DISCUSSION PAPER SERIES

No. 4422

SPANNING TESTS IN RETURN AND STOCHASTIC DISCOUNT FACTOR MEAN-VARIANCE FRONTIERS: A UNIFYING APPROACH

Francisco Peñaranda and Enrique Sentana

FINANCIAL ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP4422.asp**

ISSN 0265-8003

SPANNING TESTS IN RETURN AND STOCHASTIC DISCOUNT FACTOR MEAN-VARIANCE FRONTIERS: A UNIFYING APPROACH

Francisco Peñaranda, London School of Economics (LSE)
Enrique Sentana, Centre for Monetary and Financial Studies (CEMFI) and CEPR

Discussion Paper No. 4422
June 2004

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 4422

June 2004

ABSTRACT

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

We propose new approaches to test for spanning in the return and stochastic discount factor mean-variance frontiers, which assess if either the centred or uncentred mean and cost representing portfolios are shared by the initial and extended sets of assets. We show that our proposed tests are asymptotically equivalent to the existing spanning tests under the null and sequences of local alternatives, and analyse their asymptotic relative efficiency. We also extend the theory of optimal GMM inference to deal with the singularities that arise in some spanning tests. Finally, we include an empirical application to money markets in Europe.

JEL Classification: G11, G12, C12 and C13,
Keywords: asset pricing, asymptotic slopes, GMM, representing portfolios and singular covariance matrix

Francisco Peñaranda
FMG
LSE
Houghton Street
London
WC2A 2AE
Email: f.penaranda@lse.ac.uk

Enrique Sentana
CEMFI
Casado del Alisal 5
E-28014 Madrid
SPAIN
Tel: (34 91) 429 0551
Fax: (34 91) 429 1056
Email: sentana@cemfi.es

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=160932

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=111330

*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 Foro de Finanzas (Alicante, 2003), the III Workshop on International Finance (Malaga, 2003), as well as audiences at ANU, CEMFI, Chicago, CIDE, LSE, Montreal, Nuffield College, Pompeu Fabra, Sydney, Tor Vergata and Venice for very useful comments and suggestions. Of course, the usual caveat applies. The first version of this paper was developed while the first author was based at CEMFI.

Submitted 29 April 2004

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 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 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. 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 recent survey); (ii) gains from portfolio diversification, often arising from 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. Campbell, Lo and MacKinlay (1996) or Cochrane (2001) for textbook treatments).

Nevertheless, given the duality of the two frontiers, it is possible to develop spanning tests that are not tied down to the specific properties of either frontier. In particular, since both frontiers are spanned by the cost and mean representing portfolios (RP's) introduced by Chamberlain and Rothschild (1983), one can simply test if these two portfolios are shared by the initial and extended sets of assets. This is precisely the approach that we follow. An important advantage of our approach is that we can directly apply Hansen's

(1982) generalised method of moments (GMM) without introducing any nuisance parameters because those RP's are defined in terms of uncentred moment conditions.

Given that Chamberlain and Rothschild (1983) considered cost and mean RP's defined in terms of central moments too, we also develop alternative testing procedures based on these centred portfolios. This second approach, though, requires the introduction of additional moment conditions that define the mean returns as nuisance parameters. Unfortunately, the joint covariance matrix of the augmented set of moment conditions turns out to be singular in the population, although not necessarily in the sample, which complicates GMM inference. For that reason, we extend the theory of optimal GMM estimation in Hansen (1982) to those non-trivial situations in which the estimating functions have a singular covariance matrix along an implicit manifold in the parameter space that contains the true value. For the benefit of practitioners, we also suggest sensible consistent first-step parameter estimators that can be used 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 the estimators and testing procedures.

In addition, 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. In this respect, 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. We also present a comparison of the power against fixed alternatives of the new and existing testing procedures by using Bahadur's notion of asymptotic relative efficiency studied by Geweke (1981).

Finally, we apply our testing procedures to shed some light on the important question of whether the elimination of intra-European exchange rate risk resulting from the European Monetary Union (EMU) has had any effect on global investors, given that it has limited the extent to which they can internationally diversify their portfolios across different currencies. We do so by testing if the opportunity set of investors who diversify their holdings across the most important developed countries was the same (in a mean-variance sense) in the pre-EMU era with and without the assets of several of the current EMU members. We concentrate on the very short end of the term structure, which is the only case in which a truly PanEuropean interbank money market has been created.

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 a singular covariance matrix. 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 describe the existing spanning tests in section 5, and devote section 6 to the asymptotic comparison of all the tests. Finally, we present our empirical application to the Euro zone money markets in section 7, and summarise our conclusions in section 8. The proofs of our main results are in the appendix, while the rest are available on request.

2 Theoretical background

In this section, we first describe the RP's introduced by Chamberlain and Rothschild (1983), which we then use to characterise the RMVF and SMVF.

2.1 Cost and Mean Representing Portfolios

Consider an economy with a finite number N of primitive risky assets whose random payoffs are defined on an underlying probability space. Let $\mathbf{R} = (R_1, \ldots, R_N)'$ denote the vector of gross returns on those assets, with first and second uncentred moments given by $\boldsymbol{\nu}$ and $\boldsymbol{\Gamma}$, respectively. We assume that these moments are bounded, which implies that $R_i \in L^2$ $(i = 1, \ldots, N)$, 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 primitive asset returns, $\boldsymbol{\Sigma}$ say, as $\boldsymbol{\Gamma} - \boldsymbol{\nu}\boldsymbol{\nu}'$, which we assume has full rank. This implies that none of the primitive 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.[1] We also assume that not all expected returns are the same.[2]

[1] Spanning tests in the presence of a safe asset are studied in Peñaranda and Sentana (2004).
[2] See Peñaranda and Sentana (2004) for a brief discussion of the equal expected returns case.

Let $\mathcal{P}_N$ 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 $\mathcal{P}_N$ will be of the form $p = \sum_{i=1}^N \omega_i R_i = \boldsymbol{\omega}'\mathbf{R}$, where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_N)' \in \mathbb{R}^N$ 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 $C(p) = \boldsymbol{\omega}'\ell_N$, $E(p) = \boldsymbol{\omega}'\boldsymbol{\nu}$ and $V(p) = \boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega}$ respectively, where ℓ_N is a vector of N ones, which reflects the fact that we have normalised the price of all the original assets to 1.[3] Since $\mathcal{P}_N$ 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 $\mathcal{P}_N$, $P(q|\mathcal{P}_N)$, as the element of $\mathcal{P}_N$ that is closest to q in the mean square norm. Specifically:

$$P(q|\mathcal{P}_N) = E(q\mathbf{R})E^{-1}(\mathbf{R}\mathbf{R}')\mathbf{R}. \tag{1}$$

In this context, we can formally understand $C(.)$ and $E(.)$ as linear functionals that map the elements of $\mathcal{P}_N$ onto the real line. Since $E(p^2) \geq E^2(p)$ by the Markov inequality, the expected value functional is always continuous on L^2. Similarly, our full rank assumption on $\boldsymbol{\Sigma}$ implies that $\boldsymbol{\Gamma}$ has full rank too, and consequently, that the cost functional is also continuous on $\mathcal{P}_N$, which is tantamount to the law of one price. The Riesz representation theorem then implies that there exist two unique elements of $\mathcal{P}_N$ that represent these functionals over $\mathcal{P}_N$ (see Chamberlain and Rothschild (1983)). In particular, the uncentred cost and mean RP's, p^* and p^+, respectively, will be such that:

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

It is then straightforward to show that

$$\begin{aligned} p^* &= \boldsymbol{\phi}^{*\prime}\mathbf{R} = \ell_N'\boldsymbol{\Gamma}^{-1}\mathbf{R}, \\ p^+ &= \boldsymbol{\phi}^{+\prime}\mathbf{R} = \boldsymbol{\nu}'\boldsymbol{\Gamma}^{-1}\mathbf{R}. \end{aligned} \tag{2}$$

If $\mathcal{P}_N$ included a unit payoff, then p^+ would coincide with it. But even though it does not, it follows from (1) that $p^+ = P(1|\mathcal{P}_N)$. To interpret p^*, it is convenient to recall that a stochastic discount factor (SDF), m say, is any scalar random variable defined on the same underlying probability space which prices assets in terms of their expected cross

[3] The case of arbitrage (i.e. zero-cost) portfolios is studied in Peñaranda and Sentana (2004).

product with it. For instance, in a complete markets set-up, m would correspond to the price of each Arrow-Debreu security divided by the probability of the corresponding state. But whatever m is, we can again use (1) to interpret p^* as $P(m|\mathcal{P}_N)$. 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 $\text{A} = cov(p^{**}, p^{++}) = \boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}\ell_N = 0$.

Finally, Chamberlain and Rothschild (1983) show that an alternative valid topology on $\mathcal{P}_N$ can be defined with covariance as inner product and standard deviation as norm when there is not a safe asset in $\mathcal{P}_N$.[4] Hence, we could also represent the two functionals by means of two alternative centred RP's, p^{**} and p^{++} in $\mathcal{P}_N$, such that

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

Not surprisingly,

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

where $\text{B} = V(p^{++}) = \boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu}$. We can then define the return associated with p^{**} as $R^{**} = \ell_N'\boldsymbol{\Sigma}^{-1}\mathbf{R}/(\ell_N'\boldsymbol{\Sigma}^{-1}\ell_N)$, 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) $\text{A} \neq 0$.[5]

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

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

If there were a safe asset with gross return c^{-1}, the only SMVF portfolio would be

$$m^{MV}(c) = c + \boldsymbol{\beta}(c)'(\mathbf{R} - \boldsymbol{\nu}) = \alpha(c) + \boldsymbol{\beta}(c)'\mathbf{R},$$

$$\boldsymbol{\beta}(c) = \boldsymbol{\Sigma}^{-1}(\ell_N - c\boldsymbol{\nu}) = \boldsymbol{\phi}^* - \alpha(c)\boldsymbol{\phi}^+, \quad \alpha(c) = c - \boldsymbol{\beta}(c)'\boldsymbol{\nu}.$$

[4] If $1 \in \mathcal{P}_N$, then the covariance and standard deviation would only constitute a proper metric over the orthogonal complement to the safe portfolio in $\mathcal{P}_N$, $\mathcal{U}_N$ say (see Chamberlain and Rothschild (1983)).

[5] When $\text{A} = 0$, both p^+ and p^{++} are arbitrage portfolios, which means that neither R^+ nor R^{++} can be defined. In addition, $p^{**} = p^*$, so that $R^{**} = R^*$.

But even though no safe asset exists, we can trace the SMVF by computing the above expression for any $c \geq 0$. It is sometimes more convenient to write $m^{MV}(c)$ as:

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

which shows that all the elements of the SMVF are portfolios spanned by p^* and $1 - p^+$ alone. Note, however, that $m^{MV}(c) \notin \mathcal{P}_N$ 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)).

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

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

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

$$R^{MV}(\nu) = R^* + p^{\#} \cdot [\nu - E(R^*)]/E(p^{\#}),$$

where $p^{\#} = p^+ - C(p^+)R^*$, 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.

Given the expressions above, it is easy to show that

$$m^{MV}(c) - \alpha(c) = p^* - \alpha(c)p^+ = \boldsymbol{\beta}'(c)\mathbf{R},$$

so 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. However, there are some exceptions to such a duality. In particular, we can go from $R^{MV}(\nu)$ to $m^{MV}(c)$ for any return belonging to the RMVF except R^+. The reason is that R^+ is the return to p^+, while $m^{MV}(c)$ always holds a unit position on p^*. However, we can still establish a relationship by using a limiting argument. In particular

$$\lim_{c \to \infty} E\left[\frac{m^{MV}(c)}{c} - (1 + \text{B})(1 - p^+)\right]^2 = 0,$$

which defines the behaviour of the asymptotes of the SMVF frontier, whose slope is $\sqrt{\text{B}}$. Similarly, we can go from $m^{MV}(c)$ to $R^{MV}(\nu)$ for any point on the SMVF except for $c = \text{A}^{-1}\text{C}$, where C$= V(p^{**}) = \ell_N'\boldsymbol{\Sigma}^{-1}\ell_N$. The problem is that $m^{MV}(\text{A}^{-1}\text{C})$ holds an arbitrage position of risky assets, for which there is no counterpart in the usual RMVF.

3 Econometric Methods

We begin by briefly reviewing the inference methods proposed by Hansen (1982), which allows us to introduce all the relevant notation and assumptions required in our extension to those cases in which the covariance 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 our proposed spanning tests in section 4.

3.1 GMM inference procedures

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 $\boldsymbol{\Theta} \subseteq \mathbb{R}^k$, is implicitly defined by the (population) moment conditions:

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

where the expectation is with respect to the stationary distribution of $\mathbf{x}_t$. In this context, the unrestricted GMM estimator of $\boldsymbol{\theta}$ will be

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

where $J_T(\boldsymbol{\theta};\boldsymbol{\Upsilon}_T) = \bar{\mathbf{h}}_T'(\boldsymbol{\theta})\boldsymbol{\Upsilon}_T\bar{\mathbf{h}}_T(\boldsymbol{\theta})$ defines a particular norm of the sample moments

$$\bar{\mathbf{h}}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{h}(\mathbf{x}_t;\boldsymbol{\theta})$$

characterised by the possibly stochastic, positive semidefinite weighting matrix $\boldsymbol{\Upsilon}_T$, which we assume converges in probability to a positive semidefinite matrix $\boldsymbol{\Upsilon}$.

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, while if both dimensions coincide, we say that $\boldsymbol{\theta}$ is (seemingly) exactly identified. 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 $\boldsymbol{\theta}$, a sufficient condition for the local identifiability of $\boldsymbol{\theta}$ at $\boldsymbol{\theta}^0$ is that $\text{rank}[\mathbf{H}(\boldsymbol{\theta}^0,\boldsymbol{\Upsilon})] = k$, where $\mathbf{H}(\boldsymbol{\theta},\boldsymbol{\Upsilon}) = \mathbf{D}'(\boldsymbol{\theta})\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta})$, which requires that $\text{rank}[\mathbf{D}(\boldsymbol{\theta}^0)] = k$. Under the additional assumptions that $E(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\mathbf{h}(\mathbf{x}_t;\boldsymbol{\theta})\|) < \infty$, $\bar{\mathbf{D}}_T(\boldsymbol{\theta}^i) \xrightarrow{p} \mathbf{D}(\boldsymbol{\theta}^0)$ if $\boldsymbol{\theta}^i \xrightarrow{p} \boldsymbol{\theta}^0$, and $\sqrt{T}\bar{\mathbf{h}}_T(\boldsymbol{\theta}^0) \xrightarrow{d} N[\mathbf{0}, \mathbf{S}(\boldsymbol{\theta}^0)]$, where $\mathbf{S}(\boldsymbol{\theta}^0)$ is positive semidefinite, then

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

(see Newey and MacFadden (1994) for more primitive regularity conditions and proofs).

The expression for $\mathbf{V}(\boldsymbol{\theta}^0, \boldsymbol{\Upsilon})$ simplifies to $\mathbf{D}^{-1}(\boldsymbol{\theta}^0)\mathbf{S}(\boldsymbol{\theta}^0)\mathbf{D}'^{-1}(\boldsymbol{\theta}^0)$ in exactly identified models, as the weighting matrix $\boldsymbol{\Upsilon}_T$ becomes 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 $\boldsymbol{\Upsilon} = \mathbf{S}^{-1}(\boldsymbol{\theta}^0)$ is the "optimal" weighting matrix when the long-run covariance matrix of the moment conditions $\mathbf{S}(\boldsymbol{\theta}^0)$ has full rank, 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. The asymptotic distribution of the optimal GMM estimator of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_T[\mathbf{S}^{-1}(\boldsymbol{\theta}^0)]$, will be

$$\sqrt{T}\{\hat{\boldsymbol{\theta}}_T[\mathbf{S}^{-1}(\boldsymbol{\theta}^0)] - \boldsymbol{\theta}^0\} \xrightarrow{d} N\{\mathbf{0}, \mathbf{H}^{-1}[\boldsymbol{\theta}^0, \mathbf{S}^{-1}(\boldsymbol{\theta}^0)]\}.$$

This optimal estimator is infeasible unless we know $\mathbf{S}(\boldsymbol{\theta}^0)$, but under additional regularity conditions, we can define a feasible asymptotically equivalent two-step optimal GMM estimator as $\hat{\boldsymbol{\theta}}_T[\bar{\mathbf{S}}_T^{-1}(\dot{\boldsymbol{\theta}}_T)]$, where $\dot{\boldsymbol{\theta}}_T$ is some initial consistent estimator of $\boldsymbol{\theta}^0$, and $\bar{\mathbf{S}}_T(\dot{\boldsymbol{\theta}}_T)$ is a heteroskedasticity and autocorrelation consistent (HAC) estimator of $\mathbf{S}(\boldsymbol{\theta}^0)$ based on $\mathbf{h}(\mathbf{x}_t; \dot{\boldsymbol{\theta}}_T)$ (see e.g. de Jong and Davidson (2000) and the references therein).

The optimal weighting matrix is also required in the so-called "overidentification" restrictions test, given by $T \cdot J_T\{\hat{\boldsymbol{\theta}}_T[\mathbf{S}^{-1}(\boldsymbol{\theta}^0)]; \mathbf{S}^{-1}(\boldsymbol{\theta}^0)\}$, which asymptotically follows a χ^2 with degrees of freedom equal to the difference between the number of moments and parameters under correct specification of the original moment conditions (Hansen (1982)).

In this GMM context, it is also straightforward to carry out hypothesis tests of the $r \leq k$ implicit parametric restrictions $\mathbf{G}(\boldsymbol{\theta}^0) = \mathbf{0}$. In particular, under the additional assumptions that $\mathbf{G}(\boldsymbol{\theta})$ is continuously differentiable, with a full-rank Jacobian matrix $\mathbf{Q}(\boldsymbol{\theta}) = \partial\mathbf{G}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ in an open neighbourhood of $\boldsymbol{\theta}^0$, and rank$[\mathbf{F}(\boldsymbol{\theta}^0, \boldsymbol{\Upsilon})] = r$, where $\mathbf{F}(\boldsymbol{\theta}^0, \boldsymbol{\Upsilon}) = \mathbf{Q}(\boldsymbol{\theta}^0)\mathbf{V}(\boldsymbol{\theta}^0, \boldsymbol{\Upsilon})\mathbf{Q}'(\boldsymbol{\theta}^0)$, we can define a potentially suboptimal Wald test as:

$$W_T(\boldsymbol{\Upsilon}_T) = T \cdot \mathbf{G}'[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)]'\mathbf{F}^{-1}(\boldsymbol{\theta}^0, \boldsymbol{\Upsilon})\mathbf{G}[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)].$$

Given our assumptions, $W_T(\boldsymbol{\Upsilon}_T)$ will be asymptotically distributed as a χ^2 with r degrees of freedom under $H_0 : \mathbf{G}(\boldsymbol{\theta}) = \mathbf{0}$, and as a non-central χ^2 with the same degrees of freedom and non-centrality parameter $\boldsymbol{\delta}'\mathbf{F}^{-1}(\boldsymbol{\theta}^0, \boldsymbol{\Upsilon})\boldsymbol{\delta}$ under the Pitman sequence of local alternatives $H_l : \mathbf{G}(\boldsymbol{\theta}) = \boldsymbol{\delta}/\sqrt{T}$ (see again Newey and MacFadden (1994)). In contrast, $W_T(\boldsymbol{\Upsilon}_T)$ will diverge to infinity for fixed alternatives of the form $H_f : \mathbf{G}(\boldsymbol{\theta}) = \boldsymbol{\delta}$, which makes it a consistent test. Theorem 1 in Geweke (1981) then implies that

$$\underset{\boldsymbol{\theta}^0}{\text{plim}} \frac{1}{T} W_T(\boldsymbol{\Upsilon}_T) = \mathbf{G}(\boldsymbol{\theta}^0)'\mathbf{F}^{-1}(\boldsymbol{\theta}^0, \boldsymbol{\Upsilon})\mathbf{G}(\boldsymbol{\theta}^0) \tag{4}$$

coincides with Bahadur's definition of the approximate slope of this Wald test. Note that (4) has the same form as the non-centrality parameter derived above, except that now $\mathbf{F}(\boldsymbol{\theta}^0, \boldsymbol{\Upsilon})$ is no longer evaluated under the null. Once again, the expression for $\mathbf{F}(\boldsymbol{\theta}^0, \boldsymbol{\Upsilon})$ simplifies when either $\boldsymbol{\theta}$ is just identified, or when $\boldsymbol{\Upsilon}_T$ is optimally chosen, in which case the non-centrality parameter and/or approximate slope will achieve its maximum.

We can also base our tests on the restricted GMM estimator $\tilde{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)$, which minimises $J_T(\boldsymbol{\theta}; \boldsymbol{\Upsilon}_T)$ over $\Theta \cap \{\mathbf{G}(\boldsymbol{\theta}) = \mathbf{0}\}$. If $\tilde{\boldsymbol{\lambda}}_T'(\boldsymbol{\Upsilon}_T)$ are the Lagrange multipliers (LM) associated with the constraints $\mathbf{G}(\boldsymbol{\theta}) = \mathbf{0}$, we can define a potentially suboptimal LM test of H_0 as:

$$LM_T(\boldsymbol{\Upsilon}_T) = T \cdot \tilde{\boldsymbol{\lambda}}_T'(\boldsymbol{\Upsilon}_T)^{-1}\boldsymbol{\Xi}^{-1}(\boldsymbol{\theta}^0, \boldsymbol{\Upsilon})\tilde{\boldsymbol{\lambda}}_T'(\boldsymbol{\Upsilon}_T),$$
$$\boldsymbol{\Xi}(\boldsymbol{\theta}^0, \boldsymbol{\Upsilon}) = [\mathbf{Q}(\boldsymbol{\theta}^0)\mathbf{H}^{-1}(\boldsymbol{\theta}^0, \boldsymbol{\Upsilon})\mathbf{Q}'(\boldsymbol{\theta}^0)]^{-1}\mathbf{F}(\boldsymbol{\theta}^0, \boldsymbol{\Upsilon})[\mathbf{Q}(\boldsymbol{\theta}^0)\mathbf{H}^{-1}(\boldsymbol{\theta}^0, \boldsymbol{\Upsilon})\mathbf{Q}'(\boldsymbol{\theta}^0)]^{-1}.$$

Importantly, Property 18.2 in Gouriéroux and Monfort (1995) indicates that for any $\boldsymbol{\Upsilon}_T$, $LM_T(\boldsymbol{\Upsilon}_T) - W_T(\boldsymbol{\Upsilon}_T) \xrightarrow{p} 0$ as $T \to \infty$ under the null and local alternatives. However, such a relationship no longer holds under fixed alternatives, even though $LM_T(\boldsymbol{\Upsilon}_T)$ also diverges to infinity in that case.

It is also possible to define the GMM analogue of the likelihood ratio test as

$$DM_T(\boldsymbol{\Upsilon}_T) = T \cdot \{J_T[\tilde{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T); \boldsymbol{\Upsilon}_T] - J_T[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T); \boldsymbol{\Upsilon}_T]\}.$$

But like the overidentifying restriction test, this "distance metric" test will have an asymptotic χ^2 distribution only if $\boldsymbol{\Upsilon}_T$ is optimally chosen, in which case it will be asymptotically equivalent to the optimal versions of the W_T and LM_T tests under the null and sequences of local alternatives (see e.g. Theorem 9.2 in Newey and MacFadden (1994)).

Finally, the following result on the asymptotic and sometimes numerical invariance of GMM estimators of functions of the parameters of interest to linear transformations and reparametrisations of the original moment conditions, will prove useful below:

Lemma 1 *Let* $\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T) = \arg\min_{\boldsymbol{\theta}} \bar{\mathbf{h}}'_T(\boldsymbol{\theta})\boldsymbol{\Upsilon}_T\bar{\mathbf{h}}_T(\boldsymbol{\theta})$ *denote a GMM estimator of the* $k \times 1$ *vector of unknown parameter* $\boldsymbol{\theta}$ *defined by the* $n \geq k$ *set of moment conditions* $E[\mathbf{h}(\mathbf{x}_t;\boldsymbol{\theta})]$ $=\mathbf{0}$*. Further, let* $\mathbf{G}(\boldsymbol{\theta})$ *denote a vector of* $r \leq k$ *continuously differentiable functions of* $\boldsymbol{\theta}$ *whose Jacobian matrix has full row rank in an open neighbourhood of* $\boldsymbol{\theta}^0$*. Finally, let* $\hat{\boldsymbol{\rho}}_T(\boldsymbol{\Upsilon}_{NT}) = \arg\min_{\boldsymbol{\rho}} \bar{\mathbf{h}}'_{NT}(\boldsymbol{\rho})\boldsymbol{\Upsilon}_{NT}\bar{\mathbf{h}}_{NT}(\boldsymbol{\rho})$ *denote a GMM estimator of the* k *unknown parameters* $\boldsymbol{\rho}$ *based on the transformed set of moment conditions* $E[\mathbf{h}_N(\mathbf{x}_t;\boldsymbol{\rho})] = \mathbf{0}$*, where* $\mathbf{h}_N(\mathbf{x}_t;\boldsymbol{\rho}) = \mathbf{A}[\mathbf{P}^{-1}(\boldsymbol{\rho})]\mathbf{h}[\mathbf{x}_t;\mathbf{P}^{-1}(\boldsymbol{\rho})] = \mathbf{A}(\boldsymbol{\theta})\mathbf{h}_O(\mathbf{x}_t;\boldsymbol{\theta})$*,* $\mathbf{A}(\boldsymbol{\theta})$ *is an* $n \times n$ *matrix of continuously differentiable functions of* $\boldsymbol{\theta}$ *in an open neighbourhood of* $\boldsymbol{\theta}^0$ *such that* $\text{rank}[\mathbf{A}(\boldsymbol{\theta}^0)] = n$*, and* $\mathbf{P}(.)$ *is a regular transformation from* $\boldsymbol{\theta}$ *to* $\boldsymbol{\rho}$ *over the same open neighbourhood. If we assume that the standard regularity conditions that guarantee the asymptotic normality of* $\hat{\boldsymbol{\theta}}_{OT}(\boldsymbol{\Upsilon}_{OT})$ *hold, then*

1. $\sqrt{T}\{\mathbf{G}_N[\hat{\boldsymbol{\rho}}_{NT}(\boldsymbol{\Upsilon}_{NT})] - \mathbf{G}[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)]\} = o_p(1)$ *if* $\boldsymbol{\Upsilon}_N = \mathbf{A}^{-1\prime}(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{A}^{-1}(\boldsymbol{\theta}^0)$*, where* $\boldsymbol{\Upsilon} - \boldsymbol{\Upsilon}_T = o_p(1)$*,* $\boldsymbol{\Upsilon}_N - \boldsymbol{\Upsilon}_{NT} = o_p(1)$ *and* $\mathbf{G}_N(\boldsymbol{\rho}) = \mathbf{G}[\mathbf{P}^{-1}(\boldsymbol{\rho})]$*.*

2. $\mathbf{G}_N[\hat{\boldsymbol{\rho}}_{NT}(\boldsymbol{\Upsilon}_{NT})] = \mathbf{G}[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)]$ *for large enough* T *if* $\mathbf{A}(\boldsymbol{\theta}) = \mathbf{A}\ \forall\boldsymbol{\theta}$ *and* $\boldsymbol{\Upsilon}_{NT} = \mathbf{A}'^{-1}\boldsymbol{\Upsilon}_T\mathbf{A}^{-1}$*.*

3.2 Optimal GMM with a singular covariance matrix

Unfortunately, the previous definitions of optimal GMM estimators and tests break down when $\mathbf{S}(\boldsymbol{\theta}^0)$ is singular. In this section, we obtain both the optimal GMM estimators of $\boldsymbol{\theta}$ and the optimal tests of $H_0 : \mathbf{G}(\boldsymbol{\theta}) = \mathbf{0}$ in those non-trivial situations in which the covariance matrix of $\mathbf{h}(\mathbf{x}_t;\boldsymbol{\theta})$ is singular along a manifold in $\boldsymbol{\Theta}$ which includes $\boldsymbol{\theta}^0$. The exact nature of such a singularity, which is the relevant one for the spanning tests in sections 4.2 and 5.1, can be fully characterised by the following three assumptions:

Assumption 1 *Let* $\boldsymbol{\Pi}(\boldsymbol{\theta})$ *denote a* $n \times k_\ominus$ *matrix of continuously differentiable functions of* $\boldsymbol{\theta}$*, where* $0 \leq k_\ominus \leq k$*. Then,* $\boldsymbol{\Pi}'(\boldsymbol{\theta})\mathbf{h}(\mathbf{x}_t,\boldsymbol{\theta}) = \mathbf{0}\ \forall\mathbf{x}_t$ *if and only* $\mathbf{m}_\ominus(\boldsymbol{\theta}) = \mathbf{0}$*, where* $\mathbf{m}_\ominus(\boldsymbol{\theta})$ *is a* $k_\ominus \times 1$ *continuously differentiable transformation of* $\boldsymbol{\theta}$ *such that the rank of* $\partial\mathbf{m}_\ominus(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ *is* $k_\ominus$ *in an open neighbourhood of* $\boldsymbol{\theta}^0$*.*

Assumption 2 *If* $k_\ominus > 0$*, then* $\mathbf{m}_\ominus(\boldsymbol{\theta}^0) = \mathbf{0}$*.*

Assumption 3 $\text{rank}[\mathbf{S}(\boldsymbol{\theta}^0)] = n - k_\ominus$*.*

For the non-standard case of $k_\ominus > 0$, the first assumption implicitly defines the $k_\oplus$-dimensional manifold in $\boldsymbol{\Theta}$ over which the singularity in the contemporaneous covariance matrix of $\mathbf{h}(\mathbf{x}_t,\boldsymbol{\theta})$ takes place, where $k_\oplus = k - k_\ominus$, while the second assumption says that the true values of the parameters belong to that manifold. Finally, Assumption 3 ensures that the singularity of $\mathbf{S}(\boldsymbol{\theta}^0)$, when it exists, is fully characterised by Assumption 1.[6]

[6]Hence, we rule out trivial situations with "duplicated" moment conditions, in which some linear combinations of $\mathbf{h}(\mathbf{x}_t,\boldsymbol{\theta})$ whose coefficients do not depend on $\boldsymbol{\theta}$ are singular. Given that in those cases any HAC estimator of $\mathbf{S}(\boldsymbol{\theta}^0)$, $\bar{\mathbf{S}}_T(\dot{\boldsymbol{\theta}}_T)$, will be singular in finite samples irrespective of the choice of first-step estimator $\dot{\boldsymbol{\theta}}_T$, 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 $\bar{\mathbf{S}}_T(\dot{\boldsymbol{\theta}}_T)$.

