holly (1990): “A general Kronecker formula for the moments of the multivariate normal distribution”, Cahier No. 9002, DEEP Université de Lausanne.
- Bekaert, B., and M.S. Urias (1996): “Diversification, integration and emerging market closed-end funds”, *Journal of Finance* 51, 835-869.
- Breusch, T., H. Qian, P. Schmidt and D. Wyhowski (1999): “Redundancy of moment conditions”, *Journal of Econometrics* 91, 89-111.
- Campbell, J.Y., A.W. Lo, and A.C. MacKinlay (1997): *The Econometrics of Financial Markets*, Princeton University Press.
- Chamberlain, G. (1983): “A characterization of the distributions that imply mean-variance utility functions”, *Journal of Economic Theory* 29, 185-201.
- Chamberlain, G. and M. Rothschild (1983): “Arbitrage, factor structure, and mean-variance analysis on large asset markets”, *Econometrica* 51, 1281-1304.
- Chen, X. (2005): “Large sample sieve estimation of semi-nonparametric models”, forthcoming in J. Heckman and E. Leamer, *Handbook of Econometrics* Vol. VI, Elsevier.
- Cochrane, J.H. (2001): *Asset Pricing*, Princeton University Press.
- de Jong, R.M. and J. Davidson (2000): “Consistency of kernel estimators of heteroskedasticity and autocorrelation covariance matrices”, *Econometrica* 68, 407-423.
- De Roon, F.A., and T.E. Nijman (2001): “Testing for mean-variance spanning: a survey”, *Journal of Empirical Finance* 8, 111-155.
- De Roon, F.A., T.E. Nijman, and B.J.M. Werker (1997): “Testing for spanning with future contracts and nontraded assets: a general approach”, Center Discussion Paper 9683, Tilburg University.

- De Roon, F.A., T.E. Nijman, and B.J.M. Werker (2000): “Testing for mean-variance spanning with short sales constraints and transaction costs: The case of emerging markets”, *Journal of Finance* 56, 721-742.
- De Santis, G. (1993): “Volatility bounds for stochastic discount factors: tests and implications from international financial markets”, Ph.D. dissertation, University of Chicago.
- De Santis, G. (1995): “Volatility bounds for stochastic discount factors: tests and implications from international financial markets”, mimeo, University of Southern California.
- Díez de los Ríos, A. and E. Sentana (2010): “Testing uncovered interest parity: a continuous time approach”, forthcoming in the *International Economic Review*.
- Dovonon, P. and E. Renault (2009): “GMM overidentification test with first order underidentification”, mimeo, University of North Carolina at Chapel Hill.
- Dufour, J.-M., and P. Valéry (2009): “GMM and hypothesis tests when rank conditions fail: a regularization approach”, mimeo, McGill University.
- Errunza, V., K. Hogan and M.-W. Hung (1999): “Can the gains from international diversification be achieved without trading abroad?”, *Journal of Finance* 54, 2075-2107.
- 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.
- Ferson, W.E. (1995): “Theory and empirical testing of asset pricing models”, in Jarrow, R.A., W.T. Ziemba and V. Maksimovic, *The Finance Handbook*, 145-200, North-Holland.
- Ferson, W.E., S.R. Foerster and D.B. Keim (1993): “General tests of latent variable models and mean-variance spanning,” *Journal of Finance* 48, 131-156.
- Fleming, W. (1977): *Functions of several variables*, Springer-Verlag.
- Gallant, A.R., L.P. Hansen, and G.E. Tauchen (1990), “Using conditional moments of asset payoffs to infer the volatility of intertemporal marginal rates of substitution” *Journal of Econometrics* 45, 141-180.
- Gibbons, M.R., S.A. Ross, and J. Shanken (1989): “A test of the efficiency of a given portfolio”, *Econometrica* 57, 1121-1152.
- Gouriéroux, C., A. Monfort, and E. Renault (1996): “Two-stage generalized moment method with applications to regressions with heteroscedasticity of unknown form”, *Journal of Statistical Planning and Inference* 50, 37-63.
- Gouriéroux, C. and A. Monfort (2005): “The econometrics of efficient portfolios”, *Journal of Empirical Finance* 12, 1-41.