For notational simplicity, but without loss of generality, we shall work with the alternative $k_\oplus + k_\ominus$ parameters $(\boldsymbol{\theta}'_\oplus, \boldsymbol{\theta}'_\ominus) = [\mathbf{m}'_\oplus(\boldsymbol{\theta}), \mathbf{m}'_\ominus(\boldsymbol{\theta})] = \mathbf{m}'(\boldsymbol{\theta})$, which we can always choose to be a regular transformation on an open neighbourhood of $\boldsymbol{\theta}^0$ in view of Assumptions 1 and 2 (see e.g. Fleming (1977, p. 143)). The GMM estimators of $\boldsymbol{\theta}$ will then be obtained from the estimators of $\boldsymbol{\theta}_\oplus$ and $\boldsymbol{\theta}_\ominus$ by means of the inverse transformation $\mathbf{l}[\mathbf{m}(\boldsymbol{\theta})] = \boldsymbol{\theta}$. In this context, our proposed solution for conducting optimal GMM estimation and inference under singularity (or optimal GMMS for short) involves the following two steps:

a) replace the ordinary inverse of $\mathbf{S}(\boldsymbol{\theta}^0)$, which cannot be defined when $k_\ominus > 0$, by any of its generalised inverses, $\mathbf{S}^-(\boldsymbol{\theta}^0)$, and simultaneously

b) impose the parametric restrictions $\mathbf{m}_\ominus(\boldsymbol{\theta}) = \boldsymbol{\theta}_\ominus = \mathbf{0}$ by working with the smaller vector of parameters $\boldsymbol{\theta}_\oplus$.

In this way, we effectively decrease both the number of parameters and the number of moment conditions to avoid the singularity, but their difference remains the same.[7]

More specifically, let

$$\mathbf{S}(\boldsymbol{\theta}^0) = \begin{bmatrix} \mathbf{P}_\oplus(\boldsymbol{\theta}^0) & \mathbf{P}_\ominus(\boldsymbol{\theta}^0) \end{bmatrix} \begin{bmatrix} \boldsymbol{\Delta}_\oplus(\boldsymbol{\theta}^0) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{P}'_\oplus(\boldsymbol{\theta}^0) \\ \mathbf{P}'_\ominus(\boldsymbol{\theta}^0) \end{bmatrix} = \mathbf{P}_\oplus(\boldsymbol{\theta}^0)\boldsymbol{\Delta}_\oplus(\boldsymbol{\theta}^0)\mathbf{P}'_\oplus(\boldsymbol{\theta}^0)$$

denote the spectral decomposition of $\mathbf{S}(\boldsymbol{\theta}^0)$, where $\boldsymbol{\Delta}_\oplus(\boldsymbol{\theta}^0)$ is a positive definite diagonal matrix of order $n - k_\ominus$, so that all its generalised inverses will be of the form

$$\mathbf{S}^-(\boldsymbol{\theta}^0) = \begin{bmatrix} \mathbf{P}_\oplus(\boldsymbol{\theta}^0) & \mathbf{P}_\ominus(\boldsymbol{\theta}^0) \end{bmatrix} \begin{bmatrix} \boldsymbol{\Delta}_\oplus^{-1}(\boldsymbol{\theta}^0) & \boldsymbol{\Delta}^{\oplus\ominus}(\boldsymbol{\theta}^0) \\ \boldsymbol{\Delta}^{\ominus\oplus}(\boldsymbol{\theta}^0) & \boldsymbol{\Delta}^{\ominus\ominus}(\boldsymbol{\theta}^0) \end{bmatrix} \begin{bmatrix} \mathbf{P}'_\oplus(\boldsymbol{\theta}^0) \\ \mathbf{P}'_\ominus(\boldsymbol{\theta}^0) \end{bmatrix},$$

with $\boldsymbol{\Delta}^{\oplus\ominus}(\boldsymbol{\theta}^0)$, $\boldsymbol{\Delta}^{\ominus\oplus}(\boldsymbol{\theta}^0)$ and $\boldsymbol{\Delta}^{\ominus\ominus}(\boldsymbol{\theta}^0)$ arbitrary (see e.g. Rao and Mitra (1971)). In addition, let us define the following set of moment conditions:

$$\begin{bmatrix} \mathbf{h}_\oplus(\mathbf{x}_t; \boldsymbol{\theta}_\oplus, \boldsymbol{\theta}_\ominus|\boldsymbol{\theta}^0) \\ \mathbf{h}_\ominus(\mathbf{x}_t; \boldsymbol{\theta}_\oplus, \boldsymbol{\theta}_\ominus|\boldsymbol{\theta}^0) \end{bmatrix} = \begin{bmatrix} \mathbf{P}'_\oplus(\boldsymbol{\theta}^0)\mathbf{h}[\mathbf{x}_t, \mathbf{l}(\boldsymbol{\theta}_\oplus, \boldsymbol{\theta}_\ominus)] \\ \mathbf{P}'_\ominus(\boldsymbol{\theta}^0)\mathbf{h}[\mathbf{x}_t, \mathbf{l}(\boldsymbol{\theta}_\oplus, \boldsymbol{\theta}_\ominus)] \end{bmatrix} = \mathbf{P}'(\boldsymbol{\theta}^0)\mathbf{h}[\mathbf{x}_t, \mathbf{l}(\boldsymbol{\theta}_\oplus, \boldsymbol{\theta}_\ominus)],$$

[7] In the unlikely situation of $k_\ominus = k$, the dimension of $\boldsymbol{\theta}_\oplus$ 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 recover the true value of the parameters without any sampling variability. In contrast, if $k_\ominus = 0$, the dimension of $\boldsymbol{\theta}_\ominus$ 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 $\boldsymbol{\theta}_\oplus = \boldsymbol{\theta}$ by means of the regular GMM methods discussed in the previous section.

which is a full-rank linear transformation with constant coefficients of the original moment conditions $\mathbf{h}(\mathbf{x}_t,\boldsymbol{\theta})$. In this notation, the optimal GMMS criterion function is:

$$J_T[\mathbf{l}(\boldsymbol{\theta}_\oplus,\mathbf{0});\mathbf{S}^-(\boldsymbol{\theta}^0)] = J_T[\mathbf{l}(\boldsymbol{\theta}_\oplus,\mathbf{0});\mathbf{S}^+(\boldsymbol{\theta}^0)] + \bar{\mathbf{h}}'_{\oplus T}(\boldsymbol{\theta}_\oplus,\mathbf{0}|\boldsymbol{\theta}^0)\boldsymbol{\Delta}^{\oplus\ominus}(\boldsymbol{\theta}^0)\bar{\mathbf{h}}_{\ominus T}(\boldsymbol{\theta}_\oplus,\mathbf{0}|\boldsymbol{\theta}^0)$$
$$+\bar{\mathbf{h}}'_{\ominus T}(\boldsymbol{\theta}_\oplus,\mathbf{0}|\boldsymbol{\theta}^0)\boldsymbol{\Delta}^{\ominus\oplus}(\boldsymbol{\theta}^0)\bar{\mathbf{h}}'_{\oplus T}[\mathbf{l}(\boldsymbol{\theta}_\oplus,\mathbf{0})|\boldsymbol{\theta}^0] + \bar{\mathbf{h}}'_{\ominus T}(\boldsymbol{\theta}_\oplus,\mathbf{0}|\boldsymbol{\theta}^0)\boldsymbol{\Delta}^{\ominus\ominus}(\boldsymbol{\theta}^0)\bar{\mathbf{h}}_{\ominus T}(\boldsymbol{\theta}_\oplus,\mathbf{0}|\boldsymbol{\theta}^0),$$
$$J_T[\mathbf{l}(\boldsymbol{\theta}_\oplus,\mathbf{0});\mathbf{S}^+(\boldsymbol{\theta}^0)] = \bar{\mathbf{h}}'_{\oplus T}(\boldsymbol{\theta}_\oplus,\mathbf{0}|\boldsymbol{\theta}^0)\boldsymbol{\Delta}_\oplus^{-1}(\boldsymbol{\theta}^0)\bar{\mathbf{h}}'_{\oplus T}[\mathbf{l}(\boldsymbol{\theta}_\oplus,\mathbf{0})|\boldsymbol{\theta}^0].$$

To keep the algebra simple, we initially use the Moore-Penrose inverse $\mathbf{S}^+(\boldsymbol{\theta}^0)$ in our discussion, although as argued below, the choice of $\mathbf{S}^-(\boldsymbol{\theta}^0)$ is asymptotically inconsequential.

Importantly, note that if we simply weighted the original moment conditions by $\mathbf{S}^+(\boldsymbol{\theta}^0)$ without exploiting the restrictions implicit in $\boldsymbol{\theta}_\ominus = \mathbf{0}$, the resulting estimators and testing procedures would be generally suboptimal because they would give no weight to precisely the $k_\ominus$ linear combinations of $\mathbf{h}(\mathbf{x}_t,\boldsymbol{\theta})$ estimable without error. In fact, it may well happen that the transformed parameters $\boldsymbol{\theta}_\oplus$ and $\boldsymbol{\theta}_\ominus$ are not even identified from the reduced set of $n-k_\ominus$ moment conditions $E[\mathbf{h}_\oplus(\mathbf{x}_t;\boldsymbol{\theta}_\oplus,\boldsymbol{\theta}_\ominus|\boldsymbol{\theta}^0)] = \mathbf{0}$, because, for instance, $n-k_\ominus < k$. After imposing the restriction $\boldsymbol{\theta}_\ominus = \mathbf{0}$, on the other hand, these reduced moment conditions will locally identify $\boldsymbol{\theta}_\oplus$ at $\boldsymbol{\theta}_\oplus^0$, as the following result guarantees:

Proposition 1 *Let $\mathbf{h}(\mathbf{x}_t;\boldsymbol{\theta})$ denote a set of n continuously differentiable functions of the k dimensional vector of parameters $\boldsymbol{\theta}$. If Assumptions 1 and 2 hold, and* $\text{rank}[\mathbf{D}(\boldsymbol{\theta}^0)] = k$, *then* $\text{rank}[\mathbf{D}_{\oplus\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}_\oplus^0|\boldsymbol{\theta}^0)] = k_\oplus$, *while* $\text{rank}[\mathbf{D}_{\ominus\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}_\oplus^0|\boldsymbol{\theta}^0)] = 0$, *where* $\mathbf{D}(\boldsymbol{\theta}^0) = E[\mathbf{D}_t(\boldsymbol{\theta})]$, $\mathbf{D}_{\oplus\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}_\oplus|\boldsymbol{\theta}^0) = E[\mathbf{D}_{\oplus\boldsymbol{\theta}_\oplus t}(\boldsymbol{\theta}_\oplus|\boldsymbol{\theta}^0)]$, $\mathbf{D}_{\ominus\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}_\oplus|\boldsymbol{\theta}^0) = E[\mathbf{D}_{\ominus\boldsymbol{\theta}_\oplus t}(\boldsymbol{\theta}_\oplus|\boldsymbol{\theta}^0)]$,

$$\begin{aligned}
\mathbf{D}_t(\boldsymbol{\theta}) &= \frac{\partial\mathbf{h}(\mathbf{x}_t;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'},\\
\mathbf{D}_{\oplus\boldsymbol{\theta}_\oplus t}(\boldsymbol{\theta}_\oplus|\boldsymbol{\theta}^0) &= \frac{\partial\mathbf{h}_\oplus(\mathbf{x}_t;\boldsymbol{\theta}_\oplus,\mathbf{0}|\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}'_\oplus} = \mathbf{P}'_\oplus(\boldsymbol{\theta}^0)\mathbf{D}_t[\mathbf{l}(\boldsymbol{\theta}_\oplus,\mathbf{0})]\mathbf{L}_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}_\oplus,\mathbf{0}),\\
\mathbf{D}_{\ominus\boldsymbol{\theta}_\oplus t}(\boldsymbol{\theta}_\oplus|\boldsymbol{\theta}^0) &= \frac{\partial\mathbf{h}_\ominus(\mathbf{x}_t;\boldsymbol{\theta}_\oplus,\mathbf{0}|\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}'_\oplus} = \mathbf{P}'_\ominus(\boldsymbol{\theta}^0)\mathbf{D}_t[\mathbf{l}(\boldsymbol{\theta}_\oplus,\mathbf{0})]\mathbf{L}_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}_\oplus,\mathbf{0}),\\
\mathbf{L}(\boldsymbol{\theta}_\oplus,\boldsymbol{\theta}_\ominus) &= \frac{\partial\mathbf{l}(\boldsymbol{\theta}_\oplus,\boldsymbol{\theta}_\ominus)}{\partial(\boldsymbol{\theta}'_\oplus,\boldsymbol{\theta}'_\ominus)} = [\ \mathbf{L}_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}_\oplus,\boldsymbol{\theta}_\ominus)\quad \mathbf{L}_{\boldsymbol{\theta}_\ominus}(\boldsymbol{\theta}_\oplus,\boldsymbol{\theta}_\ominus)\ ].
\end{aligned}$$

Similarly, we can also show that if we imposed the parametric restrictions $\boldsymbol{\theta}_\ominus = \mathbf{0}$, but used a weighting matrix such that $\boldsymbol{\Upsilon} \neq \mathbf{S}^-(\boldsymbol{\theta}^0)$, then the resulting estimators and testing procedures would also be generally suboptimal. In this sense, our solution to the singular GMM case can be regarded as the natural extension of the approach discussed in Judge et al. (1985, section 12.5.2) in the context of a classical multivariate regression with a singular residual covariance matrix, since they also reduce the number of equations by using the principal components of the multivariate regression residuals, as well as the number of parameters by exploiting the parametric restrictions that give rise to the singularity.

If $\mathbf{h}(\mathbf{x}_t;\boldsymbol{\theta})$ satisfies the regularity conditions mentioned in the previous section, together with Assumptions 1 and 2, then we can easily prove that those regularity conditions will also be satisfied by $\mathbf{h}_{\oplus}(\mathbf{x}_t;\boldsymbol{\theta}_{\oplus},\mathbf{0}|\boldsymbol{\theta}^0)$ because the latter functions are a linear combination of the former, and the transformation from $\boldsymbol{\theta}$ to $(\boldsymbol{\theta}_{\oplus},\boldsymbol{\theta}_{\ominus})$ is regular over an open neighbourhood of $\boldsymbol{\theta}^0$. This fact, together with Proposition 1, allows us to derive the asymptotic distribution of the infeasible unrestricted GMMS estimator of the transformed parameters $\boldsymbol{\theta}_{\oplus}$, $\hat{\boldsymbol{\theta}}_{\oplus T}[\mathbf{S}^{+}(\boldsymbol{\theta}^0)] = \arg\min_{\boldsymbol{\theta}_{\oplus}\in\boldsymbol{\Theta}_{\oplus}} J_T[\mathbf{l}(\boldsymbol{\theta}_{\oplus},\mathbf{0});\mathbf{S}^{+}(\boldsymbol{\theta}^0)]$. Specifically,

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

We can also prove that regardless of the choice of generalised inverse $\mathbf{S}^{-}(\boldsymbol{\theta}^0)$,

$$\sqrt{T}\{\hat{\boldsymbol{\theta}}_{\oplus T}[\mathbf{S}^{-}(\boldsymbol{\theta}^0)]-\hat{\boldsymbol{\theta}}_{\oplus T}[\mathbf{S}^{+}(\boldsymbol{\theta}^0)]\} = o_p(1),$$

where $\hat{\boldsymbol{\theta}}_{\oplus T}[\mathbf{S}^{-}(\boldsymbol{\theta}^0)] = \arg\min_{\boldsymbol{\theta}_{\oplus}\in\boldsymbol{\Theta}_{\oplus}} J_T[\mathbf{l}(\boldsymbol{\theta}_{\oplus},\mathbf{0});\mathbf{S}^{-}(\boldsymbol{\theta}^0)]$. As shown by Proposition 1, the intuitive reason is that there is no identifying information whatsoever about $\boldsymbol{\theta}_{\oplus}$ in the moment conditions $E[\mathbf{h}_{\ominus}(\mathbf{x}_t;\boldsymbol{\theta}_{\oplus},\mathbf{0}|\boldsymbol{\theta}^0)]=\mathbf{0}$ because $\mathbf{h}_{\ominus}(\mathbf{x}_t;\boldsymbol{\theta}_{\oplus}^0,\mathbf{0}|\boldsymbol{\theta}^0)=\mathbf{0}$ $\forall t$.

Finally, we can use the standard delta method to show that the optimal "unrestricted" GMMS estimators of the parameters of interest, $\boldsymbol{\theta}$, which will be given by $\mathbf{l}\{\hat{\boldsymbol{\theta}}_{\oplus T}\left[\mathbf{S}^{-}(\boldsymbol{\theta}^0)\right],\mathbf{0}\}$, will have an asymptotically normal distribution, but with a singular covariance matrix of rank $k_{\oplus}$. Intuitively, the reason is simply that for large enough T

$$\mathbf{l}\{\hat{\boldsymbol{\theta}}_{\oplus T}[\mathbf{S}^{+}(\boldsymbol{\theta}^0)],\mathbf{0}\} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} J_T[\boldsymbol{\theta};\mathbf{S}^{+}(\boldsymbol{\theta}^0)] \text{ s.t. } \mathbf{m}_{\ominus}(\boldsymbol{\theta})=\mathbf{0}.$$

The following Proposition confirms our claimed optimality of $\mathbf{l}\{\hat{\boldsymbol{\theta}}_{\oplus T}\left[\mathbf{S}^{-}(\boldsymbol{\theta}^0)\right],\mathbf{0}\}$:

Proposition 2 *Let $\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T) = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} J_T(\boldsymbol{\theta};\boldsymbol{\Upsilon}_T)$ denote a 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, 2 and 3. Similarly, let $\hat{\boldsymbol{\theta}}_{\oplus T}[\mathbf{S}^{-}(\boldsymbol{\theta}^0)] = \arg\min_{\boldsymbol{\theta}_{\oplus}\in\boldsymbol{\Theta}_{\oplus}} J_T[\mathbf{l}(\boldsymbol{\theta}_{\oplus},\mathbf{0});\mathbf{S}^{-}(\boldsymbol{\theta}^0)]$. Then $\mathbf{l}\{\hat{\boldsymbol{\theta}}_{\oplus T}[\mathbf{S}^{-}(\boldsymbol{\theta}^0)],\mathbf{0}\}$ is asymptotically at least as efficient as $\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)$ regardless of $\boldsymbol{\Upsilon}_T$.*

Assuming that $k_{\oplus} \geq r$, we can also define the infeasible, Moore-Penrose-based, optimal restricted GMMS estimator of $\boldsymbol{\theta}$ as $\mathbf{l}\{\tilde{\boldsymbol{\theta}}_{\oplus T}[\mathbf{S}^{+}(\boldsymbol{\theta}^0)], \mathbf{0}\}$, where $\tilde{\boldsymbol{\theta}}_{\oplus T}[\mathbf{S}^{+}(\boldsymbol{\theta}^0)]$ minimises $J_T[\mathbf{l}(\boldsymbol{\theta}_{\oplus}, \mathbf{0}), \mathbf{S}^{+}(\boldsymbol{\theta}^0)]$ over $\boldsymbol{\Theta}_{\oplus} \cap \{\mathbf{G}_{\oplus}(\boldsymbol{\theta}_{\oplus}) = \mathbf{0}\}$, with $\mathbf{G}_{\oplus}(\boldsymbol{\theta}_{\oplus}) = \mathbf{G}[\mathbf{l}(\boldsymbol{\theta}_{\oplus}, \mathbf{0})]$. Further, we can easily show that $\mathbf{l}\{\tilde{\boldsymbol{\theta}}_{\oplus T}[\mathbf{S}^{+}(\boldsymbol{\theta}^0)], \mathbf{0}\}$ is numerically equivalent for large enough T to both $\arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} J_T[\boldsymbol{\theta}; \mathbf{S}^{+}(\boldsymbol{\theta}^0)]$ s.t. $\mathbf{m}_{\ominus}(\boldsymbol{\theta}) = \mathbf{0}$ *and* $\mathbf{G}(\boldsymbol{\theta}) = \mathbf{0}$, and an unrestricted GMMS estimator of $\boldsymbol{\theta}$ in an extended system which includes not only the n original moment conditions $E[\mathbf{h}(\mathbf{x}_t, \boldsymbol{\theta})] = \mathbf{0}$, but also $\mathbf{G}(\boldsymbol{\theta}) = \mathbf{0}$ as r additional singular "moment conditions". Therefore, we can adapt Proposition 2 to show that $\mathbf{l}\{\tilde{\boldsymbol{\theta}}_{\oplus T}[\mathbf{S}^{+}(\boldsymbol{\theta}^0)], \mathbf{0}\}$ is asymptotically at least as efficient as any other GMM estimator of $\boldsymbol{\theta}$ which imposes the restrictions $\mathbf{G}(\boldsymbol{\theta}) = \mathbf{0}$, but which either does not impose the singularity restrictions $\mathbf{m}_{\ominus}(\boldsymbol{\theta}) = \mathbf{0}$, or does not use the optimal class of weighting matrices $\mathbf{S}^{-}(\boldsymbol{\theta}^0)$.

Finally, we can use $\hat{\boldsymbol{\theta}}_{\oplus T}[\mathbf{S}^{+}(\boldsymbol{\theta}^0)]$ and $\tilde{\boldsymbol{\theta}}_{\oplus T}[\mathbf{S}^{+}(\boldsymbol{\theta}^0)]$ to define optimal GMMS versions of W_T, LM_T and DM_T for the modified null hypothesis $H_0 : \mathbf{G}_{\oplus}(\boldsymbol{\theta}_{\oplus}) = \mathbf{0}$. If we further assume that $\text{rank}[\mathbf{Q}_{\oplus}(\boldsymbol{\theta}^0_{\oplus})] = r$, where $\mathbf{Q}_{\oplus}(\boldsymbol{\theta}_{\oplus}) = \partial \mathbf{G}_{\oplus}(\boldsymbol{\theta}_{\oplus})/\partial \boldsymbol{\theta}'_{\oplus}$, then we can easily prove that those three optimal tests will be asymptotically equivalent to each other under the null and sequences of local alternatives, being distributed as a central and a non-central χ^2 with r degrees of freedom, respectively. Moreover, they will separately diverge to infinity under fixed alternatives.

In practice, the optimal GMMS 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 it by a consistent estimator. However, an estimator of $\mathbf{S}^{-}(\boldsymbol{\theta}^0)$ must be chosen with some care when $k_{\ominus} > 0$ 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 $\mathbf{S}(\boldsymbol{\theta}^0)$, $\bar{\mathbf{S}}_T(\dot{\boldsymbol{\theta}}_T)$ say, 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 $\bar{\mathbf{S}}_T(\dot{\boldsymbol{\theta}}_T)$ has full rank for finite T. Hence, a researcher who is unaware of the singularity of $\mathbf{S}(\boldsymbol{\theta}^0)$ because her choice of $\dot{\boldsymbol{\theta}}_T$ is such that $\mathbf{m}_{\ominus}(\dot{\boldsymbol{\theta}}_T) \neq \mathbf{0}$, may well end up with seemingly optimal estimators and testing procedures whose asymptotic distribution will be non-standard. For that reason, we shall restrict our attention to those consistent estimators of $\boldsymbol{\theta}^0$, $\ddot{\boldsymbol{\theta}}_T$ say, that satisfy $\mathbf{m}_{\ominus}(\ddot{\boldsymbol{\theta}}_T) = \mathbf{0}$. In this way, the rank of $\bar{\mathbf{S}}_T(\ddot{\boldsymbol{\theta}}_T)$ is guaranteed to be $k_{\oplus}$ in finite samples because $\mathbf{\Pi}'(\ddot{\boldsymbol{\theta}}_T)\mathbf{h}(\mathbf{x}_t, \ddot{\boldsymbol{\theta}}_T) = \mathbf{0}\ \forall t$.

In this respect, note that a Hansen, Heaton and Yaron (1996) continuously updated criterion function of the form $J_T\{\mathbf{l}(\boldsymbol{\theta}_\oplus, 0); \bar{\mathbf{S}}_T^-[\mathbf{l}(\boldsymbol{\theta}_\oplus, \mathbf{0})]\}$ would be numerically invariant to the choice of generalised inverse because $\mathbf{P}'_\ominus[\mathbf{l}(\boldsymbol{\theta}_\oplus, \mathbf{0})]\mathbf{h}[\mathbf{x}_t, \mathbf{l}(\boldsymbol{\theta}_\oplus, \mathbf{0})] = \mathbf{0}\ \forall \mathbf{x}_t$ *and* $\forall \boldsymbol{\theta}_\oplus$. From this perspective, $\bar{\mathbf{S}}_T^+(\ddot{\boldsymbol{\theta}}_T)$ provides the two-step choice of $\mathbf{S}^-(\boldsymbol{\theta}^0)$ that is closest to such a continuously updated estimator.

Therefore, our feasible GMMS estimators will be based on the moment conditions

$$\mathbf{h}_\oplus(\mathbf{x}_t; \boldsymbol{\theta}_\oplus, \mathbf{0}|\ddot{\boldsymbol{\theta}}_T) = \bar{\mathbf{P}}'_{\oplus T}(\ddot{\boldsymbol{\theta}}_T)\mathbf{h}[\mathbf{x}_t, \mathbf{l}(\boldsymbol{\theta}_\oplus, \mathbf{0})],$$

whose regular asymptotic covariance matrix, $\Delta_\oplus(\boldsymbol{\theta}^0)$, can be consistently estimated as $\bar{\boldsymbol{\Delta}}_T(\ddot{\boldsymbol{\theta}}_T)$, where $\bar{\mathbf{P}}_{\oplus T}(\ddot{\boldsymbol{\theta}}_T)\bar{\boldsymbol{\Delta}}_T(\ddot{\boldsymbol{\theta}}_T)\bar{\mathbf{P}}'_{\oplus T}(\ddot{\boldsymbol{\theta}}_T)$ provides the spectral decomposition of $\bar{\mathbf{S}}_T(\ddot{\boldsymbol{\theta}}_T)$.

Finally, it is worth mentioning that if the original moment conditions exactly identify $\boldsymbol{\theta}$, our proposed GMMS approach is strictly speaking unnecessary as far as the *unrestricted* estimators of $\boldsymbol{\theta}$ are concerned, because $J_T[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T); \boldsymbol{\Upsilon}_T] = 0$ for large enough T regardless of $\boldsymbol{\Upsilon}_T$. The following result makes the relationship between the two unrestricted estimators and the corresponding Wald tests explicit:

Lemma 2 *Let* $\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T) = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} J_T(\boldsymbol{\theta}; \boldsymbol{\Upsilon}_T)$ *denote a GMM estimator of the* $k \times 1$ *vector of unknown parameter* $\boldsymbol{\theta}$ *defined by the* $n = k$ *exactly identified moment conditions* $E[\mathbf{h}(\mathbf{x}_t; \boldsymbol{\theta})] = \mathbf{0}$*, which satisfy all the usual regularity conditions, together with Assumptions 1, 2 and 3. Similarly, let* $\hat{\boldsymbol{\theta}}_{\oplus T}[\mathbf{S}^-(\boldsymbol{\theta}^0)] = \arg\min_{\boldsymbol{\theta}_\oplus\in\boldsymbol{\Theta}_\oplus} J_T[\mathbf{l}(\boldsymbol{\theta}_\oplus, \mathbf{0}); \mathbf{S}^-(\boldsymbol{\theta}^0)]$*. Then*

1. $\sqrt{T}[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T) - \mathbf{l}\{\hat{\boldsymbol{\theta}}_{\oplus T}[\mathbf{S}^-(\boldsymbol{\theta}^0)], \mathbf{0}\}] = o_p(1)$ *for any* $\boldsymbol{\Upsilon}_T$ *whose probability limit is a positive definite matrix* $\boldsymbol{\Upsilon}$*,*

2. $\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T) = \mathbf{l}[\hat{\boldsymbol{\theta}}_{\oplus T}\{\bar{\mathbf{S}}_T^+[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)\}, \mathbf{0}]$ *for large enough* T*, where* $\hat{\boldsymbol{\theta}}_{\oplus T}\{\bar{\mathbf{S}}_T^+[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)\} = \arg\min_{\boldsymbol{\theta}_\oplus\in\boldsymbol{\Theta}_\oplus} J_T\{\mathbf{l}(\boldsymbol{\theta}_\oplus, \mathbf{0}); \bar{\mathbf{S}}_T^+[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)]\}$*, if* $\mathbf{m}_\ominus[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)] = \mathbf{0}$ *and* $\boldsymbol{\Upsilon}_T$ *any positive definite matrix,*

3. *The Wald tests based on* $\mathbf{G}_\oplus\{\hat{\boldsymbol{\theta}}_{\oplus T}[\mathbf{S}^+(\boldsymbol{\theta}^0)]\}$ *and* $\mathbf{G}[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)]$ *will also be asymptotically equivalent if* $\mathrm{rank}[\mathbf{Q}_\oplus(\boldsymbol{\theta}_\oplus^0)] = r$*, and*

4. *It is possible to define asymptotically valid Wald test statistics based on* $\mathbf{G}[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)]$ *and* $\mathbf{G}_\oplus[\hat{\boldsymbol{\theta}}_{\oplus T}\{\bar{\mathbf{S}}_T^+[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)\}]$ *which are also numerically identical for large enough* T*.*

In contrast, the restricted estimators $\mathbf{l}\{\tilde{\boldsymbol{\theta}}_{\oplus T}[\mathbf{S}^+(\boldsymbol{\theta}^0)], \mathbf{0}\}$ and $\tilde{\boldsymbol{\theta}}_T[\boldsymbol{\Upsilon}_T]$ will not be asymptotically equivalent in general in exactly identified cases in view of Proposition 2, because in effect the restrictions $\mathbf{G}(\boldsymbol{\theta}) = \mathbf{0}$ transform the original model into an overidentified one.

4 Representing portfolios tests for spanning

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. 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.[8] The purpose of this section is to develop spanning tests based on the cost and mean RP's described in section 2.

4.1 Uncentred cost and mean representing portfolios

Given that the cost and mean RP's span both the SMVF and RMVF, a rather natural way to test for spanning consists in studying whether these portfolios are common to $\langle \mathbf{R}_1 \rangle$ and $\langle \mathbf{R} \rangle$. In particular, if p_1^* and p_1^+ denote the cost and mean RP's corresponding to $\langle \mathbf{R}_1 \rangle$, where $p_1^* = P(m|\langle \mathbf{R}_1 \rangle) = \boldsymbol{\phi}_1^{*\prime}\mathbf{R}_1$, $p_1^+ = P(1|\langle \mathbf{R}_1 \rangle) = \boldsymbol{\phi}_1^{+\prime}\mathbf{R}_1$, $\boldsymbol{\phi}_1^* = \boldsymbol{\Gamma}_{11}^{-1}\ell_{N_1}$ and $\boldsymbol{\phi}_1^+ = \boldsymbol{\Gamma}_{11}^{-1}\boldsymbol{\nu}_1$, mean-variance spanning of $\mathbf{R}$ by $\mathbf{R}_1$ is equivalent to $p^* = p_1^*$ and $p^+ = p_1^+$.

If $\langle \mathbf{R}_1 \rangle$ and $\langle \mathbf{R} \rangle$ only share the same mean RP, and $\textsc{a} \neq 0$, then the two RMVF's are tangent at the point that corresponds to the return associated with this portfolio. 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 Figures 1a and 1b). On the other hand, if $\langle \mathbf{R}_1 \rangle$ and $\langle \mathbf{R} \rangle$ only share the same cost RP, 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 Figures 2a and 2b). Thus, if we add both conditions, the old and new frontiers will be equal.

In order to implement our econometric tests for spanning, it is convenient to write the definitions of p^* and p^+ in (2) in terms of the following moment conditions:

$$E\begin{pmatrix} \mathbf{R}_t\mathbf{R}_t'\boldsymbol{\phi}^+ - \mathbf{R}_t \\ \mathbf{R}_t\mathbf{R}_t'\boldsymbol{\phi}^* - \ell_N \end{pmatrix} = E[\mathbf{h}_U(\mathbf{R}_t;\boldsymbol{\phi})] = \mathbf{0}, \tag{5}$$

[8] A 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 $\boldsymbol{\phi} = (\boldsymbol{\phi}^{+\prime}, \boldsymbol{\phi}^{*\prime})'$. In this context, spanning imposes the $2N_2$ homogeneous parametric restrictions $H_0 : \boldsymbol{\phi}_2^+ = \mathbf{0}$, $\boldsymbol{\phi}_2^* = \mathbf{0}$, where we have partition $\boldsymbol{\phi}^+$ and $\boldsymbol{\phi}^*$ conformably with $\mathbf{R}_1$ and $\mathbf{R}_2$. Hence, we can test for spanning by using the trinity of GMM asymptotic tests discussed in section 3.1. But since the moment conditions defining $\boldsymbol{\phi}^*$ and $\boldsymbol{\phi}^+$ are exactly identified, the distance metric test will coincide with the overidentifying restrictions test. In addition, all the tests can be made numerically identical by using a common estimator of the asymptotic covariance matrix of $\sqrt{T}\bar{\mathbf{h}}_{UT}(\boldsymbol{\phi}^0)$, because both the moment conditions and the restrictions to test are linear in the parameters (see Newey and West (1987)).

Such a linearity also implies that we can obtain simple closed-form solutions for the unrestricted and restricted GMM estimators of $\boldsymbol{\phi}$. Given that the moment conditions (5) are exactly identified, the former is $\hat{\boldsymbol{\phi}}_T = \bar{\mathbf{D}}_{UT}^{-1} \cdot \bar{\mathbf{d}}_T$ for large enough T, where $\bar{\mathbf{d}}_T = (\hat{\boldsymbol{\nu}}_T', \ell_N')'$, $\hat{\boldsymbol{\nu}}_T = T^{-1}\sum_{t=1}^T \mathbf{R}_t$, $\bar{\mathbf{D}}_{UT} = \mathbf{I}_2 \otimes \hat{\boldsymbol{\Gamma}}_T$ and:

$$\hat{\boldsymbol{\Gamma}}_T = \frac{1}{T}\sum_{t=1}^{T} \begin{pmatrix} \mathbf{R}_{1t}\mathbf{R}_{1t}' & \mathbf{R}_{1t}\mathbf{R}_{2t}' \\ \mathbf{R}_{2t}\mathbf{R}_{1t}' & \mathbf{R}_{2t}\mathbf{R}_{2t}' \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{\Gamma}}_{11T} & \hat{\boldsymbol{\Gamma}}_{21T}' \\ \hat{\boldsymbol{\Gamma}}_{21T} & \hat{\boldsymbol{\Gamma}}_{22T} \end{pmatrix}.$$

On the other hand, if we impose the null hypothesis on the moment conditions (5), then we will be left with the overidentified system:

$$E\begin{pmatrix} \mathbf{R}_t\mathbf{R}_{1t}'\boldsymbol{\phi}_1^+ - \mathbf{R}_t \\ \mathbf{R}_t\mathbf{R}_{1t}'\boldsymbol{\phi}_1^* - \ell_N \end{pmatrix} = \mathbf{0}. \tag{6}$$

As a result, the optimal restricted GMM estimator of $\boldsymbol{\phi}$ from (6) will be $\tilde{\boldsymbol{\phi}}_{2T} = \mathbf{0}$ and

$$\tilde{\boldsymbol{\phi}}_{1T}[\bar{\mathbf{S}}_{UT}^{-1}(\dot{\boldsymbol{\phi}}_T)] = \left\{[\mathbf{I}_2 \otimes (\hat{\boldsymbol{\Gamma}}_{11T}, \hat{\boldsymbol{\Gamma}}_{21T}')]\bar{\mathbf{S}}_{UT}^{-1}(\dot{\boldsymbol{\phi}}_T)\left[\mathbf{I}_2 \otimes \begin{pmatrix} \hat{\boldsymbol{\Gamma}}_{11T} \\ \hat{\boldsymbol{\Gamma}}_{21T} \end{pmatrix}\right]\right\}^{-1}$$
$$\times\{[\mathbf{I}_2 \otimes (\hat{\boldsymbol{\Gamma}}_{11T}, \hat{\boldsymbol{\Gamma}}_{21T}')]\bar{\mathbf{S}}_{UT}^{-1}(\dot{\boldsymbol{\phi}}_T)\bar{\mathbf{d}}_T\},$$

where $\mathbf{I}_2 \otimes (\hat{\boldsymbol{\Gamma}}_{11T}, \hat{\boldsymbol{\Gamma}}_{21T}')$ is the sample analogue of the Jacobian of (6) with respect to $\boldsymbol{\phi}_1 = (\boldsymbol{\phi}_1^{+\prime}, \boldsymbol{\phi}_1^{*\prime})'$, and $\bar{\mathbf{S}}_{UT}(\dot{\boldsymbol{\phi}}_T)$ is some HAC estimator of the optimal weighting matrix obtained with a preliminary consistent estimator $\dot{\boldsymbol{\phi}}_T$. Although the choice of $\dot{\boldsymbol{\phi}}_T$ 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)). The following result justifies an obvious first-step estimator:

Lemma 3 *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 (6) that provide the most efficient estimators of $\boldsymbol{\phi}_1^+$ and $\boldsymbol{\phi}_1^*$ under $H_0 : \boldsymbol{\phi}_2^+ = \mathbf{0}$, $\boldsymbol{\phi}_2^* = \mathbf{0}$ will be given by*

$$E\begin{pmatrix} \mathbf{R}_{1t}\mathbf{R}_{1t}'\boldsymbol{\phi}_1^+ - \mathbf{R}_{1t} \\ \mathbf{R}_{1t}\mathbf{R}_{1t}'\boldsymbol{\phi}_1^* - \ell_{N_1} \end{pmatrix} = E[\mathbf{h}_{U1}(\mathbf{R}_t;\boldsymbol{\phi})] = \mathbf{0},$$

so that $\bar{\boldsymbol{\phi}}_{1T}^+ = \hat{\boldsymbol{\Gamma}}_{11T}^{-1}\hat{\boldsymbol{\nu}}_{1T}$ and $\bar{\boldsymbol{\phi}}_{1T}^* = \hat{\boldsymbol{\Gamma}}_{11T}^{-1}\ell_{N_1}$, where $\hat{\boldsymbol{\nu}}_{1T} = T^{-1}\sum_{t=1}^T \mathbf{R}_{1t}$.

Intuitively, this means that under those circumstances, the blocks involving $\mathbf{R}_{1t}$ exactly identify the parameters $\boldsymbol{\phi}_1^*$ and $\boldsymbol{\phi}_1^+$, while the blocks corresponding to $\mathbf{R}_{2t}$ 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 mention that $\bar{\boldsymbol{\phi}}_{1T}^+$ and $\bar{\boldsymbol{\phi}}_{1T}^*$ 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).

4.2 Centred cost and mean representing portfolios

As we discussed in section 2.1, we can define an alternative pair of mean and cost RP's, $p^{++} = \mathbf{R}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu}$ and $p^{**} = \mathbf{R}'\boldsymbol{\Sigma}^{-1}\ell_N$, respectively, in terms of central moments in the absence of a safe asset. Since these portfolios also span both SMVF and RMVF, we can also test for spanning by checking that p_1^{++} and p_1^{**} coincide with p^{++} and p^{**}.

The graphical implication of sharing the centred mean RP has already been explained in section 4.1 in terms of R^+ when $\mathrm{A} \neq 0$, because p^{++} is proportional to p^+ (see Figures 2a and 2b). 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 Figures 3a and 3b). Hence, if we add both conditions, it is once more clear that the original and expanded frontiers must be equal.

The use of central moments, though, implies that we must explicitly define $\boldsymbol{\nu}$ to estimate $\boldsymbol{\Sigma}$. The simplest way to do so is to add the moment conditions that exactly identify these parameters. Hence, the extended set of moment conditions will be

$$E\begin{bmatrix} \mathbf{R}_t - \boldsymbol{\nu} \\ (\mathbf{R}_t - \boldsymbol{\nu})(\mathbf{R}_t - \boldsymbol{\nu})'\boldsymbol{\varphi}^+ - \mathbf{R}_t \\ (\mathbf{R}_t - \boldsymbol{\nu})(\mathbf{R}_t - \boldsymbol{\nu})'\boldsymbol{\varphi}^* - \ell_N \end{bmatrix} = E\begin{bmatrix} \mathbf{h}_M(\mathbf{R}_t;\boldsymbol{\nu}) \\ \mathbf{h}_C(\mathbf{R}_t,\boldsymbol{\varphi},\boldsymbol{\nu}) \end{bmatrix} = E\left[\mathbf{h}_E(\mathbf{R}_t;\boldsymbol{\varphi},\boldsymbol{\nu})\right] = \mathbf{0}, \quad (7)$$

where $\boldsymbol{\varphi} = (\boldsymbol{\varphi}^{+\prime}, \boldsymbol{\varphi}^{*\prime})'$. Partitioning $\boldsymbol{\varphi}$ conformably with $\mathbf{R}_{1t}$ and $\mathbf{R}_{2t}$, the parametric restrictions to test become $H_0 : \boldsymbol{\varphi}_2^+ = \mathbf{0}$, $\boldsymbol{\varphi}_2^* = \mathbf{0}$.

Although the approaches based on the uncentred and centred RP's look similar, there are three important differences between the moment conditions (5) in the previous section and these ones. The first two are that (7) is no longer linear in parameters, and that some of those parameters can be regarded as "nuisance". The third one, which is far less obvious but has more serious consequences, is made explicit in the following result:

Proposition 3 *Let* $\boldsymbol{\Pi}_E(\boldsymbol{\varphi}, \boldsymbol{\nu}) = (\boldsymbol{\varphi}^{*\prime}, \boldsymbol{\varphi}^{*\prime}, -\boldsymbol{\varphi}^{+\prime})'$. *Then,* $\boldsymbol{\Pi}'_E(\boldsymbol{\varphi}, \boldsymbol{\nu})\mathbf{h}_E(\mathbf{R}_t; \boldsymbol{\varphi}, \boldsymbol{\nu}) = \mathbf{0}\ \forall \mathbf{R}_t$ *if and only if* $\mathbf{m}_{E\ominus}(\boldsymbol{\varphi}, \boldsymbol{\nu}) = \boldsymbol{\varphi}^{*\prime}\boldsymbol{\nu} - \boldsymbol{\varphi}^{+\prime}\ell_N = \boldsymbol{\varphi}_\ominus^+ = \mathbf{0}$.

Given that $\boldsymbol{\varphi}_\ominus^{+0} = \mathbf{m}_{E\ominus}(\boldsymbol{\varphi}^0, \boldsymbol{\nu}^0)$ is 0 in view of (2), Proposition 3 implies that the rank of the asymptotic covariance matrix of $\sqrt{T}\bar{\mathbf{h}}_{ET}(\boldsymbol{\varphi}^0, \boldsymbol{\nu}^0)$, $\mathbf{S}_E(\boldsymbol{\varphi}^0, \boldsymbol{\nu}^0)$ say, is $3N - 1$.

But since the above moment conditions are exactly identified under the alternative hypothesis, this singularity does not affect the unrestricted GMM estimators of $\boldsymbol{\nu}$ and $\boldsymbol{\varphi}$, which will be given by $\hat{\boldsymbol{\nu}}_T$ and $\hat{\boldsymbol{\varphi}}_T = \bar{\mathbf{D}}_{C\varphi T}^{-1} \cdot \bar{\mathbf{d}}_T$, where $\bar{\mathbf{D}}_{C\varphi T} = \mathbf{I}_2 \otimes \hat{\boldsymbol{\Sigma}}_T$ and

$$\hat{\boldsymbol{\Sigma}}_T = \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} (\mathbf{R}_{1t} - \hat{\boldsymbol{\nu}}_{1T})(\mathbf{R}'_{1t} - \hat{\boldsymbol{\nu}}'_{1T}) & (\mathbf{R}_{1t} - \hat{\boldsymbol{\nu}}_{1T})(\mathbf{R}'_{2t} - \hat{\boldsymbol{\nu}}'_{2T}) \\ (\mathbf{R}_{2t} - \hat{\boldsymbol{\nu}}_{2T})(\mathbf{R}'_{1t} - \hat{\boldsymbol{\nu}}'_{1T}) & (\mathbf{R}_{2t} - \hat{\boldsymbol{\nu}}_{2T})(\mathbf{R}'_{2t} - \hat{\boldsymbol{\nu}}'_{2T}) \end{bmatrix} = \begin{pmatrix} \hat{\boldsymbol{\Sigma}}_{11T} & \hat{\boldsymbol{\Sigma}}'_{21T} \\ \hat{\boldsymbol{\Sigma}}_{21T} & \hat{\boldsymbol{\Sigma}}_{22T} \end{pmatrix}.$$

In addition, it is easy to prove that the joint asymptotic covariance matrix of $\hat{\boldsymbol{\varphi}}_{2T}^+$ and $\hat{\boldsymbol{\varphi}}_{2T}^*$ is not singular, despite the fact that the joint asymptotic distribution of $\hat{\boldsymbol{\varphi}}_T$ and $\hat{\boldsymbol{\nu}}_T$ will be so. Therefore, we can compute a Wald test based on those unrestricted estimators.

However, the singularity described in Proposition 3 does affect the optimal restricted GMM estimator, which must be carefully defined in order to take into account the information implicit in the relationship $\mathbf{m}_E(\boldsymbol{\varphi}^0, \boldsymbol{\nu}^0) = 0$. To do so, it is convenient to write the overidentified moment conditions under the null of spanning:

$$E\begin{bmatrix} \mathbf{R}_t - \boldsymbol{\nu} \\ (\mathbf{R}_t - \boldsymbol{\nu})(\mathbf{R}_{1t} - \boldsymbol{\nu}_1)'\boldsymbol{\varphi}_1^+ - \mathbf{R}_t \\ (\mathbf{R}_t - \boldsymbol{\nu})(\mathbf{R}_{1t} - \boldsymbol{\nu}_1)'\boldsymbol{\varphi}_1^* - \ell_N \end{bmatrix} = \mathbf{0}. \tag{8}$$

The optimal GMMS procedure in section 3.2 implies that we must first reparametrise $\mathbf{h}_E(\mathbf{R}_t; \boldsymbol{\varphi}, \boldsymbol{\nu})$, for instance in terms of $\boldsymbol{\nu}_\oplus$, $\boldsymbol{\varphi}_{1\ominus}^+$, $\boldsymbol{\varphi}_{1\oplus}^+$ and $\boldsymbol{\varphi}_{1\oplus}^*$, where $\boldsymbol{\nu}_\oplus = \boldsymbol{\nu}$, $\boldsymbol{\varphi}_{1\ominus}^+ = \boldsymbol{\varphi}_1^{*\prime}\boldsymbol{\nu}_1 - \boldsymbol{\varphi}_1^{+\prime}\ell_{N_1}$, $\boldsymbol{\varphi}_{1\oplus}^+$ contains the last $N_1 - 1$ elements of $\boldsymbol{\varphi}_1^+$, and $\boldsymbol{\varphi}_{1\oplus}^* = \boldsymbol{\varphi}_1^*$. Then, we should impose the singularity constraint $\boldsymbol{\varphi}_{1\ominus}^+ = 0$, and finally estimate the remaining parameters by using a consistent estimator of the Moore-Penrose inverse of $\mathbf{S}_E(\boldsymbol{\varphi}^0, \boldsymbol{\nu}^0)$, which effectively eliminates the singular linear combination of $\mathbf{h}_E(\mathbf{R}_t; \boldsymbol{\varphi}, \boldsymbol{\nu})$. As discussed in that section, though, in order to obtain a consistent estimator of $\mathbf{S}_E^+(\boldsymbol{\varphi}^0, \boldsymbol{\nu}^0)$, we need a consistent estimator of $\mathbf{S}_E(\boldsymbol{\varphi}^0, \boldsymbol{\nu}^0)$ that is singular in finite samples. The following result justifies an obvious first-step estimator:

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 (8) that provide the most efficient estimators of $\boldsymbol{\phi}_1^+$ and $\boldsymbol{\phi}_1^*$ under $H_0 : \boldsymbol{\varphi}_2^+ = \mathbf{0}$, $\boldsymbol{\varphi}_2^* = \mathbf{0}$ will be given by*

$$E \begin{bmatrix} \mathbf{R}_{1t} - \boldsymbol{\nu}_1 \\ (\mathbf{R}_{1t} - \boldsymbol{\nu}_1)(\mathbf{R}_{1t} - \boldsymbol{\nu}_1)'\boldsymbol{\varphi}_1^+ - \mathbf{R}_{1t} \\ (\mathbf{R}_{1t} - \boldsymbol{\nu}_1)(\mathbf{R}_{1t} - \boldsymbol{\nu}_1)'\boldsymbol{\varphi}_1^* - \ell_{N_1} \end{bmatrix} = E[\mathbf{h}_{E1}(\mathbf{R}_t, \boldsymbol{\varphi}, \boldsymbol{\nu})] = \mathbf{0},$$

so that $\bar{\boldsymbol{\nu}}_{1T} = \hat{\boldsymbol{\nu}}_{1T}$, $\bar{\boldsymbol{\varphi}}_{1T}^+ = \hat{\boldsymbol{\Sigma}}_{11T}^{-1}\hat{\boldsymbol{\nu}}_{1T}$ and $\bar{\boldsymbol{\varphi}}_{1T}^* = \hat{\boldsymbol{\Sigma}}_{11T}^{-1}\ell_{N_1}$.

Intuitively, this means that under those circumstances, the blocks involving $\mathbf{R}_{1t}$ exactly identify $\boldsymbol{\nu}_1$, $\boldsymbol{\varphi}_1^+$ and $\boldsymbol{\varphi}_1^*$, while the blocks corresponding to $\mathbf{R}_{2t}$ 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 3 does indeed guarantee that $\bar{\mathbf{S}}_E(\bar{\boldsymbol{\varphi}}_T, \hat{\boldsymbol{\nu}}_T)$ will be singular because the linear combination defined in Proposition 3 only involves $\mathbf{h}_{E1}(\mathbf{R}_t; \boldsymbol{\varphi}, \boldsymbol{\nu})$ under the null of spanning, and $\bar{\boldsymbol{\varphi}}_{1T}^{+\prime}\ell_{N_1} - \bar{\boldsymbol{\varphi}}_{1T}^{*\prime}\hat{\boldsymbol{\nu}}_{1T} = \mathbf{0}$.

Finally, given that the singularity described in Proposition 3 affects $\mathbf{h}_E(\mathbf{R}_t; \boldsymbol{\varphi}, \boldsymbol{\nu})$ but not $\mathbf{h}_C(\mathbf{R}_t; \boldsymbol{\varphi}, \boldsymbol{\nu})$, and that $\hat{\boldsymbol{\nu}}_T$ is the GMM estimator of the expected returns based on the moment conditions $E[\mathbf{h}_M(\mathbf{R}_t; \boldsymbol{\nu})] = \mathbf{0}$ alone, an alternative approach that avoids the singularity of $\mathbf{S}_E(\boldsymbol{\varphi}^0, \boldsymbol{\nu}^0)$ in this context would be to use a sequential GMM (SGMM) estimator which replaces $\boldsymbol{\nu}$ by $\hat{\boldsymbol{\nu}}_T$ in $\mathbf{h}_C(\mathbf{R}_t; \boldsymbol{\varphi}, \boldsymbol{\nu})$ (see e.g. Ogaki (1993)). But since

$$\mathbf{D}_{C\boldsymbol{\nu}}(\boldsymbol{\varphi}^0, \boldsymbol{\nu}^0) = E\left[\frac{\partial \mathbf{h}_C(\mathbf{R}_t; \boldsymbol{\varphi}^0, \boldsymbol{\nu}^0)}{\partial \boldsymbol{\nu}'}\right] = E\begin{bmatrix} -(\mathbf{R}_t - \boldsymbol{\nu}^0)'\boldsymbol{\varphi}^{+0}\mathbf{I}_N - (\mathbf{R}_t - \boldsymbol{\nu}^0)\boldsymbol{\varphi}^{+0\prime} \\ -(\mathbf{R}_t - \boldsymbol{\nu}^0)'\boldsymbol{\varphi}^{*0}\mathbf{I}_N - (\mathbf{R}_t - \boldsymbol{\nu}^0)\boldsymbol{\varphi}^{*0\prime} \end{bmatrix} = \mathbf{0},$$

it is clear that $\sqrt{T}[\bar{\mathbf{h}}_{CT}(\boldsymbol{\varphi}^0, \hat{\boldsymbol{\nu}}_T) - \bar{\mathbf{h}}_{CT}(\boldsymbol{\varphi}^0, \boldsymbol{\nu}^0)] \xrightarrow{p} 0$ as $T \to \infty$. Thus, in this particular instance, it is not necessary to account for the sample variability in $\hat{\boldsymbol{\nu}}_T$ in obtaining the asymptotic covariance matrix of the unrestricted SGMM estimators, which numerically coincide with the unrestricted GMMS estimators $\hat{\boldsymbol{\varphi}}_T$. Hence, the Wald tests based on the two estimators will coincide too. In contrast, we would expect the restricted SGMM estimators to be asymptotically less efficient than the optimal restricted GMMS estimators in view of Proposition 2.

5 Two-point tests for spanning

The centred and uncentred RP's 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. In this section, we shall analyse arbitrary two-point tests for spanning in the RMVF and SMVF.