- Hall, A.R. (2000) “Covariance matrix estimation and the power of the overidentifying restrictions test”, *Econometrica* 68, 1517-1527.
- Hansen, L.P. (1982): “Large sample properties of generalized method of moments estimators”, *Econometrica* 50, 1029-1054.
- Hansen, L.P., J. Heaton, and A. Yaron (1996): “Finite sample properties of some alternative GMM estimators”, *Journal of Business and Economic Statistics* 14, 262-280.
- Hansen, L.P., and R. Jagannathan (1991): “Implications of security market data for models of dynamic economies”, *Journal of Political Economy* 99, 225-262.
- Hansen, L.P. and S.F. Richard (1987): “The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models”, *Econometrica* 55, 587-613.
- Huberman, G., and S.A. Kandel (1987): “Mean-variance spanning”, *Journal of Finance* 42, 383-388.
- Hunter, D.M., and D.P. Simon (2005): “Are TIPS the “real” deal?: a conditional assessment of their role in a nominal portfolio”, *Journal of Banking and Finance* 29, 347-368.
- Jagannathan, R., and Z. Wang (2002): “Empirical evaluation of asset-pricing models: a comparison of the SDF and beta methods”, *Journal of Finance* 57, 2337-2367.
- Jobson, J. D. and B. Korkie (1982): “Potential performance and tests of portfolio efficiency”, *Journal of Financial Economics* 10, 433-466.
- Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T.C. Lee (1985): *The Theory and Practice of Econometrics*, 2nd ed., Wiley.
- Kan, R. and G. Zhou (2001): “Tests of mean-variance spanning”, mimeo, J.M. Olin School of Business, Washington University in St. Louis.
- Lezan, G. and E. Renault (1999): “Performance estimation with stochastic volatility of unknown form”, mimeo, CREST INSEE.
- Magnus, J. and H. Neudecker (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley.
- Mardia, K.V. (1970), “Measures of multivariate skewness and kurtosis with applications”, *Biometrika* 57, 519-530.
- Marín, L. (1996): “Efectos de la diversificación internacional sobre los límites de volatilidad de las tasas de descuento estocásticas”, unpublished Master’s thesis, CEMFI.
- Markowitz, H. (1952): “Portfolio selection”, *Journal of Finance* 7, 77-99.
- Newey, W.K. and D.L. McFadden (1994): “Large sample estimation and hypothesis testing”, in Engle, R.F. and D.L. McFadden, *Handbook of Econometrics* vol. IV, 2111-2245, Elsevier.
- Newey, W.K., and R.J. Smith (2004): “Higher order properties of GMM and Generalized Em-

- pirical Likelihood estimators”, *Econometrica* 72, 219-255.
- Owen, J., and R. Rabinovitch (1983): “On the class of elliptical distributions and their applications to the theory of portfolio choice”, *Journal of Finance* 58, 745-752.
- Palacios-Huerta, I. (2003): “An empirical analysis of the risk properties of human capital returns”, *American Economic Review* 93, 948-964.
- Peñaranda, F. (1999): “Contrastes de generación en la frontera Hansen-Jagannathan: Dualidad con la frontera media-varianza”, unpublished Master’s thesis, CEMFI.
- Peñaranda, F. and E. Sentana (2004): “Tangency tests in return and stochastic discount factor mean-variance frontiers: a unifying approach”, mimeo, CEMFI.
- Peñaranda, F., and E. Sentana (2008): “Duality in mean-variance frontiers with conditioning information”, revised version of CEPR DP 6566.
- Peñaranda, F., and E. Sentana (2010): “A unifying approach to the empirical evaluation of asset pricing models”, UPF WP 1229.
- Rao, C.R. and S.K. Mitra (1971): *Generalized Inverse of Matrices and its Applications*, Wiley.
- Ross, S.A. (1978): “Mutual fund separation in financial theory: the separating distributions”, *Journal of Economic Theory* 17, 254-286.
- Sargan, J.D. (1983): “Identification and lack of identification” *Econometrica* 51, 1605-1634.
- Snow, K.N. (1991): “Diagnosing asset pricing models using the distribution of asset returns”, *Journal of Finance* 46, 955-983.
- Stevenson, S. (2000): “International real estate diversification: empirical tests using hedged indices”, *Journal of Real Estate Research* 19, 105-131.
- Vorkink, K. (2003): “Return distributions and improved tests of asset pricing models”, *Review of Financial Studies* 16, 845-874.
- Zellner, A. (1971): *An Introduction to Bayesian Inference in Econometrics*, Wiley.