5.1 RMVF Tests

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 $(0, c^{-1})$ if and only if the intercepts in the multivariate regression of $(\mathbf{R}_2 - c^{-1}\ell_{N_2})$ on a constant and $(\mathbf{R}_1 - c^{-1}\ell_{N_1})$ are all 0.[9] Therefore, a 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. Then, the null of spanning can be written as $H_0 : \mathbf{a}(c_i) = \mathbf{0}, \mathbf{a}(c_{ii}) = \mathbf{0}$, where the regression intercepts $\mathbf{a}(c_i)$ and $\mathbf{a}(c_{ii})$ are implicitly defined by the following exactly identified $2N_2(N_1 + 1)$ moment conditions:

$$
\begin{aligned}
&E\left\{\begin{array}{c}
\begin{pmatrix} 1 \\ \mathbf{R}_{1t} - c_i^{-1}\ell_{N_1} \end{pmatrix} \otimes [(\mathbf{R}_{2t} - c_i^{-1}\ell_{N_2}) - \mathbf{a}(c_i) - \mathbf{B}(c_i)(\mathbf{R}_{1t} - c_i^{-1}\ell_{N_1})] \\
\begin{pmatrix} 1 \\ \mathbf{R}_{1t} - c_{ii}^{-1}\ell_{N_1} \end{pmatrix} \otimes [(\mathbf{R}_{2t} - c_{ii}^{-1}\ell_{N_2}) - \mathbf{a}(c_{ii}) - \mathbf{B}(c_{ii})(\mathbf{R}_{1t} - c_{ii}^{-1}\ell_{N_1})]
\end{array}\right\} \\
&= E\left\{\mathbf{h}_L[\mathbf{R}_t; \mathbf{a}(c_i), \mathbf{b}(c_i), \mathbf{a}(c_{ii}), \mathbf{b}(c_{ii})]\right\} = \mathbf{0}. \qquad (9)
\end{aligned}
$$

with $\mathbf{b}(c) = vec[\mathbf{B}(c)]$. But as pointed out by Marín (1996), the asymptotic covariance matrix of the sample analogues of (9) is singular under the null. More explicitly:

[9] If we regard c^{-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)).

Proposition 4 *Let*

$$\mathbf{\Pi}_L[\mathbf{a}(c_i),\mathbf{b}(c_i),\mathbf{a}(c_{ii}),\mathbf{b}(c_{ii})]=\begin{pmatrix}\mathbf{\Phi}'^{-1}(c_i)\otimes\mathbf{I}_{N_2}\\ -\mathbf{\Phi}'^{-1}(c_{ii})\otimes\mathbf{I}_{N_2}\end{pmatrix},\ \textit{with}\ \mathbf{\Phi}(c)=\begin{pmatrix}1 & \mathbf{0}'\\ -c^{-1}\ell_{N_1} & \mathbf{I}_{N_1}\end{pmatrix}.$$

Then, $\mathbf{\Pi}'_L[\mathbf{a}(c_i),\mathbf{b}(c_i),\mathbf{a}(c_{ii}),\mathbf{b}(c_{ii})]\mathbf{h}_L[\mathbf{R}_t;\mathbf{a}(c_i),\mathbf{b}(c_i),\mathbf{a}(c_{ii}),\mathbf{b}(c_{ii})]=\mathbf{0}\quad\forall\mathbf{R}_t$

$$\Leftrightarrow \mathbf{m}_{L\ominus}[\mathbf{a}(c_i),\mathbf{b}(c_i),\mathbf{a}(c_{ii}),\mathbf{b}(c_{ii})]=\begin{bmatrix}\mathbf{a}(c_{ii})-\mathbf{a}(c_i)-c_{ii}^{-1}\left[\mathbf{B}(c_{ii})\ell_{N_1}-\ell_{N_2}\right]\\ +c_i^{-1}\left[\mathbf{B}(c_i)\ell_{N_1}-\ell_{N_2}\right]\\ \mathbf{b}(c_{ii})-\mathbf{b}(c_i)\end{bmatrix}=\mathbf{0},$$

$$\Leftrightarrow \mathbf{a}(c_i)=\mathbf{a}+c_i^{-1}\mathbf{f},\quad \mathbf{a}(c_{ii})=\mathbf{a}+c_{ii}^{-1}\mathbf{f},\ \textit{and}\quad \mathbf{b}(c_i)=\mathbf{b}(c_{ii})=\mathbf{b}=vec(\mathbf{B}),\tag{10}$$

where $\mathbf{a}$ *is a* $N_2\times 1$ *vector of parameters,* $\mathbf{B}$ *a* $N_2\times N_1$ *matrix, and* $\mathbf{f}=\ell_{N_2}-\mathbf{B}\ell_{N_1}$.

Given that $\mathbf{m}_{L\ominus}[\mathbf{a}(c_i),\mathbf{b}(c_i),\mathbf{a}(c_{ii}),\mathbf{b}(c_{ii})]$ is 0 at the true values, Proposition 4 implies that the rank of the asymptotic covariance matrix of $\bar{\mathbf{h}}_{LT}[\mathbf{a}^0(c_i),\mathbf{b}^0(c_i),\mathbf{a}^0(c_{ii}),\ \mathbf{b}^0(c_{ii})]$ is $N_2(N_1+1)$ instead of $2N_2(N_1+1)$. In this case, though, it is possible to explicitly characterise the optimal transformation of moments and parameters proposed in section 3.2 to deal with the singular linear combinations of $\mathbf{h}_L[\mathbf{R}_t;\mathbf{a}^0(c_i),\mathbf{b}^0(c_i),\mathbf{a}^0(c_{ii}),\mathbf{b}^0(c_{ii})]$.

Proposition 5 *The optimal GMM estimators of* $\mathbf{a}(c_i),\mathbf{b}(c_i),\mathbf{a}(c_{ii}),\mathbf{b}(c_{ii})$ *based on the moment conditions (9) can be obtained through (10) from the optimal GMM estimators of* $\mathbf{a}$ *and* $\mathbf{b}$ *based on the* $N_2(N_1+1)$ *moment conditions*

$$E\left[\begin{pmatrix}1\\ \mathbf{R}_{1t}\end{pmatrix}\otimes(\mathbf{R}_{2t}-\mathbf{a}-\mathbf{B}\mathbf{R}_{1t})\right]=E[\mathbf{h}_H(\mathbf{R}_t;\mathbf{a},\mathbf{b})]=\mathbf{0}.\tag{11}$$

Therefore, it is not surprising that the unrestricted GMM estimators of $\mathbf{a}(c_i),\mathbf{b}(c_i)$, $\mathbf{a}(c_{ii})$ and $\mathbf{b}(c_{ii})$, which are well defined despite the singularity of (9) because these moment conditions are exactly identified, will be given by:

$$\hat{\mathbf{B}}_T(c_i)=\hat{\mathbf{B}}_T(c_{ii})=\hat{\mathbf{\Sigma}}_{21T}\hat{\mathbf{\Sigma}}_{11T}^{-1}=\hat{\mathbf{B}}_T,$$

$$\hat{\mathbf{a}}_T(c_i)=\hat{\mathbf{a}}_T+c_i^{-1}\hat{\mathbf{f}}_T,\qquad \hat{\mathbf{a}}_T(c_{ii})=\hat{\mathbf{a}}_T+c_{ii}^{-1}\hat{\mathbf{f}}_T,$$

$$\hat{\mathbf{a}}_T=\hat{\boldsymbol{\nu}}_{2T}-\hat{\mathbf{B}}_T\hat{\boldsymbol{\nu}}_{1T},\qquad \hat{\mathbf{f}}_T=\ell_{N_2}-\hat{\mathbf{B}}_T\ell_{N_1},$$

where $\hat{\mathbf{a}}_T$ and $\hat{\mathbf{b}}_T$ are the unrestricted GMM estimators based on (11), which coincide with the OLS estimators in the multivariate regression of $\mathbf{R}_2$ on a constant and $\mathbf{R}_1$.

Further, we can compute a Wald test of $H_0:\mathbf{a}(c_i)=\mathbf{0},\mathbf{a}(c_{ii})=\mathbf{0}$ based on $\hat{\mathbf{a}}_T(c_i)$ and $\hat{\mathbf{a}}_T(c_{ii})$ because their joint asymptotic covariance matrix is not singular despite the fact that the joint asymptotic covariance matrix of $\hat{\mathbf{a}}_T(c_i)$, $\hat{\mathbf{b}}_T(c_i)$, $\hat{\mathbf{a}}_T(c_{ii})$ and $\hat{\mathbf{b}}_T(c_{ii})$ will be so. But since $\hat{\mathbf{a}}_T(c_i)$ and $\hat{\mathbf{a}}_T(c_{ii})$ are a full-rank linear transformation of $\hat{\mathbf{a}}_T$ and $\hat{\mathbf{f}}_T$ with known coefficients, we can easily prove that such a Wald test is asymptotically equivalent to the GMM-based Wald version of the Huberman and Kandel (1987) test discussed by Ferson, Foerster and Kim (1993), which assesses whether $H_0 : \mathbf{a} = \mathbf{0}, \mathbf{f} = \mathbf{0}$.

In fact, Huberman and Kandel (1987) derived a likelihood ratio test and a related F test whose finite sample distribution is exact under the assumption that the distribution of $\mathbf{R}_{2t}$ given $\mathbf{R}_{1s}$ $(s = 1, \ldots, T)$ is multivariate normal with linear mean $\mathbf{a} + \mathbf{B}\mathbf{R}_{1t}$ and constant covariance matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$.[10] The same assumption also allowed Kan and Zhou (2001) to theoretically compare the finite sample distributions of the Wald, Lagrange multiplier and likelihood ratio versions of the Huberman and Kandel (1987) testing procedure using results in Berndt and Savin (1977). The advantage of working with a GMM framework, though, is that under fairly weak regularity conditions, the tests are robust to departures from the assumption of $i.i.d.$ Gaussian returns.

As for the optimal restricted GMMS estimators of $\mathbf{a}(c_i)$, $\mathbf{b}(c_i)$, $\mathbf{a}(c_{ii})$ and $\mathbf{b}(c_{ii})$ based on (9), it follows from Proposition 5 that they can also be obtained through (10) by minimising with respect to $\mathbf{a}$ and $\mathbf{b}$ the optimal norm of the sample analogue of $E[\mathbf{h}_H(\mathbf{R}_t; \mathbf{a}, \mathbf{b})] = \mathbf{0}$ subject to the constraints $\mathbf{a} = \mathbf{0}$ and $\mathbf{f} = \ell_{N_2} - \mathbf{B}\ell_{N_1} = \mathbf{0}$. As described in section 6.2 of Campbell, Lo and MacKinlay (1997), a numerically equivalent procedure is to minimise with respect to the elements of $\mathbf{B}_2$ the optimal norm of the sample analogues of the following unrestricted set of moment conditions

$$E\left\{\left(\begin{array}{c} 1 \\ R_{1at} \\ \mathbf{R}_{1bt}\text{-}R_{1at}\ell_{N_1-1} \end{array}\right) \otimes \left[\begin{array}{c} (\mathbf{R}_{2t}\text{-}R_{1at}\ell_{N_2}) \\ -\mathbf{B}_2(\mathbf{R}_{1bt}\text{-}R_{1at}\ell_{N_1-1}) \end{array}\right]\right\} = E[\mathbf{h}_L((\mathbf{R}_t; \mathbf{b}_2)] = \mathbf{0}, \quad (12)$$

for any choice of reference portfolio R_{1at}, where we have partitioned $\mathbf{B} = (\mathbf{b}_1, \mathbf{B}_2)$ and $\ell_{N_1} = (1, \ell'_{N_1-1})'$ conformably with $\mathbf{R}_1 = (R_{1a}, \mathbf{R}'_{1b})'$, and $\mathbf{b}_2 = vec(\mathbf{B}_2)$. In practice, we need an initial consistent estimator of $\mathbf{b}_2$ to calculate the optimal weighting matrix. Our next lemma suggests some sensible ways of doing so:

Lemma 5 *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*

[10] Nevertheless, both Peñaranda (1999) and Kan and Zhou (2001) noticed a typo in the Huberman and Kandel (1987) paper, whereby a square root is missing in the ratio of determinants of the residual variances. Kan and Zhou (2001) also stress the fact that both the test statistic and the distribution to use depend on whether N_2 is equal or greater than 1.

the linear combinations of the moment conditions (12) that provide the most efficient estimators of $\mathbf{b}_2$ *under* $H_0 : \mathbf{a} = \mathbf{0}$, $\mathbf{f} = \mathbf{0}$ *will be given by*

$$E\left\{\left[\begin{array}{c}(\mathbf{R}_{1bt}-R_{1at}\ell_{N_1-1})\\+\kappa(\boldsymbol{\nu}_{1b}-\nu_{1a}\ell_{N_1-1})\end{array}\right]\otimes\left[\begin{array}{c}(\mathbf{R}_{2t}-R_{1at}\ell_{N_2})\\-\mathbf{B}_2(\mathbf{R}_{1bt}-R_{1at}\ell_{N_1-1})\end{array}\right]\right\}=\mathbf{0}. \qquad (13)$$

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 $\mathbf{B}_2$ will come from the multivariate regression of ($\mathbf{R}_2$-$R_{1a}\ell_{N_2}$) on ($\mathbf{R}_{1b}$-$R_{1a}\ell_{N_1-1}$). Alternatively, we could use the sample analogues of $\boldsymbol{\nu}$ and κ to obtain an IV estimator of $\mathbf{B}_2$ from (13).[11] In either case, such first-step estimators will remain consistent under H_0 even if those assumptions are not totally realistic in practice.

5.2 SMVF Tests

De Santis (1993, 1995) and Bekaert and Urias (1996) were the first to develop two-point GMM-based spanning tests in the SMVF. The starting point of their suggested procedure is the pricing equation obtained by using elements of the SMVF as SDF's:

$$E[\mathbf{R}m^{MV}(c)] = cov[\mathbf{R}, m^{MV}(c)] + E(\mathbf{R})E[m^{MV}(c)] = \ell_N \quad \forall c. \qquad (14)$$

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 $E[m_1^{MV}(c)] = c$. Therefore, we can develop two-point GMM spanning tests based on the moment conditions:

$$E\left\{\begin{array}{c}(\mathbf{R}_t-\boldsymbol{\nu})[c_i+(\mathbf{R}_t-\boldsymbol{\nu})'\boldsymbol{\beta}(c_i)]+c_i\mathbf{R}_t-\ell_N\\(\mathbf{R}_t-\boldsymbol{\nu})[c_{ii}+(\mathbf{R}_t-\boldsymbol{\nu})'\boldsymbol{\beta}(c_{ii})]+c_{ii}\mathbf{R}_t-\ell_N\end{array}\right\}=E\{\mathbf{h}_S[\mathbf{R}_t;\boldsymbol{\beta}(c_i),\boldsymbol{\beta}((c_{ii}),\boldsymbol{\nu}]\}=\mathbf{0},$$

where $c_i \neq c_{ii}$ are two non-negative scalars chosen by the researcher. Unfortunately, $\boldsymbol{\nu}$ is generally unknown, so that these moment conditions are not directly testable. Once more, the simplest way to handle the estimation of $\boldsymbol{\nu}$ would be to add the moment conditions $\mathbf{h}_M(\mathbf{R}_t;\boldsymbol{\nu})$ that exactly identify $\boldsymbol{\nu}$, as in (7). In particular, inference should be based on

$$E\left[\left(\begin{array}{c}\mathbf{h}_M(\mathbf{R}_t;\boldsymbol{\nu})\\\mathbf{h}_S[\mathbf{R}_t;\boldsymbol{\beta}(c^i),\boldsymbol{\beta}(c^{ii}),\boldsymbol{\nu}]\end{array}\right)\right]=E\{\mathbf{h}_D[\mathbf{R}_t;\boldsymbol{\beta}(c^i),\boldsymbol{\beta}(c^{ii}),\boldsymbol{\nu}]\}=\mathbf{0}, \qquad (15)$$

where the restrictions to test become $H_0 : \boldsymbol{\beta}_2(c_i) = \mathbf{0}$, $\boldsymbol{\beta}_2(c_{ii}) = \mathbf{0}$, and where we have partitioned $\boldsymbol{\beta}(c) = [\boldsymbol{\beta}_1'(c), \boldsymbol{\beta}_2'(c)]'$ conformably with $\mathbf{R}_{1t}$ and $\mathbf{R}_{2t}$.

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

However, it turns out that we can write the moment conditions $\mathbf{h}_D[\mathbf{R}_t;\boldsymbol{\beta}(c_i),\boldsymbol{\beta}(c_{ii}),\boldsymbol{\nu}]$ as a full rank linear transformation with *known* coefficients of the moment conditions (7) that define the cost RP. Specifically:

$$\mathbf{h}_S[\mathbf{R}_t;\boldsymbol{\beta}(c_i),\boldsymbol{\beta}(c_{ii}),\boldsymbol{\nu}] = \left[\begin{pmatrix} -c^i & 1 \\ -c^{ii} & 1 \end{pmatrix} \otimes \mathbf{I}_N\right] \mathbf{h}_C(\mathbf{R}_t;\boldsymbol{\varphi},\boldsymbol{\nu}).$$

Hence, we can use Lemma 1 to show that these two-point SMVF tests are in all respects equivalent to our centred RP tests irrespective of the validity of the null hypothesis. In particular, the covariance matrix of $\mathbf{h}_D[\mathbf{R}_t;\boldsymbol{\beta}(c^i),\boldsymbol{\beta}(c^{ii}),\boldsymbol{\nu}]$ is singular too. Note, though, that there are no finite values of c_i and c_{ii} for which $\mathbf{h}_D[\mathbf{R}_t;\boldsymbol{\beta}(c_i),\boldsymbol{\beta}(c_{ii}),\boldsymbol{\nu}]$ reduces to $\mathbf{h}_E(\mathbf{R}_t;\boldsymbol{\varphi},\boldsymbol{\nu})$, which reflects that p^{++} does not belong to the SMVF. In addition, depending on the first-stage estimators of $\boldsymbol{\beta}(c_i)$ and $\boldsymbol{\beta}(c_{ii})$, the values of c_i and c_{ii} may numerically influence the GMM tests based on (15).[12]

In fact, De Santis (1995) and Bekaert and Urias (1997) worked with the moments

$$E\left\{\begin{array}{c} \mathbf{R}_t[c_i + (\mathbf{R}_t - \boldsymbol{\nu})'\boldsymbol{\beta}\,(c_i)] - \ell_N \\ \mathbf{R}[c_{ii} + (\mathbf{R}_t - \boldsymbol{\nu})'\boldsymbol{\beta}\,(c_{ii})] - \ell_N \end{array}\right\} = E\{\mathbf{h}_B[\mathbf{R}_t;\boldsymbol{\beta}(c_i),\boldsymbol{\beta}(c_{ii}),\boldsymbol{\nu}]\} = \mathbf{0}.$$

Nevertheless, since

$$\mathbf{h}_B[\mathbf{R}_t;\boldsymbol{\beta}(c_i),\boldsymbol{\beta}(c_{ii}),\boldsymbol{\nu}] = \begin{pmatrix} \boldsymbol{\nu}\boldsymbol{\beta}'(c_i) & \mathbf{I}_N & \mathbf{0} \\ \boldsymbol{\nu}\boldsymbol{\beta}'(c_{ii}) & \mathbf{0} & \mathbf{I}_N \end{pmatrix} \left\{\begin{array}{c} \mathbf{h}_M(\mathbf{R}_t;\boldsymbol{\nu}) \\ \mathbf{h}_S[\mathbf{R}_t;\boldsymbol{\beta}(c_i),\boldsymbol{\beta}(c_{ii}),\boldsymbol{\nu}] \end{array}\right\},$$

Lemma 1 can again be used to show that the difference between the two-point spanning tests based on $\mathbf{h}_B[\mathbf{R}_t;\boldsymbol{\beta}(c_i),\boldsymbol{\beta}(c_{ii}),\boldsymbol{\nu}]$ and $\mathbf{h}_S[\mathbf{R}_t;\boldsymbol{\beta}(c_i),\boldsymbol{\beta}(c_{ii}),\boldsymbol{\nu}]$ converges in probability to 0 as $T \to \infty$ irrespective of the validity of the null hypothesis. In this context, an advantage of working with $\mathbf{h}_S[\mathbf{R}_t;\boldsymbol{\beta}(c_i),\boldsymbol{\beta}(c_{ii}),\boldsymbol{\nu}]$ instead of $\mathbf{h}_B[\mathbf{R}_t;\boldsymbol{\beta}(c_i),\boldsymbol{\beta}(c_{ii}),\boldsymbol{\nu}]$ is that sequential GMM can be applied without the need to make any adjustment to the estimators of the asymptotic covariance of $\bar{\mathbf{h}}_{ST}[\boldsymbol{\beta}(c_i),\boldsymbol{\beta}(c_{ii}),\hat{\boldsymbol{\nu}}_T]$ in order to reflect the sample variability in $\hat{\boldsymbol{\nu}}_T$ for the reasons explained at the end of section 4.2.

More recently, Kan and Zhou (2001) have discussed an alternative two-point spanning test for the SMVF frontier. Specifically, they suggest to use the expressions obtained in section 2.2 to reparametrise $m^{MV}(c)$ in terms of α instead of c as:

$$m^{MV}(\alpha) = \alpha + \mathbf{R}'\boldsymbol{\beta}[c(\alpha)] = \alpha + \mathbf{R}'\boldsymbol{\gamma}(\alpha),$$

[12] The second part of Lemma 1 provides sufficient conditions for the equality of explicit estimators of $\boldsymbol{\beta}(c_i)$ and $\boldsymbol{\beta}(c_{ii})$ obtained from $\mathbf{h}_D[\mathbf{R}_t;\boldsymbol{\beta}(c_i),\boldsymbol{\beta}(c_{ii}),\boldsymbol{\nu}]$, and implicit estimators obtained from estimators of $\boldsymbol{\varphi}$ based on (7) through the theoretical relationship $\boldsymbol{\beta}(c_i) = \boldsymbol{\varphi}^* - c_i^+\boldsymbol{\varphi}^+$.

where $c(\alpha) = (\alpha + \mathrm{A})/(1 + \mathrm{B})$ and $\boldsymbol{\gamma}(\alpha) = \boldsymbol{\Gamma}^{-1}(\ell_N - \alpha\boldsymbol{\nu}) = \boldsymbol{\phi}^* - \alpha\boldsymbol{\phi}^+$. Given the properties of the SMVF, though, it is again clear that there will be spanning if and only if the above condition is satisfied for any two distinct $\alpha's$, which we shall call α_i and α_{ii}. Therefore, the moment conditions that Kan and Zhou (2001) analyse are

$$E\left\{\begin{array}{c} \mathbf{R}_t[\alpha_i + \mathbf{R}_t'\boldsymbol{\gamma}(\alpha_i)] - \ell_N \\ \mathbf{R}_t[\alpha_{ii} + \mathbf{R}_t'\boldsymbol{\gamma}(\alpha_{ii})] - \ell_N \end{array}\right\} = E\{\mathbf{h}_K[\mathbf{R}_t; \boldsymbol{\gamma}(\alpha_i), \boldsymbol{\gamma}(\alpha_{ii})]\} = \mathbf{0}, \tag{16}$$

where $\alpha_i \neq \alpha_{ii}$ are two scalars chosen by the researcher. In this context, the null becomes $H_0 : \boldsymbol{\gamma}_2(\alpha_i) = \mathbf{0}$, $\boldsymbol{\gamma}_2(\alpha_{ii}) = \mathbf{0}$. But like in the case of the tests based on two $c's$, we can also write these estimating functions as a full rank linear transformation with known coefficients of the estimating functions that define the uncentred RP. Specifically

$$\mathbf{h}_K\left[\mathbf{R}_t; \boldsymbol{\gamma}(\alpha_i), \boldsymbol{\gamma}(\alpha_{ii})\right] = \left[\begin{pmatrix} -\alpha_i & 1 \\ -\alpha_{ii} & 1 \end{pmatrix} \otimes \mathbf{I}_N\right] \mathbf{h}_U\left(\mathbf{R}_t; \boldsymbol{\phi}\right).$$

Therefore, we can again use Lemma 1 to show that tests based on (16) would be equivalent to our uncentred RP tests irrespective of the validity of the null hypothesis. However, there is no finite value of α_i and α_{ii} for which the moment conditions $\mathbf{h}_K[\mathbf{R}_t; \boldsymbol{\gamma}(\alpha_i), \boldsymbol{\gamma}(\alpha_{ii})]$ reduce to $\mathbf{h}_U(\mathbf{R}, \boldsymbol{\phi})$, which reflects that p^+ does not belong to the SMVF either. Further, depending on the first-stage estimators of $\boldsymbol{\gamma}(\alpha_i)$ and $\boldsymbol{\gamma}(\alpha_{ii})$, the values of α^i and α^{ii} may numerically influence the GMM tests based on (16).[13]

For all these reasons, in what follows we shall not separately discuss the different two-point SMVF tests, concentrating instead on the centred and uncentred RP tests.

6 Asymptotic comparisons of spanning tests

So far, we have presented three separate families of spanning tests: centred and uncentred RP's, and regression versions. In this section, we shall extensively compare them.

6.1 Equivalence of the parametric restrictions

As we have already seen, the parametric restrictions involved in the novel testing procedures proposed in section 4 simply mean that the centred or uncentred cost and mean RP's of $\mathbf{R}_t$ do not depend on $\mathbf{R}_{2t}$. Given that the SMVF is spanned by either

[13] Lemma 1 also provides sufficient conditions for the numerical equality of explicit estimators of $\boldsymbol{\gamma}(\alpha_i)$ and $\boldsymbol{\gamma}(\alpha_{ii})$ obtained from $\mathbf{h}_K[\mathbf{R}_t; \boldsymbol{\gamma}(\alpha_i), \boldsymbol{\gamma}(\alpha_{ii})]$, and implicit estimators obtained from estimators of $\boldsymbol{\phi}$ based on (5) through $\boldsymbol{\gamma}(\alpha) = \boldsymbol{\phi}^* - \alpha\boldsymbol{\phi}^+$ (see footnote 29 in Kan and Zhou (2001) for an example).

pair of RP's, 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 $\mathbf{R}_t$ does not depend on $\mathbf{R}_{2t}$. 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 $\mathbf{R}_{2t}$ can be written as a unit cost portfolio of $\mathbf{R}_{1t}$, plus an orthogonal arbitrage portfolio with zero mean.

These equivalences can be seen more formally if we write:

$$\begin{aligned} p_t^+ &= p_{1t}^+ + (1+\mathrm{B}_1)^{-1}\mathbf{a}'\mathbf{\Lambda}^{-1}\mathbf{v}_t, \\ p_t^* &= p_{1t}^* + (\ell_{N_2} - \mathbf{C}\ell_{N_1})'\,\mathbf{\Lambda}^{-1}\mathbf{v}_t, \\ p_t^{++} &= p_{1t}^{++} + (1+\mathrm{B}_1)^{-1}\mathbf{a}'\mathbf{\Omega}^{-1}\mathbf{w}_t, \\ p_t^{**} &= p_{1t}^{**} + (\ell_{N_2} - \mathbf{B}\ell_{N_1}')\mathbf{\Omega}^{-1}\mathbf{w}_t, \end{aligned} \tag{17}$$

where $\mathbf{v}_t = \mathbf{R}_{2t} - \mathbf{C}\mathbf{R}_{1t}$, $\mathbf{C} = \mathbf{\Gamma}_{21}\mathbf{\Gamma}_{11}^{-1}$ and $\mathbf{\Lambda} = \mathbf{\Gamma}_{22} - \mathbf{\Gamma}_{21}\mathbf{\Gamma}_{11}^{-1}\mathbf{\Gamma}_{21}'$ are related to the least squares projection of $\mathbf{R}_2$ on $\langle \mathbf{R}_1 \rangle$, while $\mathbf{w}_t = \mathbf{R}_{2t} - \mathbf{a} - \mathbf{B}\mathbf{R}_{1t}$, $\mathbf{B}$ and $\mathbf{\Omega}$ are related to 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

$$\mathbf{f} = (\ell_{N_2} - \mathbf{C}\ell_{N_1}) + \mathrm{A}_1\mathbf{a}/(1+\mathrm{B}_1).$$

6.2 Equivalence of the tests under the null and local alternatives

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 is the same. The purpose of this subsection is to fill the gap in the literature by investigating the asymptotic performance of the different optimal versions of all the previously discussed GMM tests under the null and sequences of local alternatives. The following proposition, which is one of the key results of our paper, provides a very precise answer:

Proposition 6 *The trinity of optimal asymptotic tests based on each of the following sets of moment conditions and restrictions:*

$$E[\mathbf{h}_U(\mathbf{R}_t;\boldsymbol{\phi})] = \mathbf{0}, \quad \mathbf{H}_0 : \boldsymbol{\phi}_2^+ = \mathbf{0}, \boldsymbol{\phi}_2^* = \mathbf{0},$$
$$E[\mathbf{h}_E(\mathbf{R}_t;\boldsymbol{\varphi},\boldsymbol{\nu})] = \mathbf{0}, \quad \mathbf{H}_\mathbf{0} : \boldsymbol{\varphi}_2^+ = \mathbf{0}, \boldsymbol{\varphi}_2^* = \mathbf{0},$$
$$E[\mathbf{h}_H(\mathbf{R}_t;\mathbf{a},\mathbf{b})] = \mathbf{0}, \quad \mathbf{H}_\mathbf{0} : \mathbf{a} = \mathbf{0}, \mathbf{f} = \mathbf{0},$$

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

Therefore, there is no basis to prefer one test to the other from this perspective because all the statistics asymptotically converge to exactly the same random variable. In this respect, note that our equivalence 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 3.1. Nevertheless, it is only valid under the null of spanning, and alternatives arbitrarily close to it.

6.3 Relative Performance under Fixed Alternatives

We are going to use Bahadur's definition of asymptotic relative efficiency (ARE) of two testing procedures as the ratio of their approximate slopes (AS), which we described in section 3.1. Although W_T, LM_T and DM_T are not necessarily equivalent in terms of AS, except of course when they are numerically equivalent, for the sake of brevity we shall only compare the approximate slopes of those versions of the Wald test statistics in which the asymptotic covariance matrix of the restrictions evaluated at the unrestricted parameter estimators has been computed by using the long-run variance of the centred second moments under the alternative, as suggested by Hall (2000).

In principle, we can use (17) to obtain the required AS expressions, which indicate that in general, the three families of spanning tests in Proposition 6 are not asymptotically equivalent under fixed alternatives. However, it is virtually impossible to compare the different AS without making specific assumptions about the true values of the parameters, and/or the distribution of returns. In this respect, we can show that:

Lemma 6 *If* $\mathbf{a} = \mathbf{0}$*, so that* $\mathbf{R}_1$ *and* $\mathbf{R}$ *share the mean RP's, then the ARE of the Wald tests based on the centred and uncentred RP's is 1 regardless of distributional assumptions.*

Additional results can be obtained when returns are *i.i.d.* elliptical:

Lemma 7 *If* $\mathbf{R}_t$ *is an i.i.d. elliptical random vector with bounded fourth moments, then the AS of the Wald version of the regression test is at least as large as the AS of the Wald version of the centred RP test regardless of the values of the parameters.*

In contrast, it is possible to find parametric configurations for which the AS of the uncentred RP test is either bigger or smaller than the AS of the GMM version of the Huberman-Kandel (1987) test. For instance, we can show that the uncentred RP test is always asymptotically more powerful than the regression test when the distribution of returns is *i.i.d.* normal and $\mathbf{C}\ell_{N_1} = \ell_{N_2}$, so that $\mathbf{R}_1$ and $\mathbf{R}$ share the uncentred cost RP.

Although these results are fairly specific, they can rationalise Monte Carlo results obtained under the commonly made assumption that $\mathbf{R}_t$ is an *i.i.d.* multivariate normal or multivariate t random vector (see e.g. Kan and Zhou (2001)).

7 Empirical Application

Although Euro notes and coins started to circulate on Jan. 1st, 2002, the third stage of the European Monetary Union (EMU) began on Jan. 1st, 1999, when the exchange rates of the participating currencies were irrevocably set. During these years, the European Commission has envisaged the creation of EMU as a cornerstone to the realisation of a Europe in which people, services, capital and goods can move freely. However, EMU is not without its costs. Specifically, it is not clear a priori that the elimination of intra-European exchange rate risk is necessarily beneficial for investors, given that it affects their opportunities for diversification. In this section, we try to indirectly shed some light on this issue by answering a related but simpler question: would the mean-variance investment opportunity set of global investors who diversify their speculative investments across the most important developed countries remain unaffected by not being able to invest in the assets of several EMU members? We concentrate on the very short end of the term structure, which is the only case in which a truly PanEuropean integrated financial market has been created under the form of an interbank money market.

Our data consists of US dollar prices of Eurodeposits for 1 week, as well as 1 and 3 months from Jan. 4th, 1984 to Dec. 27th, 1995 for Canada, Japan, Switzerland, the UK, the US, Germany, Belgium, France and Italy,[14] which we transform in weekly (Wednesday to Wednesday) returns. The reason why we stop our sample a few years before the actual

[14] The Eurodeposit data comes from Datastream. The typical code is ECXXXYY, where XXX denotes the currency and YY the term. Specifically, CAD stands for Canadian dollar, JAP for Japanese yen, SWF for Swiss franc, UKP for British pound, USD for US dollar, WGM for Deutsche Mark, BFR for Belgian franc, FFR for French franc and ITL for Italian lira. Similarly, 1W, 1M or 3M stand for 1-week, 1-month or 3-month rates, respectively. The exchange rate data are taken from the Bank of Spain.

creation of EMU is to avoid biasing our results in favour of the null hypothesis by using data over a period in which there was a rapid convergence of the short end of the term structure of the likely candidate members towards German levels. In this respect, we take Germany as our representative EMU country, and consider the effects on the mean-variance frontiers of excluding from the asset base the other three EMU countries.

The three families of spanning tests are reported in Table 1a. Specifically, we have computed the centred and uncentred RP tests introduced in section 4, together with the GMM version of the regression test discussed in section 5.1. To keep the ratio of assets to observations small, we have only used the 1-week and 3-month rates, although qualitatively similar results are obtained by also considering data on 1-month Eurodeposits. As can been seen, our results are not sensitive either to the choice of test family, or weighting matrix, and clearly reject the null hypothesis of spanning.[15] In this respect, Table 1b contains the uncentred RP tests on a country by country basis, which suggest that the evidence against spanning seems to be much higher for Belgium or Italy than for France.

Therefore, we can fairly confidently argue that during the second half of the 1980's, and the first half of the 1990's, a global investor with speculative, short-term positions was better off by investing not only in the money markets of Germany and other major developed economies, but also in the Belgian, French and Italian money markets.

8 Conclusions

We have proposed a unifying approach to test for spanning in the return and stochastic discount factor mean-variance frontiers, which is not tied down to the properties of either frontier. Specifically, given that the uncentred cost and mean RP's introduced by Chamberlain and Rothschild (1983) span both frontiers, our testing procedure is based on assessing if these two portfolios remain the same when we increase the number of assets that we analyse. Since those RP's are defined in terms of uncentred moment conditions, GMM can be directly applied for testing without the need for nuisance parameters.

We have also proposed analogous spanning tests based on the centred cost and mean

[15] For illustrative purposes, we have computed the asymptotic slopes of the three testing families along the lines of section 6.3, under the maintained assumption that returns are *i.i.d.* elliptical. If we replace the first and second population moments of returns, together with the coefficient of multivariate excess kurtosis, by their sample counterparts, then we find that the asymptotic slopes are .0762 and .0707 for the uncentred and centred RP tests, respectively, and .0941 for the regression test.

RP's suggested by the same authors, which require the introduction of additional moment conditions that define mean returns as nuisance parameters. However, since this results in an unusual GMM framework, we have extended the theory of optimal GMM inference in Hansen (1982) to those non-trivial situations in which the estimating functions have a singular covariance matrix along an implicit manifold in the parameter space that contains the true value. For the benefit of practitioners, we have suggested sensible consistent first-step parameter estimators that can be used to obtain feasible versions of the optimal GMM estimators with potentially better finite sample properties. The choice of first-step estimator is of the utmost importance in our singular GMM set-up to avoid asymptotic discontinuities in the distributions of the estimators and testing procedures.

We have related our proposed tests to the existing ones, and showed that they can be grouped in three families: our two RP tests, and the regression tests introduced by Huberman and Kandel (1987). We have also proved that their parametric restrictions are equivalent, and more importantly, that all the tests are asymptotically equivalent under the null and compatible sequences of local alternatives. The latter result has been obtained under fairly weak assumptions on the distribution of asset returns. In particular, we do not require that they are *i.i.d.* Gaussian or elliptical random vectors. Moreover, we have compared the asymptotic power of the three families of spanning tests against fixed alternatives by using Bahadur's notion of asymptotic relative efficiency, and obtained some specific results for certain parameter configurations and commonly made assumptions on distributions. However, since our comparisons rely on asymptotic results, they have little to say about the small sample performance of the different tests (see Bekaert and Urias (1996) or Kan and Zhou (2001) for some Monte Carlo evidence on these issues). Therefore, it would be useful to obtain higher-order expansions of all the test statistics, which, however, are beyond the scope of this paper.

Finally, we have applied these procedures to the Eurodeposit market, with special emphasis on the recently created interbank money market for the Euro zone, and concluded that the during the second half of the 80's, and the first half of the 90's, a global investor with speculative, short-term positions was better off by investing in the money markets of Belgium, France and Italy, as well as Germany and other major developed countries.

There are three situations in which the structure of the RMVF and SMVF imply 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 also 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 unconditional 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 spanning under more general expected utility functions (see also Gouriéroux and Monfort (2001)), as well as nontraded assets.

We have not discussed either spanning tests in the conditional versions of the RMVF or SMVF (see Hansen and Richard (1987) and Gallant, Hansen and Tauchen (1990), respectively). This issue is partly addressed in De Santis (1995), Bekaert and Urias (1996), De Roon, Nijman, and Werker (1997) or Sentana (2004) by scaling returns with instruments, which can be interpreted as the payoffs to managed portfolios. Alternative partial approaches are discussed by Ferson, Foerster, and Keim (1993) and Cochrane (2001) (see also De Roon and Nijman (2001)). Given that Hansen and Richard (1987) derive conditional analogues to the centred and uncentred RP's, our unifying approach provides a rather natural and comprehensive way to test for spanning in those situations. However, since the weights of the conditional mean and cost RP portfolios will generally be functions of the relevant information set, the conditional analogues to our spanning tests should be conditional moment tests, as opposed to the parametric restrictions tests based on unconditional moments considered so far in the literature.

Finally, spanning tests are partly related to mutual fund separation. In fact, the only additional restriction in a RMVF 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 e.g. Chamberlain (1983) or Ferson, Foerster, and Keim (1993)). However, testing for mean independence also involves conditional moment restrictions, which is again qualitatively different from a standard parametric test. Given the practical relevance of all these issues, though, they constitute obvious avenues for further research.

Appendix

A Proofs of Propositions

Proposition 1:

Given that $\mathbf{P}(\boldsymbol{\theta}^0)$ is an orthogonal matrix, and $\mathbf{m}(\boldsymbol{\theta})$ is regular in an open neighbourhood of $\boldsymbol{\theta}^0$, so that $\text{rank}[\mathbf{L}(\boldsymbol{\theta}^0_\oplus, \mathbf{0})] = k$ by the inverse function theorem, then it follows that $\text{rank}[\mathbf{D}(\boldsymbol{\theta}^0)] = \text{rank}[\mathbf{P}'(\boldsymbol{\theta}^0)\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}(\boldsymbol{\theta}^0_\oplus, \mathbf{0})]$, which in turn equals

$$\text{rank}\begin{bmatrix} \mathbf{P}'_\oplus(\boldsymbol{\theta}^0)\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus, \mathbf{0}) & \mathbf{P}'_\oplus(\boldsymbol{\theta}^0)\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\theta}_\ominus}(\boldsymbol{\theta}^0_\oplus, \mathbf{0}) \\ \mathbf{P}'_\ominus(\boldsymbol{\theta}^0)\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus, \mathbf{0}) & \mathbf{P}'_\ominus(\boldsymbol{\theta}^0)\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\theta}_\ominus}(\boldsymbol{\theta}^0_\oplus, \mathbf{0}) \end{bmatrix} = k. \tag{A1}$$

Now, $\boldsymbol{\Pi}'[\mathbf{l}(\boldsymbol{\theta}_\oplus, \mathbf{0})]\mathbf{h}[\mathbf{x}_t; \mathbf{l}(\boldsymbol{\theta}_\oplus, \mathbf{0})] = \mathbf{0}\ \forall t$ from Assumption 1. If we differentiate this equation with respect to $\boldsymbol{\theta}_\oplus$, and evaluate the derivatives at $\boldsymbol{\theta}^0_\oplus$, we will have

$$\begin{aligned} & \{\mathbf{h}[\mathbf{x}_t; \mathbf{l}(\boldsymbol{\theta}^0_\oplus, \mathbf{0})] \otimes \mathbf{I}_{k_\ominus}\}\frac{\partial vec\{\boldsymbol{\Pi}'[\mathbf{l}(\boldsymbol{\theta}^0_\oplus, \mathbf{0})]\}}{\partial \boldsymbol{\theta}_\oplus} + \boldsymbol{\Pi}'(\boldsymbol{\theta}^0)\frac{\partial \mathbf{h}[\mathbf{x}_t; \mathbf{l}(\boldsymbol{\theta}^0_\oplus, \mathbf{0})]}{\partial \boldsymbol{\theta}_\oplus} \\ = \ & [\mathbf{h}(\mathbf{x}_t; \boldsymbol{\theta}^0) \otimes \mathbf{I}_{k_\ominus}]\frac{\partial vec[\boldsymbol{\Pi}'(\boldsymbol{\theta}^0)]}{\partial \boldsymbol{\theta}'}\mathbf{L}_\oplus(\boldsymbol{\theta}^0_\oplus, \mathbf{0}) + \boldsymbol{\Pi}'(\boldsymbol{\theta}^0)\frac{\partial \mathbf{h}(\mathbf{x}_t; \boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}'}\mathbf{L}_\oplus(\boldsymbol{\theta}^0_\oplus, \mathbf{0}) = \mathbf{0}, \end{aligned}$$

since $\boldsymbol{\theta}^0 = \mathbf{l}(\boldsymbol{\theta}^0_\oplus, \mathbf{0})$ from Assumption 2. If we take expectations of this expression, and bearing in mind that $E[\mathbf{h}(\mathbf{x}_t; \boldsymbol{\theta}^0)] = \mathbf{0}$ by definition of $\boldsymbol{\theta}^0$, then we can easily show that $\boldsymbol{\Pi}'(\boldsymbol{\theta}^0)\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus, \mathbf{0}) = \mathbf{0}$. Finally, given that $\mathbf{P}_\ominus(\boldsymbol{\theta}^0)$ must be a full-column rank linear transformation of $\boldsymbol{\Pi}(\boldsymbol{\theta}^0)$ because $\mathbf{P}'_\ominus(\boldsymbol{\theta}^0)\mathbf{h}(\mathbf{x}_t; \boldsymbol{\theta}^0) = \mathbf{0}\ \forall t$, we can also show that

$$\mathbf{P}'_\ominus(\boldsymbol{\theta}^0)\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus, \mathbf{0}) = \mathbf{0}, \tag{A2}$$

As a result, $\text{rank}[\mathbf{P}'_\oplus(\boldsymbol{\theta}^0)\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus, \mathbf{0})]$ must indeed be $k_\oplus$ for (A1) to be true. $\square$

Proposition 2:

Let us start by computing the joint asymptotic covariance matrix of $\sqrt{T}\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\bar{\mathbf{h}}_T(\boldsymbol{\theta}^0)$ and $\sqrt{T}\mathbf{L}'_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus, \mathbf{0})\mathbf{D}'(\boldsymbol{\theta}^0)\mathbf{S}^-(\boldsymbol{\theta}^0)\bar{\mathbf{h}}_T(\boldsymbol{\theta}^0)$, where irrespective of our choice of generalised inverse, the second expression is equal to $\sqrt{T}\mathbf{D}'_{\oplus\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus)\boldsymbol{\Delta}^{-1}_\oplus(\boldsymbol{\theta}^0)\mathbf{P}'_\oplus(\boldsymbol{\theta}^0)\bar{\mathbf{h}}_T(\boldsymbol{\theta}^0)$ because $\mathbf{L}'_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus, \mathbf{0})\mathbf{D}'(\boldsymbol{\theta}^0)\mathbf{S}^-(\boldsymbol{\theta}^0)$ equals

$$\begin{aligned} & \left[\ \mathbf{L}'_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus, \mathbf{0})\mathbf{D}'(\boldsymbol{\theta}^0)\mathbf{P}_\oplus(\boldsymbol{\theta}^0) \quad \mathbf{L}'_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus, \mathbf{0})\mathbf{D}'(\boldsymbol{\theta}^0)\mathbf{P}_\ominus(\boldsymbol{\theta}^0)\ \right]\begin{bmatrix} \boldsymbol{\Delta}^{-1}_\oplus(\boldsymbol{\theta}^0) & \boldsymbol{\Delta}^{\oplus\ominus}(\boldsymbol{\theta}^0) \\ \boldsymbol{\Delta}^{\ominus\oplus}(\boldsymbol{\theta}^0) & \boldsymbol{\Delta}^{\ominus\ominus}(\boldsymbol{\theta}^0) \end{bmatrix} \\ \times & \begin{bmatrix} \mathbf{P}'_\oplus(\boldsymbol{\theta}^0) \\ \mathbf{P}'_\ominus(\boldsymbol{\theta}^0) \end{bmatrix} = \left[\ \mathbf{D}'_{\oplus\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus, \mathbf{0}) \quad \mathbf{0}\ \right]\begin{bmatrix} \boldsymbol{\Delta}^{-1}_\oplus(\boldsymbol{\theta}^0) & \boldsymbol{\Delta}^{\oplus\ominus}(\boldsymbol{\theta}^0) \\ \boldsymbol{\Delta}^{\ominus\oplus}(\boldsymbol{\theta}^0) & \boldsymbol{\Delta}^{\ominus\ominus}(\boldsymbol{\theta}^0) \end{bmatrix}\begin{bmatrix} \mathbf{P}'_\oplus(\boldsymbol{\theta}^0) \\ \mathbf{P}'_\ominus(\boldsymbol{\theta}^0) \end{bmatrix}, \end{aligned}$$

and $\mathbf{P}'_{\ominus}(\boldsymbol{\theta}^0)\mathbf{h}[\mathbf{x}_t;\mathbf{l}(\boldsymbol{\theta}^0_{\oplus},\mathbf{0})]=\mathbf{0}\ \forall t$ in view of Assumptions 1 and 2. Specifically,

$$\lim_{T\to\infty} V\begin{bmatrix} \sqrt{T}\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\bar{\mathbf{h}}_T(\boldsymbol{\theta}^0) \\ \sqrt{T}\mathbf{D}'_{\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus})\boldsymbol{\Delta}^{-1}_{\oplus}(\boldsymbol{\theta}^0)\mathbf{P}'_{\oplus}(\boldsymbol{\theta}^0)\bar{\mathbf{h}}_T(\boldsymbol{\theta}^0)\end{bmatrix}=\begin{bmatrix} \mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{S}(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0) & \mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{S}(\boldsymbol{\theta}^0)\mathbf{P}_{\oplus}(\boldsymbol{\theta}^0)\boldsymbol{\Delta}^{-1}_{\oplus}(\boldsymbol{\theta}^0)\mathbf{D}_{\oplus}(\boldsymbol{\theta}^0_{\oplus}) \\ \mathbf{D}'_{\oplus}(\boldsymbol{\theta}^0_{\oplus})\boldsymbol{\Delta}^{-1}_{\oplus}(\boldsymbol{\theta}^0)\mathbf{P}'_{\oplus}(\boldsymbol{\theta}^0)\mathbf{S}(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0) & \mathbf{D}'_{\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus})\boldsymbol{\Delta}^{-1}_{\oplus}(\boldsymbol{\theta}^0)\mathbf{P}'_{\oplus}(\boldsymbol{\theta}^0)\mathbf{S}(\boldsymbol{\theta}^0)\mathbf{P}_{\oplus}(\boldsymbol{\theta}^0)\boldsymbol{\Delta}^{-1}_{\oplus}(\boldsymbol{\theta}^0)\mathbf{D}_{\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus})\end{bmatrix}$$

$$=\begin{bmatrix} \mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{S}(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0) & \mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\oplus}(\boldsymbol{\theta}^0_{\oplus},\mathbf{0}) \\ \mathbf{L}'_{\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\mathbf{0})\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0) & \mathbf{D}'_{\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus})\boldsymbol{\Delta}^{-1}_{\oplus}(\boldsymbol{\theta}^0)\mathbf{D}_{\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus})\end{bmatrix},$$

because $\mathbf{S}(\boldsymbol{\theta}^0)=\mathbf{P}_{\oplus}(\boldsymbol{\theta}^0)\boldsymbol{\Delta}_{\oplus}(\boldsymbol{\theta}^0)\mathbf{P}'_{\oplus}(\boldsymbol{\theta}^0)$, and

$$\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{S}(\boldsymbol{\theta}^0)\mathbf{P}_{\oplus}(\boldsymbol{\theta}^0)\boldsymbol{\Delta}^{-1}_{\oplus}(\boldsymbol{\theta}^0)\mathbf{D}_{\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus})=\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{P}_{\oplus}(\boldsymbol{\theta}^0)\mathbf{P}'_{\oplus}(\boldsymbol{\theta}^0)\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\mathbf{0})$$
$$=\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}[\mathbf{I}_n-\mathbf{P}_{\ominus}(\boldsymbol{\theta}^0)\mathbf{P}'_{\ominus}(\boldsymbol{\theta}^0)]\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\mathbf{0})=\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\mathbf{0})$$

in view of (A2) and the orthogonality of $\mathbf{P}(\boldsymbol{\theta}^0)$, which simultaneously guarantees that

$$\mathbf{P}'(\boldsymbol{\theta}^0)\mathbf{P}(\boldsymbol{\theta}^0)=\mathbf{P}_{\oplus}(\boldsymbol{\theta}^0)\mathbf{P}'_{\oplus}(\boldsymbol{\theta}^0)+\mathbf{P}_{\ominus}(\boldsymbol{\theta}^0)\mathbf{P}'_{\ominus}(\boldsymbol{\theta}^0)=\mathbf{I}_n=\mathbf{P}(\boldsymbol{\theta}^0)\mathbf{P}'(\boldsymbol{\theta}^0).$$

The delta method then implies that

$$\mathbf{L}(\boldsymbol{\theta}^0_{\oplus},\mathbf{0})\left\{\begin{array}{cc} \mathbf{V}_{\oplus}\left[\boldsymbol{\theta}^0,\boldsymbol{\Delta}^{-1}_{\oplus}(\boldsymbol{\theta}^0)\right] & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{array}\right\}\mathbf{L}'(\boldsymbol{\theta}^0_{\oplus},\mathbf{0})$$

will be the asymptotic covariance matrix of $\mathbf{l}\left\{\hat{\boldsymbol{\theta}}_{\oplus T}\left[\mathbf{S}^{-}(\boldsymbol{\theta}^0)\right],\mathbf{0}\right\}$, where

$$\mathbf{V}_{\oplus}\left[\boldsymbol{\theta}^0,\boldsymbol{\Delta}^{-1}_{\oplus}(\boldsymbol{\theta}^0)\right]=\left[\mathbf{D}'_{\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus})\boldsymbol{\Delta}^{-1}_{\oplus}(\boldsymbol{\theta}^0)\mathbf{D}_{\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus})\right]^{-1}$$
$$=[\mathbf{L}'_{\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus})\mathbf{D}'(\boldsymbol{\theta}^0)\mathbf{S}^{-}(\boldsymbol{\theta}^0)\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus})]^{-1}.$$

Similarly, the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)$ is

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

Hence, the difference between those two matrices will be positive semidefinite if so is the matrix

$$\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{S}(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0) - \mathbf{H}(\boldsymbol{\theta}^0,\boldsymbol{\Upsilon})\mathbf{L}(\boldsymbol{\theta}^0_\oplus,\mathbf{0})\left\{\begin{array}{cc} \mathbf{V}_\oplus\left[\boldsymbol{\theta}^0,\boldsymbol{\Delta}^{-1}_\oplus(\boldsymbol{\theta}^0)\right] & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{array}\right\}$$

$$\times\mathbf{L}'(\boldsymbol{\theta}^0_\oplus,\mathbf{0})\mathbf{H}(\boldsymbol{\theta}^0,\boldsymbol{\Upsilon}) = \mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{S}(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0)$$

$$-\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\left[\begin{array}{cc}\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus,\mathbf{0}) & \mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\theta}_\ominus}(\boldsymbol{\theta}^0_\oplus,\mathbf{0})\end{array}\right]\left\{\begin{array}{cc} \mathbf{V}_\oplus\left[\boldsymbol{\theta}^0,\boldsymbol{\Delta}^{-1}_\oplus(\boldsymbol{\theta}^0)\right] & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{array}\right\}$$

$$\times\left[\begin{array}{c}\mathbf{L}'_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus,\mathbf{0})\mathbf{D}'(\boldsymbol{\theta}^0) \\ \mathbf{L}'_{\boldsymbol{\theta}_\ominus}(\boldsymbol{\theta}^0_\oplus,\mathbf{0})\mathbf{D}'(\boldsymbol{\theta}^0)\end{array}\right]\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0) = \mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{S}(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0)$$

$$-\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus,\mathbf{0})[\mathbf{D}'_{\oplus\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus)\boldsymbol{\Delta}^{-1}_\oplus(\boldsymbol{\theta}^0)\mathbf{D}_{\oplus\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus)]^{-1}\mathbf{L}'_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus,\mathbf{0})\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0).$$

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 $\sqrt{T}\mathbf{L}'_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus,\mathbf{0})\mathbf{D}'(\boldsymbol{\theta}^0)\mathbf{S}^-(\boldsymbol{\theta}^0)\bar{\mathbf{h}}_T(\boldsymbol{\theta}^0)$. □

Proposition 3:

Trivial because $\boldsymbol{\Pi}'_E(\boldsymbol{\varphi},\boldsymbol{\nu})\mathbf{h}_E(\mathbf{R}_t;\boldsymbol{\varphi},\boldsymbol{\nu}) = \boldsymbol{\varphi}^{*\prime}\boldsymbol{\nu} - \boldsymbol{\varphi}^{+\prime}\ell_N\ \forall t$ for any $\boldsymbol{\nu}$, $\boldsymbol{\varphi}^+$, and $\boldsymbol{\varphi}^*$. □

Proposition 4:

The matrices $\boldsymbol{\Phi}(c)$ are such that $(1,\mathbf{R}'_{1t}\text{-}c^{-1}\ell'_{N_1})\boldsymbol{\Phi}^{-1\prime}(c) = (1,\mathbf{R}'_{1t})$. As a result, $\boldsymbol{\Pi}'_L\left[\mathbf{a}(c_i),\mathbf{b}(c_i),\mathbf{a}(c_{ii}),\mathbf{b}(c_{ii})\right]\mathbf{h}_L\left[\mathbf{R}_t;\mathbf{a}(c_i),\mathbf{b}(c_i),\mathbf{a}(c_{ii}),\mathbf{b}(c_{ii})\right]$ will be

$$\left\{\left[\boldsymbol{\Phi}^{-1}(c_i)\otimes\mathbf{I}_{N_2}\right],\left[-\boldsymbol{\Phi}^{-1}(c_{ii})\otimes\mathbf{I}_{N_2}\right]\right\}\left\{\begin{array}{c}\left(\begin{array}{c}1\\ \mathbf{R}_{1t}\text{-}c_i^{-1}\ell_{N_1}\end{array}\right)\otimes\left[\begin{array}{c}(\mathbf{R}_{2t}\text{-}c_i^{-1}\ell_{N_2})-\mathbf{a}(c_i)\\ -\mathbf{B}(c_i)(\mathbf{R}_{1t}\text{-}c_i^{-1}\ell_{N_1})\end{array}\right]\\ \left(\begin{array}{c}1\\ \mathbf{R}_{1t}\text{-}c_{ii}^{-1}\ell_{N_1}\end{array}\right)\otimes\left[\begin{array}{c}(\mathbf{R}_{2t}\text{-}c_{ii}^{-1}\ell_{N_2})-\mathbf{a}(c_{ii})\\ -\mathbf{B}(c_{ii})(\mathbf{R}_{1t}\text{-}c_{ii}^{-1}\ell_{N_1})\end{array}\right]\end{array}\right\}$$

$$=\left\{\left(\begin{array}{c}1\\ \mathbf{R}_{1t}\end{array}\right)\otimes\left[\begin{array}{c}\mathbf{a}(c_{ii})-\mathbf{a}(c_i)+c_i^{-1}[\mathbf{B}(c_i)\ell_{N_1}-\ell_{N_2}]\\ -c_{ii}^{-1}[\mathbf{B}(c_{ii})\ell_{N_1}-\ell_{N_2}]+[\mathbf{B}(c_{ii})-\mathbf{B}(c_i)]\mathbf{R}_{1t}\end{array}\right]\right\}.$$

Finally, it is trivial to see that such an expression will be equal to 0 $\forall t$ regardless of c_i and c_{ii} if and only if the conditions in (10) simultaneously hold. □

Proposition 5:

If we use (10) to reparametrise $\mathbf{a}(c_i),\mathbf{b}(c_i),\mathbf{a}(c_{ii})$ and $\mathbf{b}(c_{ii})$ in terms of $\mathbf{a}$ and $\mathbf{b}$, we can write $\mathbf{h}_L[\mathbf{R}_t;\mathbf{a}(c_i),\mathbf{b}(c_i),\mathbf{a}(c_{ii}),\mathbf{b}(c_{ii})]$ as

$$\left[\begin{array}{c}\boldsymbol{\Phi}(c_i)\otimes\mathbf{I}_{N_2}\\ \boldsymbol{\Phi}(c_{ii})\otimes\mathbf{I}_{N_2}\end{array}\right]\mathbf{h}_H(\mathbf{R}_t;\mathbf{a},\mathbf{b}) = \left\{\left[\begin{array}{cc}\boldsymbol{\Phi}(c_i) & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Phi}(c_{ii})\end{array}\right]\otimes\mathbf{I}_{N_2}\right\}[\ell_2\otimes\mathbf{h}_H(\mathbf{R}_t;\mathbf{a},\mathbf{b})],$$

where $\mathbf{h}_H(\mathbf{R}_t;\mathbf{a},\mathbf{b})$ is defined in (11). It is then straightforward to show that the asymptotic covariance matrix of $\sqrt{T}\bar{\mathbf{h}}_{LT}[\mathbf{a}(c_i),\mathbf{b}(c_i),\mathbf{a}(c_{ii}),\mathbf{b}(c_{ii})]$ will be

$$\left\{\begin{bmatrix}\boldsymbol{\Phi}(c_i) & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Phi}(c_{ii})\end{bmatrix}\otimes\mathbf{I}_{N_2}\right\}[\ell_2\ell_2'\otimes\mathbf{S}_H(\mathbf{a}^0,\mathbf{b}^0)]\left\{\begin{bmatrix}\boldsymbol{\Phi}'(c_i) & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Phi}'(c_{ii})\end{bmatrix}\otimes\mathbf{I}_{N_2}\right\},$$

a generalised inverse of which is

$$\left\{\begin{bmatrix}\boldsymbol{\Phi}^{-1\prime}(c_i) & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Phi}^{-1\prime}(c_{ii})\end{bmatrix}\otimes\mathbf{I}_{N_2}\right\}\left[\frac{\ell_2\ell_2'}{4}\otimes\mathbf{S}_H^{-1}(\mathbf{a}^0,\mathbf{b}^0)\right]\left\{\begin{bmatrix}\boldsymbol{\Phi}^{-1}(c_i) & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Phi}^{-1}(c_{ii})\end{bmatrix}\otimes\mathbf{I}_{N_2}\right\}.$$

Hence, it is clear that applying the optimal singular GMM approach developed in section 3.2 to $E\{\mathbf{h}_L[\mathbf{R}_t;\mathbf{a}(c_i),\mathbf{b}(c_i),\mathbf{a}(c_{ii}),\mathbf{b}(c_{ii})]\}=\mathbf{0}$ is equivalent to applying the standard optimal GMM approach to $E[\mathbf{h}_H(\mathbf{R}_t;\mathbf{a},\mathbf{b})]=\mathbf{0}$. □

Proposition 6:

For simplicity of exposition, but without loss of generality, we shall focus on the Wald versions of the different spanning tests in view of asymptotic equivalence of the Lagrange multiplier and distance metric tests under the null and sequences of local alternatives in the framework presented in section 3.