Table 1

Tests of spanning for the six size and book-to-market sorted portfolios by the two Fama-French portfolios and the market.

Monthly data from 1952 to 2007

Lags	SM, SH, BM	SL	SM	SH	BL	BM	BH
0	57.68 (.00)	57.24 (.00)	26.65 (.00)	5.76 (.06)	12.25 (.00)	7.12 (.03)	40.46 (.00)
5	26.18 (.00)	29.40 (.00)	17.27 (.00)	5.02 (.08)	9.53 (.01)	5.73 (.06)	22.24 (.00)
10	27.33 (.00)	20.30 (.00)	14.72 (.00)	4.65 (.10)	8.45 (.01)	4.87 (.09)	15.86 (.00)

Note: CU-GMM overidentification test statistics, with p-values in brackets. The number of lags refers to the window length used in the Newey-West estimator of the long-run covariance matrix of the moment conditions.

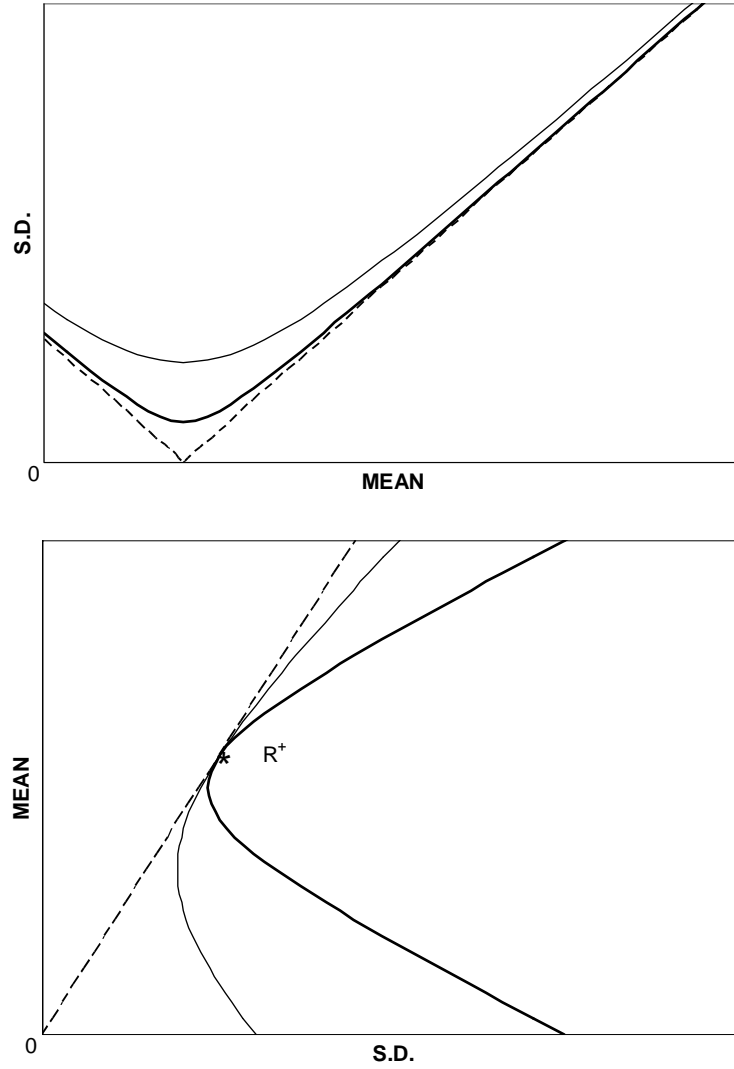


Figure 1: SDF and return frontiers such that \mathbf{R} and \mathbf{R}_1 share the mean representing portfolio.

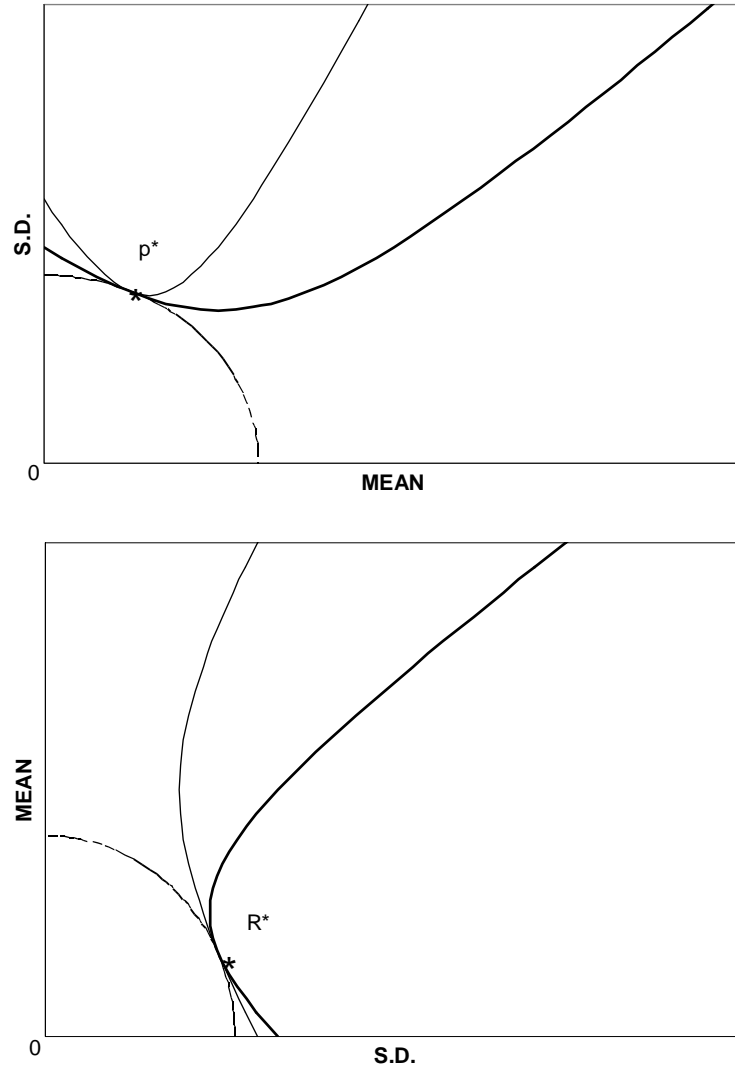


Figure 2: SDF and return frontiers such that \mathbf{R} and \mathbf{R}_1 share the uncentred cost representing portfolio.