Let $\mathbf{G}_U(\boldsymbol{\phi})=[\mathbf{I}_2\otimes(\mathbf{0},\mathbf{I}_{N_2})\mathbf{0})]\boldsymbol{\phi}=\mathbf{Q}_U\boldsymbol{\phi}$, $\mathbf{G}_C(\boldsymbol{\varphi})=[\mathbf{I}_2\otimes(\mathbf{0},\mathbf{I}_{N_2})]\boldsymbol{\varphi}=\mathbf{Q}_C\boldsymbol{\varphi}$, and

$$\mathbf{G}_H(\mathbf{a},\mathbf{b})=\left[\begin{pmatrix}1 & \mathbf{0}'\\ 0 & -\ell_{N_1}\end{pmatrix}\otimes\mathbf{I}_{N_2}\right]\begin{pmatrix}\mathbf{a}\\ \mathbf{b}\end{pmatrix}+\begin{pmatrix}\mathbf{0}\\ \ell_{N_2}\end{pmatrix}=\mathbf{Q}_H\begin{pmatrix}\mathbf{a}\\ \mathbf{b}\end{pmatrix}+\mathbf{q}_H,$$

so that the three null hypotheses can be written as $\mathbf{G}_U(\boldsymbol{\phi})=\mathbf{0}$, $\mathbf{G}_C(\boldsymbol{\varphi})=\mathbf{0}$ and $\mathbf{G}_H(\mathbf{a},\mathbf{b})=\mathbf{0}$, respectively.

In this notation, the Wald version of the uncentred RP spanning test will be based on

$$\mathbf{G}_U'(\hat{\boldsymbol{\phi}}_T)=(\hat{\boldsymbol{\phi}}_{2T}^{+\prime},\hat{\boldsymbol{\phi}}_{2T}^{*\prime})=[(\hat{\boldsymbol{\nu}}_{2T}-\hat{\mathbf{C}}_T\hat{\boldsymbol{\nu}}_{1T})',(\ell_{N_2}-\hat{\mathbf{C}}_T\ell_{N_1})'](\mathbf{I}_2\otimes\hat{\boldsymbol{\Lambda}}_T^{-1})$$

with $\hat{\mathbf{C}}_T=\hat{\boldsymbol{\Gamma}}_{21T}\hat{\boldsymbol{\Gamma}}_{11T}^{-1}$ and $\hat{\boldsymbol{\Lambda}}_T=\hat{\boldsymbol{\Gamma}}_{22T}-\hat{\boldsymbol{\Gamma}}_{21T}\hat{\boldsymbol{\Gamma}}_{11T}^{-1}\hat{\boldsymbol{\Gamma}}_{21T}'$, where the last expression has being obtained after applying the partitioned inverse formula to $(\hat{\boldsymbol{\phi}}_T^{+\prime},\hat{\boldsymbol{\phi}}_T^{*\prime})=(\hat{\boldsymbol{\nu}}_T',\ell_N')(\mathbf{I}_2\otimes\hat{\boldsymbol{\Gamma}}_T^{-1})$.

On the other hand, the Wald version of the centred RP spanning test is based on

$$\begin{aligned}\mathbf{G}_C(\hat{\boldsymbol{\varphi}}_T) &= \begin{pmatrix}\hat{\boldsymbol{\varphi}}_{2T}^{+}\\ \hat{\boldsymbol{\varphi}}_{2T}^{*}\end{pmatrix}=(\mathbf{I}_2\otimes\hat{\boldsymbol{\Omega}}_T^{-1})\begin{pmatrix}\hat{\mathbf{a}}_T\\ \hat{\mathbf{f}}_T\end{pmatrix}=\left[\begin{pmatrix}1+\hat{\text{B}}_{1T} & 0\\ \hat{\text{A}}_{1T} & 1\end{pmatrix}\otimes\hat{\boldsymbol{\Omega}}_T^{-1}\hat{\boldsymbol{\Lambda}}_T\right]\mathbf{G}_U(\hat{\boldsymbol{\phi}}_T)\\ &= \hat{\boldsymbol{\Psi}}_{CT}\mathbf{G}_U(\hat{\boldsymbol{\phi}}_T),\end{aligned}$$

where the first expression has been obtained after applying the partitioned inverse formula to $(\hat{\boldsymbol{\varphi}}_T^{+\prime}, \hat{\boldsymbol{\varphi}}_T^{*\prime}) = (\hat{\boldsymbol{\nu}}_T^{\prime}, \ell_N^{\prime})(\mathbf{I}_2 \otimes \hat{\boldsymbol{\Sigma}}_T^{-1})$, while the last expression follows from the numerical relationships $\hat{\mathbf{a}}_T = (1+\hat{\text{B}}_{1T})(\hat{\boldsymbol{\nu}}_{2T} - \hat{\mathbf{C}}_T\hat{\boldsymbol{\nu}}_{1T})$ and $\hat{\mathbf{f}}_T = (\ell_{N_2} - \mathbf{C}_T\ell_{N_1}) + \hat{\text{A}}_{1T}(\hat{\boldsymbol{\nu}}_{2T} - \hat{\mathbf{C}}_T\hat{\boldsymbol{\nu}}_{1T})$, with $\hat{\text{A}}_{1T} = \ell_{N_1}^{\prime}\hat{\boldsymbol{\Sigma}}_{11T}^{-1}\hat{\boldsymbol{\nu}}_{1T}$ and $\hat{\text{B}}_{1T} = \hat{\boldsymbol{\nu}}_{1T}^{\prime}\hat{\boldsymbol{\Sigma}}_{11T}^{-1}\hat{\boldsymbol{\nu}}_{1T}$.

Finally, the Wald version of the regression spanning test is based on

$$\begin{aligned}
\mathbf{G}_H(\hat{\mathbf{a}}_T, \hat{\mathbf{b}}_T) &= \begin{pmatrix} \hat{\mathbf{a}}_T \\ \hat{\mathbf{f}}_T \end{pmatrix} = (\mathbf{I}_2 \otimes \hat{\boldsymbol{\Omega}}_T)\mathbf{G}_C(\hat{\boldsymbol{\varphi}}_T) = \left[\begin{pmatrix} 1+\hat{\text{B}}_{1T} & 0 \\ \hat{\text{A}}_{1T} & 1 \end{pmatrix} \otimes \hat{\boldsymbol{\Lambda}}_T\right]\mathbf{G}_U(\hat{\boldsymbol{\phi}}_T) \\
&= \hat{\boldsymbol{\Psi}}_{HT}\mathbf{G}_U(\hat{\boldsymbol{\phi}}_T).
\end{aligned}$$

Hence, both $\mathbf{G}_C(\hat{\boldsymbol{\varphi}}_T)$ and $\mathbf{G}_H(\hat{\mathbf{a}}_T, \hat{\mathbf{b}}_T)$ can be written as full-rank linear combinations of $\mathbf{G}_U(\hat{\boldsymbol{\phi}}_T)$. However, since $\hat{\boldsymbol{\Psi}}_{CT}$ and $\hat{\boldsymbol{\Psi}}_{HT}$ depend on sample data, the three Wald tests will not be numerically equal in general (cf. Lemma 1).

In this context, we can define the different alternative hypothesis to be compatible if $\mathbf{G}_C(\boldsymbol{\varphi}^0) = \boldsymbol{\Psi}_C\mathbf{G}_U(\boldsymbol{\phi}^0)$ and $\mathbf{G}_H(\mathbf{a}^0, \mathbf{b}^0) = \boldsymbol{\Psi}_H\mathbf{G}_U(\boldsymbol{\phi}^0)$, where

$$\boldsymbol{\Psi}_C = \left[\begin{pmatrix} 1+\text{B}_1 & 0 \\ \text{A}_1 & 1 \end{pmatrix} \otimes \boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}\right] = p\lim_{T\to\infty} \hat{\boldsymbol{\Psi}}_{CT} \tag{A3}$$

and

$$\boldsymbol{\Psi}_H = \left[\begin{pmatrix} 1+\text{B}_1 & 0 \\ \text{A}_1 & 1 \end{pmatrix} \otimes \boldsymbol{\Lambda}\right] = p\lim_{T\to\infty} \hat{\boldsymbol{\Psi}}_{HT}. \tag{A4}$$

Then, given that we can always write

$$\begin{aligned}
\sqrt{T}[\mathbf{G}_C(\hat{\boldsymbol{\varphi}}_T) - \mathbf{G}_C(\boldsymbol{\varphi}^0)] &= \sqrt{T}[\hat{\boldsymbol{\Psi}}_{CT}\mathbf{G}_U(\hat{\boldsymbol{\phi}}_T) - \boldsymbol{\Psi}_C\mathbf{G}_U(\boldsymbol{\phi}^0)] \\
&= \hat{\boldsymbol{\Psi}}_{CT}\sqrt{T}[\mathbf{G}_U(\hat{\boldsymbol{\phi}}_T) - \mathbf{G}_U(\boldsymbol{\phi}^0)] + \sqrt{T}(\hat{\boldsymbol{\Psi}}_{CT} - \boldsymbol{\Psi}_C)\mathbf{G}_U(\boldsymbol{\phi}^0)
\end{aligned}$$

regardless of the true values of the parameters, a straightforward application of Cramér's theorem implies that $\sqrt{T}[\mathbf{G}_C(\hat{\boldsymbol{\varphi}}_T) - \boldsymbol{\Psi}_C\boldsymbol{\delta}/\sqrt{T}] = \boldsymbol{\Psi}_C\sqrt{T}[\mathbf{G}_U(\hat{\boldsymbol{\phi}}_T) - \boldsymbol{\delta}/\sqrt{T}] + o_p(1)$ for the case of compatible local alternatives of the form $\mathbf{G}_U(\boldsymbol{\phi}) = \boldsymbol{\delta}/\sqrt{T}$ and $\mathbf{G}_C(\boldsymbol{\varphi}^0) = \boldsymbol{\Psi}_C\boldsymbol{\delta}/\sqrt{T}$.

Finally, we can use an analogous argument to show that $\sqrt{T}[\mathbf{G}_H(\hat{\mathbf{a}}_T, \hat{\mathbf{b}}_T) - \boldsymbol{\Psi}_H\boldsymbol{\delta}/\sqrt{T}] = \boldsymbol{\Psi}_H\sqrt{T}[\mathbf{G}_U(\hat{\boldsymbol{\phi}}_T) - \boldsymbol{\delta}/\sqrt{T}] + o_p(1)$ for the case of compatible local alternatives of the form $\mathbf{G}_U(\boldsymbol{\phi}) = \boldsymbol{\delta}/\sqrt{T}$ and $\mathbf{G}_H(\boldsymbol{\varphi}^0) = \boldsymbol{\Psi}_H\boldsymbol{\delta}/\sqrt{T}$. $\square$

B Proofs of Lemmata

Lemma 1:

Let $\hat{\boldsymbol{\theta}}_{AT}(\boldsymbol{\Upsilon}_{AT}) = \arg\min_{\boldsymbol{\theta}} \bar{\mathbf{h}}'_{AT}(\boldsymbol{\theta})\boldsymbol{\Upsilon}_{NT}\bar{\mathbf{h}}_{AT}(\boldsymbol{\theta})$ denote another GMM estimator of $\boldsymbol{\theta}$ based on the alternative set of moment conditions $E[\mathbf{h}_A(\mathbf{x}_t;\boldsymbol{\theta})] = \mathbf{0}$, where $\mathbf{h}_A(\mathbf{x}_t;\boldsymbol{\theta}) = \mathbf{A}(\boldsymbol{\theta})\mathbf{h}(\mathbf{x}_t;\boldsymbol{\theta})$. Our assumptions on $\mathbf{A}(\boldsymbol{\theta})$ guarantee that all the required regularity conditions for the asymptotic normality of $\hat{\boldsymbol{\theta}}_{AT}(\boldsymbol{\Upsilon}_{NT})$ will hold if they hold for $\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)$. Therefore, we will have that

$$\sqrt{T}[\hat{\boldsymbol{\theta}}_{AT}(\boldsymbol{\Upsilon}_{NT}) - \boldsymbol{\theta}^0] - [\mathbf{D}'_A(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}_N\mathbf{D}_A(\boldsymbol{\theta}^0)]^{-1}\mathbf{D}'_A(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}_N\sqrt{T}\bar{\mathbf{h}}_{AT}(\boldsymbol{\theta}^0) = o_p(1)$$

and

$$\sqrt{T}[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T) - \boldsymbol{\theta}^0] - [\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\mathbf{D}(\boldsymbol{\theta}^0)]^{-1}\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\sqrt{T}\bar{\mathbf{h}}_T(\boldsymbol{\theta}^0) = o_p(1),$$

where $\mathbf{D}(\boldsymbol{\theta}) = E[\partial\mathbf{h}(\mathbf{x}_t;\boldsymbol{\theta})/\partial\boldsymbol{\theta}']$ and $\mathbf{D}_A(\boldsymbol{\theta}) = E[\partial\mathbf{h}_A(\mathbf{x}_t;\boldsymbol{\theta})/\partial\boldsymbol{\theta}']$.

Importantly, these results continue to be valid even if there is a linear combination of $\mathbf{h}(\mathbf{x}_t;\boldsymbol{\theta}^0)$ that is 0 for all t.

But since $\sqrt{T}\bar{\mathbf{h}}_{AT}(\boldsymbol{\theta}^0) - \mathbf{A}(\boldsymbol{\theta}^0)\sqrt{T}\bar{\mathbf{h}}_T(\boldsymbol{\theta}^0) = o_p(1)$, and

$$\mathbf{D}_A(\boldsymbol{\theta}) = E[\mathbf{h}(\mathbf{x}_t;\boldsymbol{\theta}) \otimes \mathbf{I}]\partial vec[\mathbf{A}(\boldsymbol{\theta})]\partial\boldsymbol{\theta}' + \mathbf{A}(\boldsymbol{\theta})\mathbf{D}(\boldsymbol{\theta}),$$

so that $\mathbf{D}_A(\boldsymbol{\theta}^0) = \mathbf{A}(\boldsymbol{\theta}^0)\mathbf{D}_O(\boldsymbol{\theta}^0)$ because $E[\mathbf{h}_O(\mathbf{x}_t;\boldsymbol{\theta}^0)] = \mathbf{0}$ by definition of $\boldsymbol{\theta}^0$, then we will have that

$$\sqrt{T}[\hat{\boldsymbol{\theta}}_{AT}(\boldsymbol{\Upsilon}_{NT}) - \hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)] = o_p(1).$$

Next, note that our assumptions on $\mathbf{P}(\boldsymbol{\theta})$ ensure that the usual regularity conditions for asymptotic normality of $\hat{\boldsymbol{\rho}}_{NT}(\boldsymbol{\Upsilon}_{NT})$ will hold if they hold for $\hat{\boldsymbol{\theta}}_{AT}(\boldsymbol{\Upsilon}_{NT})$. Therefore, we can easily show that

$$\sqrt{T}[\hat{\boldsymbol{\rho}}_{NT}(\boldsymbol{\Upsilon}_{NT}) - \boldsymbol{\rho}^0] - \{\mathbf{D}'_N(\boldsymbol{\rho}^0)\boldsymbol{\Upsilon}_N\mathbf{D}_N(\boldsymbol{\rho}^0)]^{-1}\mathbf{D}'_N(\boldsymbol{\rho}^0)\boldsymbol{\Upsilon}_N\sqrt{T}\bar{\mathbf{h}}_{NT}(\boldsymbol{\rho}^0) = o_p(1),$$

regardless of whether or not there is a linear combination of $\mathbf{h}(\mathbf{x}_t;\boldsymbol{\theta}^0)$ that is 0 for all t.

If we then apply the standard delta method, we can show that both

$$\sqrt{T}\{\mathbf{G}_N[\hat{\boldsymbol{\rho}}_{NT}(\boldsymbol{\Upsilon}_{NT})] - \mathbf{G}_N(\boldsymbol{\rho}^0)\} - \mathbf{Q}_N(\boldsymbol{\rho}^0)[\mathbf{D}'_N(\boldsymbol{\rho}^0)\boldsymbol{\Upsilon}_N\mathbf{D}_N(\boldsymbol{\rho}^0)]^{-1}\mathbf{D}'_N(\boldsymbol{\rho}^0)\boldsymbol{\Upsilon}_N\sqrt{T}\bar{\mathbf{h}}_{NT}(\boldsymbol{\rho}^0)$$

and

$$\sqrt{T}\{\mathbf{G}[\hat{\boldsymbol{\theta}}_{AT}(\boldsymbol{\Upsilon}_{NT})] - \mathbf{G}(\boldsymbol{\theta}^0)\} - \mathbf{Q}(\boldsymbol{\theta}^0)[\mathbf{D}'_A(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}_N\mathbf{D}_A(\boldsymbol{\theta}^0)]^{-1}\mathbf{D}'_A(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}_N\sqrt{T}\bar{\mathbf{h}}_{AT}(\boldsymbol{\theta}^0)$$

will be $o_p(1)$. But since $\sqrt{T}\bar{\mathbf{h}}_{NT}(\boldsymbol{\rho}^0) - \sqrt{T}\bar{\mathbf{h}}_{AT}[\mathbf{P}^{-1}(\boldsymbol{\rho}^0)]$ is also $o_p(1)$, and

$$\frac{\partial \mathbf{h}_N(\mathbf{x}_t;\boldsymbol{\rho}^0)}{\partial \boldsymbol{\rho}'} = \frac{\partial \mathbf{h}_A[\mathbf{x}_t;\mathbf{P}^{-1}(\boldsymbol{\rho}^0)]}{\partial \boldsymbol{\theta}'}\frac{\partial \mathbf{P}^{-1}(\boldsymbol{\rho}^0)}{\partial \boldsymbol{\rho}'},$$

so that

$$\mathbf{D}_N(\boldsymbol{\rho}^0) = \mathbf{D}_A[\mathbf{P}^{-1}\left(\boldsymbol{\rho}^0\right)]\frac{\partial \mathbf{P}^{-1}(\boldsymbol{\rho}^0)}{\partial \boldsymbol{\rho}'},$$

and

$$\mathbf{Q}_N(\boldsymbol{\rho}^0) = \frac{\partial \mathbf{G}_N(\boldsymbol{\rho}^0)}{\partial \boldsymbol{\rho}'} = \frac{\partial \mathbf{G}[\mathbf{P}^{-1}(\boldsymbol{\rho}^0)]}{\partial \boldsymbol{\theta}'}\frac{\partial \mathbf{P}^{-1}(\boldsymbol{\rho}^0)}{\partial \boldsymbol{\rho}'},$$

the first part follows because $\boldsymbol{\rho}^0 = \mathbf{P}\left(\boldsymbol{\theta}^0\right)$ and $\partial\mathbf{P}^{-1}(\boldsymbol{\rho}^0)/\partial\boldsymbol{\rho}' = [\partial\mathbf{P}\left(\boldsymbol{\theta}^0\right)/\partial\boldsymbol{\theta}']^{-1}$ has full rank.

To prove the second part, note that the sample first-order conditions that define $\hat{\boldsymbol{\theta}}_{AT}(\boldsymbol{\Upsilon}_{NT})$ and $\hat{\boldsymbol{\theta}}(\boldsymbol{\Upsilon}_T)$ are given by

$$\bar{\mathbf{D}}'_{AT}[\hat{\boldsymbol{\theta}}_{AT}(\boldsymbol{\Upsilon}_{NT})] \cdot \boldsymbol{\Upsilon}_{NT} \cdot \bar{\mathbf{h}}_{AT}[\hat{\boldsymbol{\theta}}_{AT}(\boldsymbol{\Upsilon}_{NT})] = \mathbf{0}$$

and

$$\bar{\mathbf{D}}'_{T}[\hat{\boldsymbol{\theta}}_{T}(\boldsymbol{\Upsilon}_{T})] \cdot \boldsymbol{\Upsilon}_{T} \cdot \bar{\mathbf{h}}_{T}[\hat{\boldsymbol{\theta}}_{T}(\boldsymbol{\Upsilon}_{T})] = \mathbf{0},$$

respectively. But since

$$\bar{\mathbf{D}}'_{AT}[\hat{\boldsymbol{\theta}}_{AT}(\boldsymbol{\Upsilon}_{NT})] \cdot \boldsymbol{\Upsilon}_{NT} \cdot \bar{\mathbf{h}}_{AT}[\hat{\boldsymbol{\theta}}_{AT}(\boldsymbol{\Upsilon}_{NT})] = \bar{\mathbf{D}}'_{T}[\hat{\boldsymbol{\theta}}_{AT}(\boldsymbol{\Upsilon}_{NT})]\mathbf{A}' \cdot \boldsymbol{\Upsilon}_{NT} \cdot \mathbf{A}\bar{\mathbf{h}}_{T}[\hat{\boldsymbol{\theta}}_{AT}(\boldsymbol{\Upsilon}_{NT})],$$

because of the chain rule for Jacobian matrices and the definition of $\mathbf{h}_A(\mathbf{x}_t;\boldsymbol{\theta})$, then the condition $\boldsymbol{\Upsilon}_{NT} = \mathbf{A}'^{-1}\boldsymbol{\Upsilon}_T\mathbf{A}^{-1}$ guarantees that $\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)$ will also satisfy the sample first-order conditions that define $\hat{\boldsymbol{\theta}}_{AT}(\boldsymbol{\Upsilon}_{NT})$ for large enough T.

Finally, note that the sample first-order conditions that define $\hat{\boldsymbol{\rho}}_{NT}(\boldsymbol{\Upsilon}_{NT})$ are

$$\bar{\mathbf{D}}'_{NT}[\hat{\boldsymbol{\rho}}_{NT}(\boldsymbol{\Upsilon}_{NT})] \cdot \boldsymbol{\Upsilon}_{NT} \cdot \bar{\mathbf{h}}_{NT}[\hat{\boldsymbol{\rho}}_{NT}(\boldsymbol{\Upsilon}_{NT})] = \mathbf{0}.$$

But since

$$\begin{aligned} &\bar{\mathbf{D}}'_{NT}[\hat{\boldsymbol{\rho}}_{NT}(\boldsymbol{\Upsilon}_{NT})] \cdot \boldsymbol{\Upsilon}_{NT} \cdot \bar{\mathbf{h}}_{NT}[\hat{\boldsymbol{\rho}}_{NT}(\boldsymbol{\Upsilon}_{NT})] \\ = &\frac{\partial \mathbf{P}^{-1\prime}[\hat{\boldsymbol{\rho}}_{NT}(\boldsymbol{\Upsilon}_{NT})]}{\partial \boldsymbol{\rho}'}\bar{\mathbf{D}}'_{AT}\{\mathbf{P}^{-1}[\hat{\boldsymbol{\rho}}_{NT}(\boldsymbol{\Upsilon}_{NT})]\} \cdot \boldsymbol{\Upsilon}_{NT} \cdot \bar{\mathbf{h}}_{AT}[\mathbf{P}^{-1}[\hat{\boldsymbol{\rho}}_{NT}(\boldsymbol{\Upsilon}_{NT})]\}, \end{aligned}$$

because of the chain rule for Jacobian matrices and the definition of $\mathbf{h}_N(\mathbf{x}_t;\boldsymbol{\rho})$, then $\mathbf{P}[\hat{\boldsymbol{\theta}}_{AT}(\boldsymbol{\Upsilon}_{NT})] = \mathbf{P}[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)]$ will satisfy the sample first-order conditions that define $\hat{\boldsymbol{\rho}}_{NT}(\boldsymbol{\Upsilon}_{NT})$. Finally, the result follows because $\mathbf{G}_N\{\mathbf{P}[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)]\} = \mathbf{G}[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)]$ for large enough T. $\square$

Lemma 2:

The first part of this lemma follows immediately from the proof of Proposition 2 if we note that in the exactly identified case both $\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}$ and $\mathbf{D}'_{\oplus\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus,\mathbf{0})\boldsymbol{\Delta}^{-1}_\oplus(\boldsymbol{\theta}^0)$ are full-rank square matrices of orders k and $k_\oplus$ respectively, and also that

$$\mathbf{h}(\mathbf{x}_t;\boldsymbol{\theta}^0)=\mathbf{P}(\boldsymbol{\theta}^0)\mathbf{P}'(\boldsymbol{\theta}^0)\mathbf{h}(\mathbf{x}_t;\boldsymbol{\theta}^0)=\mathbf{P}_\oplus(\boldsymbol{\theta}^0)\mathbf{h}_{\oplus T}(\mathbf{x}_t;\boldsymbol{\theta}^0_\oplus,\mathbf{0})$$

because the singularity described in Assumption 1 is such that $\mathbf{P}'_\ominus(\boldsymbol{\theta}^0)\mathbf{h}_t(\mathbf{x}_t;\boldsymbol{\theta}^0)=\mathbf{0}\ \forall t$ by virtue of Assumption 2. Hence, the asymptotic residual variance in the limiting projection of $\sqrt{T}\mathbf{D}'(\boldsymbol{\theta}^0)\boldsymbol{\Upsilon}\bar{\mathbf{h}}_T(\boldsymbol{\theta}^0)$ on $\sqrt{T}\mathbf{L}'_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus,\mathbf{0})\mathbf{D}'(\boldsymbol{\theta}^0)\mathbf{S}^-(\boldsymbol{\theta}^0)\bar{\mathbf{h}}_T(\boldsymbol{\theta}^0)$ will be identically $\mathbf{0}$.