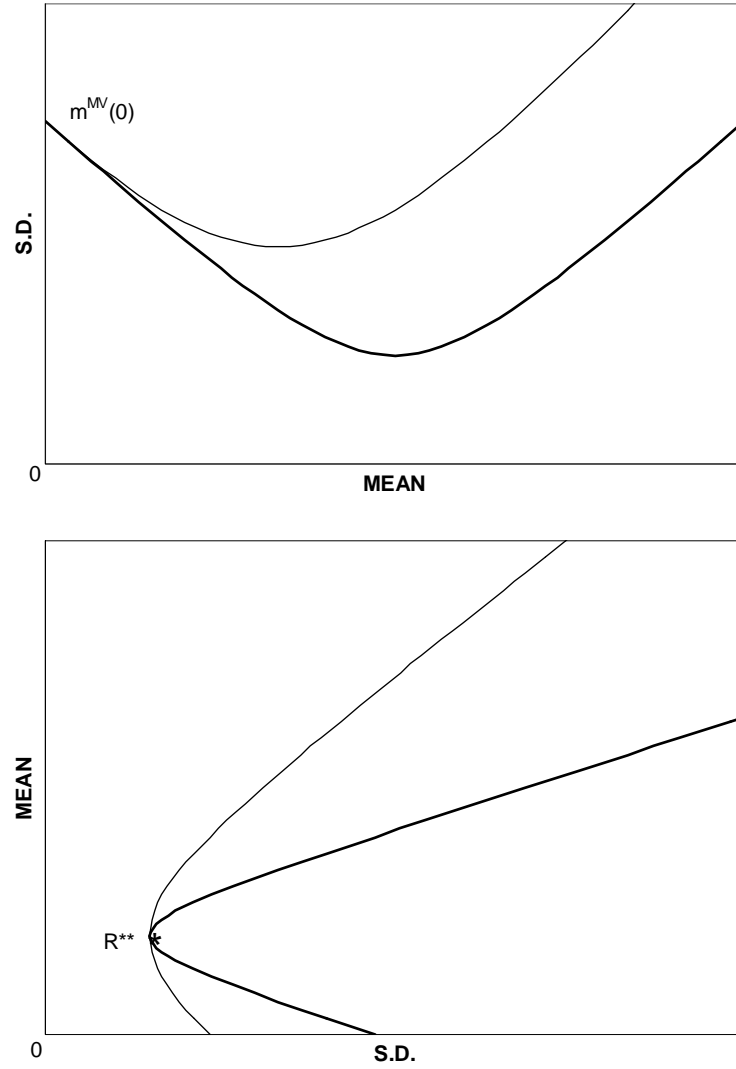


Figure 3: SDF and return frontiers such that \mathbf{R} and \mathbf{R}_1 share the centred cost representing portfolio.

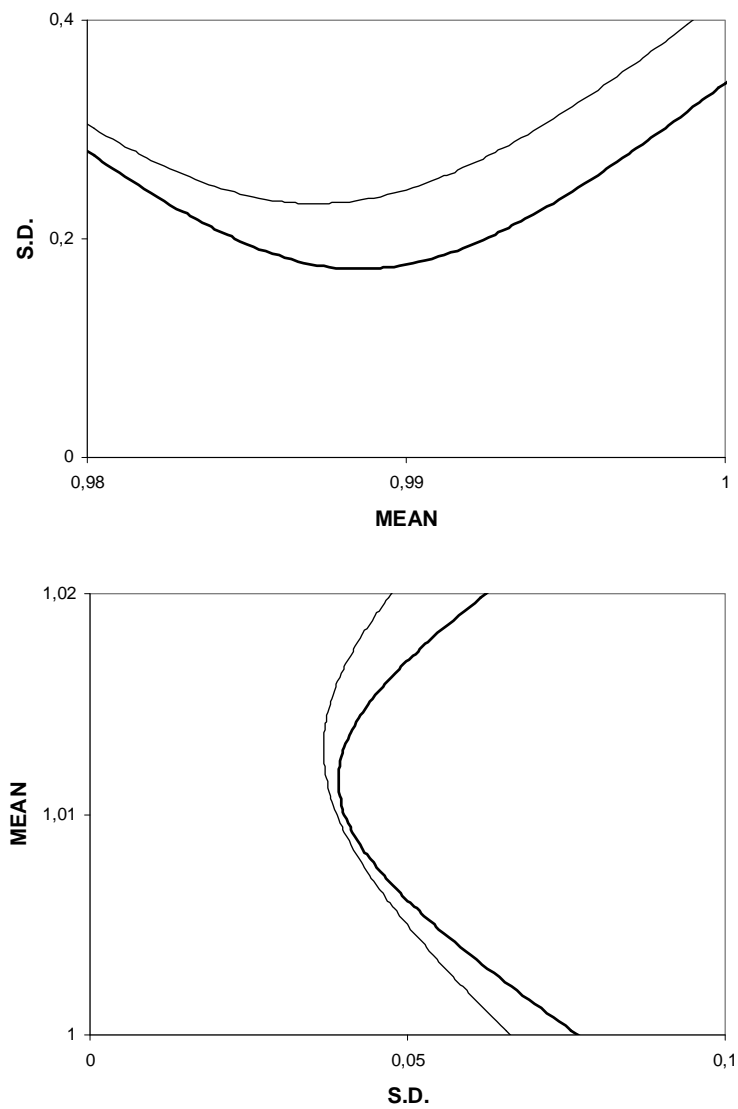


Figure 4: SDF and return frontiers for size and book-to-market sorted portfolios. \mathbf{R}_1 contains the market and the two Fama-French portfolios that capture the size and value effects. Monthly data from 1952 to 2007.