Now, given that the original functions $\mathbf{h}(\mathbf{x}_t;\boldsymbol{\theta})$ exactly identify $\boldsymbol{\theta}$, which in this case requires $\boldsymbol{\Upsilon}$ to have full rank, the unrestricted GMM estimator $\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)$ can be obtained for large enough T by solving the non-linear equation system $\bar{\mathbf{h}}_T[\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)]=\mathbf{0}$ regardless of $\boldsymbol{\Upsilon}_T$. For that reason, we shall refer to $\hat{\boldsymbol{\theta}}_T(\boldsymbol{\Upsilon}_T)$ as $\hat{\boldsymbol{\theta}}_T$ in what follows. If we make the additional assumption that $\mathbf{m}_\oplus(\hat{\boldsymbol{\theta}}_T)=\mathbf{0}$, then the regularity of $\mathbf{m}(.)$ on an open neighbourhood of $\boldsymbol{\theta}^0$ together with the consistency of $\hat{\boldsymbol{\theta}}_T$ imply that in sufficiently long samples there will be a unique value of $\boldsymbol{\theta}_\oplus$, $\ddot{\boldsymbol{\theta}}_{\oplus T}$ say, such that $\hat{\boldsymbol{\theta}}_T=\mathbf{l}(\ddot{\boldsymbol{\theta}}_{\oplus T},\mathbf{0})$, which in turn implies that $\bar{\mathbf{h}}_T[\mathbf{l}(\ddot{\boldsymbol{\theta}}_{\oplus T},\mathbf{0})]=\mathbf{0}$ for large enough T.

The condition $\mathbf{m}_\ominus(\hat{\boldsymbol{\theta}}_T)=\mathbf{0}$ also implies that $\boldsymbol{\Pi}'(\hat{\boldsymbol{\theta}}_T)\mathbf{h}(\mathbf{x}_t;\hat{\boldsymbol{\theta}}_T)=\mathbf{0}\ \forall t$ in view of Assumption 1, and consequently, that $\bar{\mathbf{S}}_T(\hat{\boldsymbol{\theta}}_T)$ has necessarily rank $k_\oplus$. Let

$$\begin{bmatrix}\bar{\mathbf{P}}_{\oplus T}(\hat{\boldsymbol{\theta}}_T) & \bar{\mathbf{P}}_{\ominus T}(\hat{\boldsymbol{\theta}}_T)\end{bmatrix}\begin{bmatrix}\bar{\boldsymbol{\Delta}}_{\oplus T}(\hat{\boldsymbol{\theta}}_T) & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\bar{\mathbf{P}}'_{\oplus T}(\hat{\boldsymbol{\theta}}_T)\\ \bar{\mathbf{P}}'_{\ominus T}(\hat{\boldsymbol{\theta}}_T)\end{bmatrix}=\bar{\mathbf{P}}_{\oplus T}(\hat{\boldsymbol{\theta}}_T)\bar{\boldsymbol{\Delta}}_{\oplus T}(\hat{\boldsymbol{\theta}}_T)\bar{\mathbf{P}}'_{\oplus T}(\hat{\boldsymbol{\theta}}_T)$$

denote the spectral decomposition of this matrix. Since $\boldsymbol{\theta}_\oplus$ will also be exactly identified from the transformed functions $\mathbf{h}_\oplus(\mathbf{x}_t;\boldsymbol{\theta}_\oplus,\mathbf{0}|\hat{\boldsymbol{\theta}}_T)=\bar{\mathbf{P}}'_\oplus(\hat{\boldsymbol{\theta}}_T)\mathbf{h}[\mathbf{x}_t;\mathbf{l}(\boldsymbol{\theta}_\oplus,\mathbf{0})]$ under our assumptions, the feasible unrestricted GMMS estimator, $\hat{\boldsymbol{\theta}}_{\oplus T}[\bar{\mathbf{S}}^+_T(\hat{\boldsymbol{\theta}}_T)]$, can be obtained for large enough T by solving the non-linear equation system $\bar{\mathbf{h}}_\oplus(\hat{\boldsymbol{\theta}}_{\oplus T},\mathbf{0}|\hat{\boldsymbol{\theta}}_T)=\bar{\mathbf{P}}'_\oplus(\hat{\boldsymbol{\theta}}_T)\bar{\mathbf{h}}_T[\mathbf{l}\{\hat{\boldsymbol{\theta}}_{\oplus T}[\bar{\mathbf{S}}^+_T(\hat{\boldsymbol{\theta}}_T)],\mathbf{0})\}]=\mathbf{0}$. But since we saw before that $\bar{\mathbf{h}}_T[\mathbf{l}(\ddot{\boldsymbol{\theta}}_{\oplus T},\mathbf{0})]=\mathbf{0}$ in sufficiently long samples, $\hat{\boldsymbol{\theta}}_{\oplus T}[\bar{\mathbf{S}}^+_T(\hat{\boldsymbol{\theta}}_T)]$ must be numerically equal to $\ddot{\boldsymbol{\theta}}_{\oplus T}$ for large enough T. In fact, the numerical equality between $\ddot{\boldsymbol{\theta}}_T$ and $\mathbf{l}\{\hat{\boldsymbol{\theta}}_{\oplus T}[\bar{\mathbf{S}}^+_T(\hat{\boldsymbol{\theta}}_T),\mathbf{0}\}$ will continue to hold in this exactly identified context if in the definition of $\hat{\boldsymbol{\theta}}_{\oplus T}[\bar{\mathbf{S}}^+_T(\hat{\boldsymbol{\theta}}_T)]$ we replace $\bar{\mathbf{P}}_\oplus(\hat{\boldsymbol{\theta}}_T)$ by any matrix whose probability limit is such that none of its columns belong to the column span of $\boldsymbol{\Pi}(\boldsymbol{\theta}^0)$. For ease of notation, we refer to $\hat{\boldsymbol{\theta}}_{\oplus T}[\bar{\mathbf{S}}^+_T(\hat{\boldsymbol{\theta}}_T)]$ as $\hat{\boldsymbol{\theta}}_{\oplus T}$ in what follows.

To prove the third part, we can apply the delta method to show that

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

and

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

Therefore, it is clear that

$$\begin{aligned}\sqrt{T}[\mathbf{G}(\hat{\boldsymbol{\theta}}_T)-\mathbf{G}_\oplus(\hat{\boldsymbol{\theta}}_{\oplus T})]=o_p(1)\\ \Leftrightarrow \mathbf{Q}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus,\mathbf{0})\left[\mathbf{P}'_\oplus(\boldsymbol{\theta}^0)\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus,\mathbf{0})\right]^{-1}-\mathbf{Q}(\boldsymbol{\theta}^0)\mathbf{D}^{-1}(\boldsymbol{\theta}^0)\mathbf{P}_\oplus(\boldsymbol{\theta}^0)=\mathbf{0}\\ \Leftrightarrow \mathbf{Q}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus,\mathbf{0})=\mathbf{Q}(\boldsymbol{\theta}^0)\mathbf{D}^{-1}(\boldsymbol{\theta}^0)\mathbf{P}_\oplus(\boldsymbol{\theta}^0)\mathbf{P}'_\oplus(\boldsymbol{\theta}^0)\mathbf{D}(\boldsymbol{\theta}^0)\mathbf{L}_{\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus,\mathbf{0}).\end{aligned} \tag{B1}$$

But the last condition is indeed satisfied in view of (A2) and the orthogonality of $\mathbf{P}(\boldsymbol{\theta}^0)$.

Finally, we can define an asymptotically valid Wald test of $H_0:\mathbf{G}_\oplus(\boldsymbol{\theta}_\oplus)=\mathbf{0}$ based on $\hat{\boldsymbol{\theta}}_{\oplus T}$ as

$$W_{\oplus T}=T\mathbf{G}'_\oplus(\hat{\boldsymbol{\theta}}_{\oplus T})[\mathbf{Q}_\oplus(\hat{\boldsymbol{\theta}}_{\oplus T})\bar{\mathbf{D}}^{-1}_{\oplus\boldsymbol{\theta}_\oplus T}(\hat{\boldsymbol{\theta}}_{\oplus T}|\hat{\boldsymbol{\theta}}_T)\bar{\boldsymbol{\Delta}}_{\oplus T}(\hat{\boldsymbol{\theta}}_T)\bar{\mathbf{D}}^{-1\prime}_{\oplus\boldsymbol{\theta}_\oplus}(\hat{\boldsymbol{\theta}}_{\oplus T}|\hat{\boldsymbol{\theta}}_T)\mathbf{Q}'_\oplus(\hat{\boldsymbol{\theta}}_{\oplus T})]^{-1}\mathbf{G}_\oplus(\hat{\boldsymbol{\theta}}_{\oplus T}).$$

We can also define an asymptotically valid Wald test of $H_0:\mathbf{G}(\boldsymbol{\theta})=\mathbf{0}$ based on $\hat{\boldsymbol{\theta}}_T$ as

$$W_T=T\mathbf{G}'(\hat{\boldsymbol{\theta}}_T)[\mathbf{Q}(\hat{\boldsymbol{\theta}}_T)\bar{\mathbf{D}}^{-1}_T(\hat{\boldsymbol{\theta}}_T)\bar{\mathbf{S}}_T(\hat{\boldsymbol{\theta}}_T)\bar{\mathbf{D}}^{-1\prime}_T(\hat{\boldsymbol{\theta}}_T)\mathbf{Q}'(\hat{\boldsymbol{\theta}}_T)]^{-1}\mathbf{G}(\hat{\boldsymbol{\theta}}_T).$$

Given that we have defined $\mathbf{G}_\oplus(\boldsymbol{\theta}_\oplus)$ as $\mathbf{G}[\mathbf{l}(\boldsymbol{\theta}_\oplus,\mathbf{0})]$, it follows from the second part of the lemma that $\mathbf{G}(\hat{\boldsymbol{\theta}}_T)=\mathbf{G}_\oplus(\hat{\boldsymbol{\theta}}_{\oplus T})$ for large enough T. As a result, W_T and $W_{\oplus T}$ will be numerically identical if the matrices defining the two quadratic forms are also numerically identical. We can express those matrices as

$$\begin{aligned}&\mathbf{Q}(\hat{\boldsymbol{\theta}}_T)\bar{\mathbf{D}}_T(\hat{\boldsymbol{\theta}}_T)^{-1}\bar{\mathbf{S}}_T(\hat{\boldsymbol{\theta}}_T)\bar{\mathbf{D}}^{-1\prime}_T(\hat{\boldsymbol{\theta}}_T)\mathbf{Q}'(\hat{\boldsymbol{\theta}}_T)\\ =\;&\mathbf{Q}(\hat{\boldsymbol{\theta}}_T)\bar{\mathbf{D}}^{-1}_T(\hat{\boldsymbol{\theta}}_T)\bar{\mathbf{P}}_{\oplus T}(\hat{\boldsymbol{\theta}}_T)\bar{\boldsymbol{\Delta}}_{\oplus T}(\hat{\boldsymbol{\theta}}_T)\bar{\mathbf{P}}'_{\oplus T}(\hat{\boldsymbol{\theta}}_T)\bar{\mathbf{D}}^{-1\prime}_T(\hat{\boldsymbol{\theta}}_T)\mathbf{Q}'(\hat{\boldsymbol{\theta}}_T)\end{aligned}$$

and

$$\begin{aligned}\mathbf{Q}_\oplus(\hat{\boldsymbol{\theta}}_{\oplus T})\bar{\mathbf{D}}^{-1}_{\oplus\boldsymbol{\theta}_\oplus T}(\hat{\boldsymbol{\theta}}_{\oplus T},\mathbf{0}|\hat{\boldsymbol{\theta}}_T)\bar{\boldsymbol{\Delta}}_{\oplus T}(\hat{\boldsymbol{\theta}}_T)\bar{\mathbf{D}}'^{-1}_{\oplus\boldsymbol{\theta}_\oplus}(\hat{\boldsymbol{\theta}}_{\oplus T},\mathbf{0})\mathbf{Q}'_\oplus(\hat{\boldsymbol{\theta}}_{\oplus T})=\\ \mathbf{Q}(\hat{\boldsymbol{\theta}}_T)\mathbf{L}_{\boldsymbol{\theta}_\oplus}(\hat{\boldsymbol{\theta}}_{\oplus T},\mathbf{0})[\bar{\mathbf{P}}'_{\oplus T}(\hat{\boldsymbol{\theta}}_T)\bar{\mathbf{D}}_T(\hat{\boldsymbol{\theta}}_T)\mathbf{L}(\hat{\boldsymbol{\theta}}_{\oplus T})]^{-1}\bar{\boldsymbol{\Delta}}_{\oplus T}(\hat{\boldsymbol{\theta}}_T)\\ \times[\mathbf{L}'_{\boldsymbol{\theta}_\oplus}(\hat{\boldsymbol{\theta}}_{\oplus T},\mathbf{0})\bar{\mathbf{D}}_T(\hat{\boldsymbol{\theta}}_T)'\bar{\mathbf{P}}'_{\oplus T}(\hat{\boldsymbol{\theta}}_T)]^{-1}\mathbf{L}'_\oplus(\hat{\boldsymbol{\theta}}_{\oplus T},\mathbf{0})\mathbf{Q}'(\hat{\boldsymbol{\theta}}_T),\end{aligned}$$

respectively, where we have repeatedly used again the chain rule for Jacobian matrices. Hence, both quadratic forms will be numerically equal for large enough T irrespective of the value of $\bar{\mathbf{\Delta}}_{\oplus T}(\hat{\boldsymbol{\theta}}_T)$ if and only if

$$\mathbf{Q}(\hat{\boldsymbol{\theta}}_T)\bar{\mathbf{D}}^{-1}(\hat{\boldsymbol{\theta}}_T)\bar{\mathbf{P}}_{\oplus T}(\hat{\boldsymbol{\theta}}_T) = \mathbf{Q}(\hat{\boldsymbol{\theta}}_T)\mathbf{L}_{\boldsymbol{\theta}_\oplus}(\hat{\boldsymbol{\theta}}_{\oplus T}, \mathbf{0})\left[\bar{\mathbf{P}}'_{\oplus T}(\hat{\boldsymbol{\theta}}_T)\bar{\mathbf{D}}(\hat{\boldsymbol{\theta}}_T)\mathbf{L}_{\boldsymbol{\theta}_\oplus}(\hat{\boldsymbol{\theta}}_{\oplus T}, \mathbf{0})\right]^{-1}.$$

Given that $\hat{\boldsymbol{\theta}}_T = \mathbf{l}(\hat{\boldsymbol{\theta}}_{\oplus T}, \mathbf{0})$, an argument analogous to the one we used to prove (B1) allows us to show that the above condition will indeed be satisfied because (i) $\bar{\mathbf{h}}_T(\hat{\boldsymbol{\theta}}_T) = \mathbf{0}$, so that $\mathbf{\Pi}'(\hat{\boldsymbol{\theta}}_T)\bar{\mathbf{D}}_T(\hat{\boldsymbol{\theta}}_T)\mathbf{L}_{\boldsymbol{\theta}_\oplus}(\hat{\boldsymbol{\theta}}_{\oplus T}, \mathbf{0}) = \mathbf{0}$, and (ii) our choice of $\bar{\mathbf{S}}_T(\hat{\boldsymbol{\theta}}_T)$ as an estimator of $\mathbf{S}(\boldsymbol{\theta}^0)$ guarantees that $\bar{\mathbf{P}}_{\ominus T}(\hat{\boldsymbol{\theta}}_T)$ is a full column rank linear transformation of $\mathbf{\Pi}(\hat{\boldsymbol{\theta}}_T)$. □

Lemma 3:

Let us re-order the estimating functions in (6) as

$$\begin{pmatrix} \mathbf{R}_{1t}\mathbf{R}'_t\boldsymbol{\phi}^+ - \mathbf{R}_{1t} \\ \mathbf{R}_{1t}\mathbf{R}'_t\boldsymbol{\phi}^* - \ell_{N_1} \\ \mathbf{R}_{2t}\mathbf{R}'_t\boldsymbol{\phi}^+ - \mathbf{R}_{2t} \\ \mathbf{R}_{2t}\mathbf{R}'_t\boldsymbol{\phi}^* - \ell_{N_2} \end{pmatrix} = \begin{bmatrix} \mathbf{h}_{U1}(\mathbf{R}_t; \boldsymbol{\phi}) \\ \mathbf{h}_{U2}(\mathbf{R}_t; \boldsymbol{\phi}) \end{bmatrix}.$$

We just need to check that the condition (C1) in Lemma C1 (see Appendix C below) is satisfied for $k_{1\ominus} = k_{2\ominus} = 0$, 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 $\mathbf{D}_{U2}\mathbf{D}_{U1}^{-1}\mathbf{S}_{U11}(\boldsymbol{\phi})$ will be equal to

$$\left\{ \begin{array}{cc} (1+\textsc{b})^{-1}\left[1+\kappa\textsc{b}\,(1+\textsc{b})^{-1}\right] & \textsc{a}\,(1+\textsc{b})^{-2}\,\kappa \\ \textsc{a}\,(1+\textsc{b})^{-1}\,\kappa & \left[\textsc{c}-\textsc{a}^2\,(1+\textsc{b})^{-1}\right]+\kappa\left(\left[\textsc{c}-\textsc{a}^2\,(1+\textsc{b})^{-1}\right]-\textsc{a}^2\,(1+\textsc{b})^{-2}\right) \end{array} \right\} \otimes \mathbf{\Gamma}_{21}$$

$$+\left\{ \begin{array}{cc} -2\,(1+\textsc{b})^{-2}+\left(3\textsc{b}^2\,(1+\textsc{b})^{-2}-5\textsc{b}\,(1+\textsc{b})^{-1}+2\right)\kappa & \textsc{a}\,(1+\textsc{b})^{-2}\,(2-3\kappa) \\ \textsc{a}\,(1+\textsc{b})^{-2}\,(2-3\kappa) & -2\textsc{a}^2\,(1+\textsc{b})^{-2}+\{3\textsc{a}^2\,(1+\textsc{b})^{-2}-\left[\textsc{c}-\textsc{a}^2\,(1+\textsc{b})^{-1}\right]\}\kappa \end{array} \right\} \otimes \mathbf{\Gamma}_{21}\mathbf{\Gamma}_{11}^{-1}\boldsymbol{\nu}_1\boldsymbol{\nu}_1'$$

$$+\begin{pmatrix} 0 & 0 \\ 0 & 1+2\kappa \end{pmatrix} \otimes \mathbf{\Gamma}_{21}\mathbf{\Gamma}_{11}^{-1}\ell_{N_1}\ell_{N_1}'$$

$$+\begin{bmatrix} 0 & (1+\textsc{b})^{-1}\,\kappa \\ (1+\textsc{b})^{-1}\,\kappa & -2\textsc{a}\,(1+\textsc{b})^{-1}\,\kappa \end{bmatrix} \otimes \mathbf{\Gamma}_{21}\mathbf{\Gamma}_{11}^{-1}\left(\boldsymbol{\nu}_1\ell_{N_1}'+\ell_{N_1}\boldsymbol{\nu}_1'\right),$$

where $\mathbf{S}_{U11}(\boldsymbol{\phi})$ has been obtained from Lemma D1 in Appendix D below. 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 $\mathbf{D}_{U2}\mathbf{D}_{U1}^{-1}\mathbf{S}_{U11}(\boldsymbol{\phi}^0) = \mathbf{S}_{U21}(\boldsymbol{\phi}^0)$, where $\mathbf{S}_{U21}(\boldsymbol{\phi})$ can also be obtained from Lemma D1. $\square$

Lemma 4:

Let us begin again by re-ordering the estimating functions (8) as

$$\begin{bmatrix} \mathbf{R}_{1t}-\boldsymbol{\nu}_1 \\ (\mathbf{R}_{1t}-\boldsymbol{\nu}_1)(\mathbf{R}_t-\boldsymbol{\nu})'\boldsymbol{\varphi}^+-\mathbf{R}_{1t} \\ (\mathbf{R}_{1t}-\boldsymbol{\nu}_1)(\mathbf{R}_t-\boldsymbol{\nu})'\boldsymbol{\varphi}^*-\ell_{N_1} \\ \mathbf{R}_{2t}-\boldsymbol{\nu}_2 \\ (\mathbf{R}_{2t}-\boldsymbol{\nu}_2)(\mathbf{R}_t-\boldsymbol{\nu})'\boldsymbol{\varphi}^+-\mathbf{R}_{2t} \\ (\mathbf{R}_{2t}-\boldsymbol{\nu}_2)(\mathbf{R}_t-\boldsymbol{\nu})'\boldsymbol{\varphi}^*-\ell_{N_2} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_{M1}(\mathbf{R}_{1t};\boldsymbol{\nu}_1) \\ \mathbf{h}_{C1}(\mathbf{R}_{1t};\boldsymbol{\varphi}_1^+,\boldsymbol{\varphi}_1^*,\boldsymbol{\nu}_1) \\ \mathbf{h}_{M2}(\mathbf{R}_{2t};\boldsymbol{\nu}_2) \\ \mathbf{h}_{C2}(\mathbf{R}_t;\boldsymbol{\varphi}_1^+,\boldsymbol{\varphi}_1^*,\boldsymbol{\nu}_1,\boldsymbol{\nu}_2) \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{h}_{E1}(\mathbf{R}_{1t};\boldsymbol{\varphi}_1^+,\boldsymbol{\varphi}_1^*,\boldsymbol{\nu}_1) \\ \mathbf{h}_{E2}(\mathbf{R}_t;\boldsymbol{\varphi}_1^+,\boldsymbol{\varphi}_1^*,\boldsymbol{\nu}_1,\boldsymbol{\nu}_2) \end{bmatrix},$$

Once more, we have to check that condition (C1) in Lemma C1 is satisfied under the null hypothesis $H_0: \boldsymbol{\varphi}_2^+ = \mathbf{0}$, $\boldsymbol{\varphi}_2^* = \mathbf{0}$, where in this case $\boldsymbol{\theta} = (\boldsymbol{\nu}_1,\ \boldsymbol{\varphi}_1^{+\prime},\ \boldsymbol{\varphi}_1^{*\prime})'$ and $\boldsymbol{\rho} = \boldsymbol{\nu}_2$. To do so, we must first orthogonalise the two blocks of moment conditions by regressing the second set of estimating functions evaluated at $\boldsymbol{\theta}^0$ and $\boldsymbol{\rho}^0$ onto the first one evaluated at $\boldsymbol{\theta}^0$. The regression coefficients, though, are not uniquely defined since the singularity in $\mathbf{h}_E(\mathbf{R}_t;\boldsymbol{\varphi},\boldsymbol{\nu})$ is confined to $\mathbf{h}_{E1}(\mathbf{R}_t;\boldsymbol{\theta})$ under H_0. Nevertheless, it is easy to see that $\mathbf{g}_2(\mathbf{R}_t;\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)$, where $\mathbf{g}_2(\mathbf{R}_t;\boldsymbol{\theta},\boldsymbol{\rho}) = \mathbf{h}_2(\mathbf{R}_t;\boldsymbol{\theta},\boldsymbol{\rho}) - (\mathbf{I}_3 \otimes \mathbf{B}^0)\mathbf{h}_{E1}(\mathbf{R}_t;\boldsymbol{\theta})$, will be orthogonal to $\mathbf{h}_{E1}\left(\mathbf{R}_t;\boldsymbol{\theta}^0\right)$ because $\mathbf{S}_{21}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0) = (\mathbf{I}_3 \otimes \mathbf{B}^0)\mathbf{S}_{11}(\boldsymbol{\theta}^0)$. In this respect, note that $\mathbf{B}$ can be interpreted as the coefficients in the multivariate regression of $\mathbf{R}_{2t} - \boldsymbol{\nu}_2$ on $\mathbf{R}_{1t} - \boldsymbol{\nu}_1$, while $(\mathbf{I}_2 \otimes \mathbf{B})$ can be interpreted as the coefficients in the multivariate regression of

$$\begin{bmatrix} (\mathbf{R}_{2t} - \boldsymbol{\nu}_2)(\mathbf{R}_{1t} - \boldsymbol{\nu}_1)'\boldsymbol{\varphi}_1^+ - \mathbf{R}_{2t} \\ (\mathbf{R}_{2t} - \boldsymbol{\nu}_2)(\mathbf{R}_{1t} - \boldsymbol{\nu}_1)'\boldsymbol{\varphi}_1^* - \ell_{N_2} \end{bmatrix}$$

on

$$\begin{bmatrix} (\mathbf{R}_{1t} - \boldsymbol{\nu}_1)(\mathbf{R}_{1t} - \boldsymbol{\nu}_1)'\boldsymbol{\varphi}_1^+ - \mathbf{R}_{1t} \\ (\mathbf{R}_{1t} - \boldsymbol{\nu}_1)(\mathbf{R}_{1t} - \boldsymbol{\nu}_1)'\boldsymbol{\varphi}_1^* - \ell_{N_1} \end{bmatrix}.$$

The next step is to find out the appropriate reparametrisation that exploits the singularity in $\mathbf{h}_{E1}\left(\mathbf{R}_t;\boldsymbol{\theta}\right)$, together with the Moore-Penrose inverses of the covariance matrices of $\mathbf{h}_{E1}\left(\mathbf{R}_t;\boldsymbol{\theta}^0\right)$ and $\mathbf{g}_2\left(\mathbf{R}_t;\boldsymbol{\theta}^0,\boldsymbol{\rho}^0\right)$. Given that the latter covariance matrix has full rank, we should concentrate on the first block. Specifically, we could work with the reparametrisation $\boldsymbol{\theta} = \mathbf{l}_1(\boldsymbol{\theta}_\oplus,\mathbf{0}) = (\ \boldsymbol{\nu}_{1\oplus}' \ \ \boldsymbol{\varphi}_{1\ominus}^+ \ \ \boldsymbol{\varphi}_{1\oplus}^{+\prime} \ \ \boldsymbol{\varphi}_{1\oplus}^{*\prime}\ )'$, where $\boldsymbol{\nu}_{1\oplus} = \boldsymbol{\nu}_1$, $\boldsymbol{\varphi}_{1\oplus}^+$ contains the last $N_1 - 1$ elements of $\boldsymbol{\varphi}_1^+$, $\boldsymbol{\varphi}_{1\oplus}^* = \boldsymbol{\varphi}_1^*$, and $\boldsymbol{\varphi}_{1\ominus}^+ = \boldsymbol{\varphi}_1^{*\prime}\boldsymbol{\nu}_1 - \boldsymbol{\varphi}_1^{+\prime}\ell_{N_1}$, together with the non-singular set of moment conditions

$$\begin{aligned} \mathbf{h}_{1\oplus}[\mathbf{R}_t;\boldsymbol{\theta}_\oplus,\boldsymbol{\theta}_\ominus] &= \mathbf{P}_{1\oplus}'(\boldsymbol{\theta}^0)\mathbf{h}_1[\mathbf{R}_t;\mathbf{l}_1(\boldsymbol{\theta}_\oplus,\boldsymbol{\theta}_\ominus)], \\ \mathbf{g}_{2\oplus}[\mathbf{R}_t;\boldsymbol{\theta}_\oplus,\boldsymbol{\theta}_\ominus,\boldsymbol{\rho}] &= \mathbf{P}_2'(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{g}_2[\mathbf{R}_t;\mathbf{l}_1(\boldsymbol{\theta}_\oplus,\boldsymbol{\theta}_\ominus),\boldsymbol{\rho}], \end{aligned}$$

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

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

$$E\left[\frac{\partial \mathbf{g}_{2\oplus}\left(\mathbf{R}_t;\boldsymbol{\theta}_\oplus^0,\mathbf{0},\boldsymbol{\rho}^0\right)}{\partial \boldsymbol{\theta}_\oplus'}\right] \in \left\langle E\left[\frac{\partial \mathbf{g}_{2\oplus}\left(\mathbf{R}_t;\boldsymbol{\theta}_\oplus^0,\mathbf{0},\boldsymbol{\rho}^0\right)}{\partial \boldsymbol{\rho}'}\right]\right\rangle,$$

But since $\mathbf{P}_2(\boldsymbol{\theta}^0, \boldsymbol{\rho}^0)$ has full rank, an equivalent condition in terms of the original moment conditions and parameters is

$$E\left[\frac{\partial \mathbf{h}_2\left(\mathbf{R}_t; \boldsymbol{\theta}^0, \boldsymbol{\rho}^0\right)}{\partial \boldsymbol{\theta}'} - (\mathbf{I}_3 \otimes \mathbf{B}^0)\frac{\partial \mathbf{h}_1\left(\mathbf{R}_t; \boldsymbol{\theta}^0\right)}{\partial \boldsymbol{\theta}'}\right]\mathbf{L}_{1\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus, \mathbf{0}) \in \left\langle E\left[\frac{\partial \mathbf{h}_2\left(\mathbf{R}_t; \boldsymbol{\theta}^0, \boldsymbol{\rho}^0\right)}{\partial \boldsymbol{\rho}'}\right]\right\rangle,$$

where $\mathbf{L}_{1\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus, \mathbf{0}) = \partial \mathbf{l}_1(\boldsymbol{\theta}_\oplus, \mathbf{0})/\partial \boldsymbol{\theta}'_\oplus$. Tedious algebra then shows that

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

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

so that

$$E\left[\frac{\partial \mathbf{h}_2\left(\mathbf{R}_t; \boldsymbol{\theta}^0, \boldsymbol{\rho}^0\right)}{\partial \boldsymbol{\theta}'} - (\mathbf{I}_3 \otimes \mathbf{B})\frac{\partial \mathbf{h}_1\left(\mathbf{R}_t; \boldsymbol{\theta}^0\right)}{\partial \boldsymbol{\theta}'}\right] = \begin{pmatrix} \mathbf{B} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

In addition, since

$$\mathbf{L}_{1\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus, \mathbf{0}) = \begin{pmatrix} \mathbf{I}_{N_1} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\varphi}_1^{*\prime} & -\boldsymbol{\ell}'_{N_1-1} & \boldsymbol{\nu}'_1 \\ \mathbf{0} & \mathbf{I}_{N_1-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{N_1} \end{pmatrix},$$

then

$$E\left[\frac{\partial \mathbf{h}_2\left(\mathbf{R}_t; \boldsymbol{\theta}^0, \boldsymbol{\rho}^0\right)}{\partial \boldsymbol{\theta}'} - (\mathbf{I}_3 \otimes \mathbf{B})\frac{\partial \mathbf{h}_1\left(\mathbf{R}_t; \boldsymbol{\theta}^0\right)}{\partial \boldsymbol{\theta}'}\right]\mathbf{L}_\oplus(\boldsymbol{\theta}^0_\oplus, \mathbf{0}) = \begin{pmatrix} \mathbf{B} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where the only change with respect to the previous expression is that the row dimension of the second row of blocks is reduced by one. Finally, the result follows because $E\left[\partial \mathbf{h}'_2\left(\mathbf{R}_t; \boldsymbol{\theta}^0, \boldsymbol{\rho}^0\right)/\partial \boldsymbol{\rho}\right] = \begin{pmatrix} -\mathbf{I}_{N_2} & \mathbf{0} & \mathbf{0} \end{pmatrix}$. □

Lemma 5:

First of all, it is important to note that under the null hypothesis, the relationship between the estimating functions $\mathbf{h}_L(\mathbf{R}_t; \mathbf{b}_2)$ and $\mathbf{h}_H(\mathbf{R}_t; \mathbf{a}, \mathbf{b})$, where $\mathbf{b}_2 = vec(\mathbf{B}_2)$, will be given by the following full rank linear transformation with constant coefficients

$$\mathbf{h}_L(\mathbf{R}_t;\mathbf{b}_2) = (\digamma \otimes \mathbf{I}_{N_2})\mathbf{h}_H(\mathbf{R}_t;\mathbf{a},\mathbf{b}),$$

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

Hence, we will have that $\mathbf{S}_L(\mathbf{b}_2^0) = (\digamma \otimes \mathbf{I}_{N_2})\mathbf{S}_H(\mathbf{a}^0,\mathbf{b}^0)(\digamma'\otimes\mathbf{I}_{N_2})$ under H_0, where the expression for $\mathbf{S}_H(\mathbf{a},\mathbf{b})$ is given in Lemma D3 below.

Similarly, the Jacobian of the moment conditions (12) will be

$$\mathbf{D}_L(\mathbf{b}_2) = -\left[\digamma \begin{pmatrix} 1 & \boldsymbol{\nu}_1' \\ \boldsymbol{\nu}_1 & \boldsymbol{\Gamma}_{11} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ -\ell_{N_1-1}' \\ \mathbf{I}_{N_1-1} \end{pmatrix} \otimes \mathbf{I}_{N_2}\right].$$

As a result, $\mathbf{D}_L'(\mathbf{b}_2)\mathbf{S}_L^{-1}(\mathbf{b}_2)$ will be equal to

$$-\left\{\begin{pmatrix} \mathbf{0} & -\ell_{N_1-1} & \mathbf{I}_{N_1-1} \end{pmatrix} \begin{pmatrix} 1 & \boldsymbol{\nu}_1' \\ \boldsymbol{\nu}_1 & \boldsymbol{\Gamma}_{11} \end{pmatrix} \begin{bmatrix} 1 & \boldsymbol{\nu}_1' \\ \boldsymbol{\nu}_1 & (\kappa+1)\boldsymbol{\Sigma}_{11} + \boldsymbol{\nu}_1\boldsymbol{\nu}_1' \end{bmatrix}^{-1} \digamma^{-1} \otimes \boldsymbol{\Omega}^{-1}\right\}$$

$$= -\left\{\begin{pmatrix} \mathbf{0} & -\ell_{N_1-1} & \mathbf{I} \end{pmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ \kappa(\kappa+1)^{-1}\boldsymbol{\nu}_1 & (\kappa+1)^{-1}\mathbf{I}_{N_1} \end{bmatrix} \begin{pmatrix} 1 & 0 & \mathbf{0} \\ 0 & 1 & \mathbf{0} \\ \mathbf{0} & \ell_{N_1-1} & \mathbf{I}_{N_1-1} \end{pmatrix} \otimes \boldsymbol{\Omega}^{-1}\right\}$$

$$= -\left\{(\kappa+1)^{-1}[\ \kappa\,(\boldsymbol{\nu}_{1b} - \nu_{1a}\ell_{N_1-1}) \quad \mathbf{0} \quad \mathbf{I}_{N_1-1}\ ] \otimes \boldsymbol{\Omega}^{-1}\right\}.$$

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

Lemma 6:

Using the notation of section 3.1, we can write the approximate slope of the Wald version of the uncentred RP spanning test as $\mathbf{G}_U'(\boldsymbol{\phi}^0)\mathbf{F}_U^{-1}(\boldsymbol{\phi}^0)\mathbf{G}_U(\boldsymbol{\phi}^0)$, where

$$\mathbf{F}_U(\boldsymbol{\phi}^0) = \lim_{T\to\infty} V\{\sqrt{T}[\mathbf{G}_U(\hat{\boldsymbol{\phi}}_T) - \mathbf{G}_U(\boldsymbol{\phi}^0)]\}.$$

However, it is more convenient to express this matrix as $\lim_{T\to\infty} V[\sqrt{T}\bar{\mathbf{g}}_{UT}(\boldsymbol{\phi}^0)]$, where $\mathbf{g}_U(\mathbf{R}_t;\boldsymbol{\phi}) = \mathbf{Q}_U(\boldsymbol{\phi})\mathbf{D}_U^{-1\prime}(\boldsymbol{\phi})\mathbf{h}_U(\mathbf{R}_t;\boldsymbol{\phi})$, so that

$$\mathbf{F}_U(\boldsymbol{\phi}^0) = \mathbf{Q}_U(\boldsymbol{\phi}^0)\mathbf{D}_U^{-1\prime}(\boldsymbol{\phi}^0)\mathbf{S}_U(\boldsymbol{\phi}^0)\mathbf{D}_U^{-1}(\boldsymbol{\phi}^0)\mathbf{Q}_U'(\boldsymbol{\phi}^0).$$

But since

$$\mathbf{Q}_U(\boldsymbol{\phi}^0)\mathbf{D}_U^{-1\prime}(\boldsymbol{\phi}^0) = [\mathbf{I}_2 \otimes \begin{pmatrix} \mathbf{0} & \mathbf{I}_{N_2} \end{pmatrix}](\mathbf{I}_2 \otimes \boldsymbol{\Gamma}^{-1}) = \mathbf{I}_2 \otimes \begin{pmatrix} -\boldsymbol{\Lambda}^{-1}\mathbf{C} & \boldsymbol{\Lambda}^{-1} \end{pmatrix}, \tag{B2}$$

and

$$\mathbf{h}_U\left(\mathbf{R}_t;\boldsymbol{\phi}\right) = \begin{bmatrix} \mathbf{R}_t\left(p_t^+ - 1\right) \\ \mathbf{R}_t p_t^* - \ell_N \end{bmatrix},$$

we can write

$$\mathbf{g}_U\left(\mathbf{R}_t;\boldsymbol{\phi}\right) = \left[\mathbf{I}_2 \otimes \boldsymbol{\Lambda}^{-1}\right] \begin{bmatrix} \mathbf{v}_t\left(p_t^+ - 1\right) \\ \mathbf{v}_t p_t^* - (\ell_{N_2} - \mathbf{C}\ell_{N_1}) \end{bmatrix}.$$

We can proceed in the same manner with the centred RP test. Given that H_0 does not involve $\boldsymbol{\nu}$, and the Jacobian matrix is diagonal, the approximate slope of this test reduces to $\mathbf{G}_C'(\boldsymbol{\varphi}^0)\mathbf{F}_C^{-1}(\boldsymbol{\varphi}^0)\mathbf{G}_C(\boldsymbol{\varphi}^0) = \mathbf{G}_U'(\boldsymbol{\phi}^0)\boldsymbol{\Psi}_C'\mathbf{F}_C^{-1}(\boldsymbol{\varphi}^0)\boldsymbol{\Psi}_C\mathbf{G}_U(\boldsymbol{\phi}^0)$, where $\boldsymbol{\Psi}_C$ is given by (A3). Specifically, we can interpret $\boldsymbol{\Psi}_C^{-1}\mathbf{F}_C(\boldsymbol{\varphi}^0)\boldsymbol{\Psi}_C^{-1\prime}$ as $\lim_{T\to\infty} V[\sqrt{T}\bar{\mathbf{g}}_{CT}(\boldsymbol{\varphi}^0)]$, where $\mathbf{g}_C\left(\mathbf{R}_t;\boldsymbol{\varphi}^0\right) = \boldsymbol{\Psi}_C^{-1}\mathbf{Q}_C\mathbf{D}_C^{-1\prime}\mathbf{h}_C\left(\mathbf{R}_t;\boldsymbol{\varphi}^0\right)$. But since

$$\mathbf{Q}_C(\boldsymbol{\varphi}^0)\mathbf{D}_C^{-1\prime}(\boldsymbol{\varphi}^0) = [\mathbf{I}_2 \otimes \begin{pmatrix} \mathbf{0} & \mathbf{I}_{N_2} \end{pmatrix}](\mathbf{I}_2 \otimes \boldsymbol{\Sigma}^{-1}) = \mathbf{I}_2 \otimes \begin{pmatrix} -\boldsymbol{\Omega}^{-1}\mathbf{B} & \boldsymbol{\Omega}^{-1} \end{pmatrix}, \tag{B3}$$

and

$$\begin{aligned}\mathbf{h}_C(\mathbf{R}_t;\boldsymbol{\varphi}^0) &= \begin{pmatrix} (1+\mathrm{B})\mathbf{R}_t(p_t^+ - 1) \\ (\mathbf{R}_t p_t^* - \ell_N) + \mathrm{A}\mathbf{R}_t(p_t^+ - 1) \end{pmatrix} + \left\{ \begin{bmatrix} \mathrm{B} - (1+\mathrm{B})p_t^+ \\ \mathrm{A} - p_t^* - \mathrm{A}p_t^+ \end{bmatrix} \otimes \boldsymbol{\nu} \right\} \\ &= \left[\begin{pmatrix} 1+\mathrm{B} & 0 \\ \mathrm{A} & 1 \end{pmatrix} \otimes \mathbf{I}_N \right] \left[\begin{pmatrix} \mathbf{R}_t(p_t^+ - 1) \\ \mathbf{R}_t p_t^* - \ell_N \end{pmatrix} - \left\{ \begin{bmatrix} p_t^+ - \mathrm{B}(1+\mathrm{B})^{-1} \\ p_t^* - \mathrm{A}(1+\mathrm{B})^{-1} \end{bmatrix} \otimes \boldsymbol{\nu} \right\} \right]\end{aligned}$$

because of the relationships between centred RP and uncentred RP described in section 2.1, we eventually get

$$\mathbf{g}_C\left(\mathbf{R}_t;\boldsymbol{\varphi}^0\right) = \left[\begin{pmatrix} 1+\mathrm{B}_1 & 0 \\ \mathrm{A}_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1+\mathrm{B} & 0 \\ \mathrm{A} & 1 \end{pmatrix} \otimes \boldsymbol{\Lambda}^{-1} \right] \begin{pmatrix} \mathbf{w}_t(p_t^+ - 1) - (1+\mathrm{B})^{-1}\mathbf{a} \\ \mathbf{w}_t p_t^* - \mathbf{f} + \mathrm{A}(1+\mathrm{B})^{-1}\mathbf{a} \end{pmatrix}.$$

Finally, the approximate slope of the regression test will be

$$\mathbf{G}_H'(\mathbf{a}^0,\mathbf{b}^0)\mathbf{F}_H^{-1}(\mathbf{a}^0,\mathbf{b}^0)\mathbf{G}_H(\mathbf{a}^0,\mathbf{b}^0) = \mathbf{G}_U'(\boldsymbol{\phi}^0)\boldsymbol{\Psi}_H'\mathbf{F}_H^{-1}(\mathbf{a}^0,\mathbf{b}^0)\boldsymbol{\Psi}_H\mathbf{G}_U(\boldsymbol{\phi}^0),$$

where $\boldsymbol{\Psi}_H$ is given by (A4). Once more, we can interpret $\boldsymbol{\Psi}_H^{-1}\mathbf{F}_H(\mathbf{a}^0,\mathbf{b}^0)\boldsymbol{\Psi}_H^{-1\prime}$ as the asymptotic variance of $\sqrt{T}\bar{\mathbf{g}}_{HT}(\mathbf{a}^0,\mathbf{b}^0)]$, where

$$\mathbf{g}_H\left(\mathbf{R}_t;\mathbf{a}^0,\mathbf{b}^0\right) = \boldsymbol{\Psi}_H^{-1}\mathbf{Q}_H(\mathbf{a}^0,\mathbf{b}^0)\mathbf{D}_H^{-1\prime}(\mathbf{a}^0,\mathbf{b}^0)\mathbf{h}_H(\mathbf{R}_t;\mathbf{a}^0,\mathbf{b}^0).$$

But since

$$\mathbf{Q}_H(\mathbf{a}^0,\mathbf{b}^0)\mathbf{D}_H^{-1\prime}(\mathbf{a}^0,\mathbf{b}^0) = \left[\begin{pmatrix} -1 & \mathbf{0}' \\ 0 & \ell'_{N_1}\end{pmatrix}\begin{pmatrix} 1 & \boldsymbol{\nu}'_1 \\ \boldsymbol{\nu}_1 & \boldsymbol{\Gamma}_{11}\end{pmatrix}^{-1} \otimes \mathbf{I}_{N_2}\right]$$

$$= \begin{pmatrix} -(1+\mathrm{B}_1) & \boldsymbol{\nu}'_1\boldsymbol{\Sigma}_{11}^{-1} \\ -\mathrm{A}_1 & \ell'_{N_1}\boldsymbol{\Sigma}_{11}^{-1}\end{pmatrix} \otimes \mathbf{I}_{N_2}, \qquad \text{(B4)}$$

and $\mathbf{h}_H(\mathbf{R}_t;\mathbf{a}^0,\mathbf{b}^0) = (1,\mathbf{R}'_{1t})' \otimes \mathbf{w}_t$, we will end up with

$$\mathbf{g}_H(\mathbf{R}_t;\mathbf{a}^0,\mathbf{b}^0) = [\mathbf{I}_2 \otimes \boldsymbol{\Lambda}^{-1}] \begin{bmatrix} \mathbf{w}_t(p_{1t}^+ - 1) \\ \mathbf{w}_t p_{1t}^* \end{bmatrix}.$$

Hence, we have expressed $\mathbf{g}_U(\mathbf{R}_t;\boldsymbol{\phi}^0)$, $\mathbf{g}_C(\mathbf{R}_t;\boldsymbol{\varphi}^0)$ and $\mathbf{g}_H(\mathbf{R}_t;\mathbf{a}^0,\mathbf{b}^0)$ as simple bilinear functions of $\mathbf{w}_t$, p_t^+ and p_t^*, and their counterparts under the null, i.e. $\mathbf{v}_t$, p_{1t}^+ and p_{1t}^*, respectively. The relationship between these random quantities is given by (17) together with $\mathbf{w}_t = \mathbf{v}_t - \mathbf{a}(1-p_{1t}^+)$, $\boldsymbol{\Lambda} = \boldsymbol{\Omega} + (1{+}\mathrm{B}_1)^{-1}\mathbf{a}\mathbf{a}'$, $\mathrm{A}{=}\mathrm{A}_1 + \mathbf{a}'\boldsymbol{\Omega}^{-1}\mathbf{f}$, and $\mathrm{B}{=}\mathrm{B}_1 + \mathbf{a}'\boldsymbol{\Omega}^{-1}\mathbf{a}$.

Finally, it is straightforward to see that $\mathbf{g}_C(\mathbf{R}_t;\boldsymbol{\varphi}^0) = \mathbf{g}_U(\mathbf{R}_t;\boldsymbol{\phi}^0)$ when $\mathbf{a}^0 = \mathbf{0}$. As a result, the corresponding approximate slopes will be the same too. $\square$

Lemma 7:

Given that H_0 does not involve $\boldsymbol{\nu}$, and that the Jacobian matrix is diagonal, we can treat the elements of $\boldsymbol{\nu}$ as if they were known without loss of generality. If we combine the expression (B3) for $\mathbf{Q}_C(\boldsymbol{\varphi}^0)\mathbf{D}_C^{-1\prime}(\boldsymbol{\varphi}^0)$ with the expression for $\mathbf{S}_C(\boldsymbol{\varphi}^0)$ in Lemma D2 below, we end up with

$$\mathbf{F}_C(\boldsymbol{\varphi}^0) = \mathbf{Q}_C(\boldsymbol{\varphi}^0)\mathbf{D}_C^{-1\prime}(\boldsymbol{\varphi}^0)\mathbf{S}_C(\boldsymbol{\varphi}^0)\mathbf{Q}_C(\boldsymbol{\varphi}^0)\mathbf{D}_C^{-1}(\boldsymbol{\varphi}^0)\mathbf{Q}'_C(\boldsymbol{\varphi}^0)$$

$$= \left[\mathbf{I}_2 \otimes \begin{pmatrix} -\boldsymbol{\Omega}^{-1}\mathbf{B} & \boldsymbol{\Omega}^{-1}\end{pmatrix}\right] \left\{ \begin{bmatrix} 1+(\kappa+1)\mathrm{B} & (\kappa+1)\mathrm{A} \\ (\kappa+1)\mathrm{A} & (\kappa+1)\mathrm{C}\end{bmatrix} \otimes \boldsymbol{\Sigma} \right.$$

$$\left. + \begin{pmatrix} (2\kappa+1)\boldsymbol{\nu}\boldsymbol{\nu}' & (\kappa+1)\ell_N\boldsymbol{\nu}' + \kappa\boldsymbol{\nu}\ell'_N \\ (\kappa+1)\boldsymbol{\nu}\ell'_N + \kappa\ell_N\boldsymbol{\nu}' & (2\kappa+1)\ell_N\ell'_N \end{pmatrix} \right\} \left[\mathbf{I}_2 \otimes \begin{pmatrix} -\mathbf{B}'\boldsymbol{\Omega}^{-1} \\ \boldsymbol{\Omega}^{-1}\end{pmatrix}\right]$$

$$= \begin{pmatrix} 1+(\kappa+1)\mathrm{B} & (\kappa+1)\mathrm{A} \\ (\kappa+1)\mathrm{A} & (\kappa+1)\mathrm{C}\end{pmatrix} \otimes \boldsymbol{\Omega}^{-1}$$

$$+ \begin{pmatrix} (2\kappa+1)\,\boldsymbol{\varphi}_2^+\boldsymbol{\varphi}_2^{+\prime} & (\kappa+1)\boldsymbol{\varphi}_2^*\boldsymbol{\varphi}_2^{+\prime} + \kappa\boldsymbol{\varphi}_2^+\boldsymbol{\varphi}_2^{*\prime} \\ (\kappa+1)\boldsymbol{\varphi}_2^+\boldsymbol{\varphi}_2^{*\prime} + \kappa\boldsymbol{\varphi}_2^*\boldsymbol{\varphi}_2^{+\prime} & (2\kappa+1)\,\boldsymbol{\varphi}_2^*\boldsymbol{\varphi}_2^{*\prime}\end{pmatrix}.$$

Similarly, if we combine the expression for $\mathbf{Q}_H(\mathbf{a}^0,\mathbf{b}^0)\mathbf{D}_H^{-1\prime}(\mathbf{a}^0,\mathbf{b}^0)$ in (B1) with the expression for $\mathbf{S}_H$ in Lemma D3 (see again Appendix D), we end up with

$$\mathbf{F}_H(\mathbf{a}^0,\mathbf{b}^0)=\mathbf{Q}_H(\mathbf{a}^0,\mathbf{b}^0)\mathbf{D}_H^{-1\prime}(\mathbf{a}^0,\mathbf{b}^0)\mathbf{S}_H(\mathbf{a}^0,\mathbf{b}^0)\mathbf{Q}_H(\mathbf{a}^0,\mathbf{b}^0)\mathbf{D}_H^{-1}(\mathbf{a}^0,\mathbf{b}^0)\mathbf{Q}_H^{\prime}(\mathbf{a}^0,\mathbf{b}^0)$$
$$=\left[\begin{pmatrix}-(1+\mathrm{B}_1) & \boldsymbol{\nu}_1'\boldsymbol{\Sigma}_{11}^{-1}\\ -\mathrm{A}_1 & \ell_{N_1}'\boldsymbol{\Sigma}_{11}^{-1}\end{pmatrix}\otimes\mathbf{I}_{N_2}\right]\left[\begin{pmatrix}1 & \boldsymbol{\nu}_1'\\ \boldsymbol{\nu}_1 & (\kappa+1)\boldsymbol{\Sigma}_{11}+\boldsymbol{\nu}_1\boldsymbol{\nu}_1'\end{pmatrix}\otimes\boldsymbol{\Omega}\right]$$
$$\times\left[\begin{pmatrix}-(1+\mathrm{B}_1) & -\mathrm{A}_1\\ \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\nu}_1 & \boldsymbol{\Sigma}_{11}^{-1}\ell_{N_1}\end{pmatrix}\otimes\mathbf{I}_{N_2}\right]=\begin{bmatrix}1+(\kappa+1)\mathrm{B}_1 & (\kappa+1)\mathrm{A}_1\\ (\kappa+1)\mathrm{A}_1 & (\kappa+1)\mathrm{C}_1\end{bmatrix}\otimes\boldsymbol{\Omega}.$$

In addition,

$$\mathbf{G}_C(\boldsymbol{\varphi})=\begin{pmatrix}\boldsymbol{\varphi}_2^{+}\\ \boldsymbol{\varphi}_2^{*}\end{pmatrix}=(\mathbf{I}_2\otimes\boldsymbol{\Omega}^{-1})\begin{pmatrix}\mathbf{a}\\ \mathbf{f}\end{pmatrix}=(\mathbf{I}_2\otimes\boldsymbol{\Omega}^{-1})\mathbf{G}_H(\mathbf{a},\mathbf{b}).$$

so that we can write

$$\mathbf{G}_H'(\mathbf{a},\mathbf{b})\mathbf{F}_H^{-1}(\mathbf{a},\mathbf{b})\mathbf{G}_H(\mathbf{a},\mathbf{b})-\mathbf{G}_C'(\boldsymbol{\varphi})\mathbf{F}_C^{-1}(\boldsymbol{\varphi})\mathbf{G}_C(\boldsymbol{\varphi})$$
$$=\mathbf{G}_H'(\mathbf{a},\mathbf{b})\mathbf{F}_H^{-1}(\mathbf{a},\mathbf{b})\mathbf{G}_H(\mathbf{a},\mathbf{b})-\mathbf{G}_H'(\mathbf{a},\mathbf{b})(\mathbf{I}_2\otimes\boldsymbol{\Omega}^{-1})\mathbf{F}_C^{-1}(\boldsymbol{\varphi})(\mathbf{I}_2\otimes\boldsymbol{\Omega}^{-1})\mathbf{G}_H(\mathbf{a},\mathbf{b})$$
$$=\mathbf{G}_H'(\mathbf{a},\mathbf{b})[\mathbf{F}_H^{-1}(\mathbf{a},\mathbf{b})-(\mathbf{I}_2\otimes\boldsymbol{\Omega}^{-1})\mathbf{F}_C^{-1}(\boldsymbol{\varphi})(\mathbf{I}_2\otimes\boldsymbol{\Omega}^{-1})]\mathbf{G}_H(\mathbf{a},\mathbf{b}).$$

Hence, a sufficient condition for the desired result would be that $(\mathbf{I}_2\otimes\boldsymbol{\Omega})\mathbf{F}_C(\boldsymbol{\varphi})(\mathbf{I}_2\otimes\boldsymbol{\Omega})-\mathbf{F}_H(\mathbf{a},\mathbf{b})$ is a positive semidefinite matrix. But this difference is

$$\left[\begin{pmatrix}1+(\kappa+1)\mathrm{B} & (\kappa+1)\mathrm{A}\\ (\kappa+1)\mathrm{A} & (\kappa+1)\mathrm{C}\end{pmatrix}\otimes\boldsymbol{\Omega}\right]+\begin{pmatrix}(2\kappa+1)\mathbf{a}\mathbf{a}' & (\kappa+1)\mathbf{f}\mathbf{a}'+\kappa\mathbf{a}\mathbf{f}'\\ (\kappa+1)\mathbf{a}\mathbf{f}'+\kappa\mathbf{f}\mathbf{a}' & (2\kappa+1)\mathbf{f}\mathbf{f}'\end{pmatrix}$$
$$-\begin{pmatrix}1+(\kappa+1)\mathrm{B}_1 & (\kappa+1)\mathrm{A}_1\\ (\kappa+1)\mathrm{A}_1 & (\kappa+1)\mathrm{C}_1\end{pmatrix}\otimes\boldsymbol{\Omega}$$
$$=(\kappa+1)\left[\begin{pmatrix}\mathbf{a}'\boldsymbol{\Omega}^{-1}\mathbf{a} & \mathbf{a}'\boldsymbol{\Omega}^{-1}\mathbf{f}\\ \mathbf{a}'\boldsymbol{\Omega}^{-1}\mathbf{f} & \mathbf{f}'\boldsymbol{\Omega}^{-1}\mathbf{f}\end{pmatrix}\otimes\boldsymbol{\Omega}\right]+\begin{bmatrix}(2\kappa+1)\mathbf{a}\mathbf{a}' & (\kappa+1)\mathbf{f}\mathbf{a}'+\kappa\mathbf{a}\mathbf{f}'\\ (\kappa+1)\mathbf{a}\mathbf{f}'+\kappa\mathbf{f}\mathbf{a}' & (2\kappa+1)\mathbf{f}\mathbf{f}'\end{bmatrix}.$$

If we then pre-multiply this expression by $(\mathbf{I}_2\otimes\boldsymbol{\Omega}^{-1/2})$, and post-multiply by its transpose, we end up with

$$\begin{bmatrix}(\kappa+1)(\mathring{\mathbf{a}}'\mathring{\mathbf{a}})\mathbf{I}_{N_2}+(2\kappa+1)\mathring{\mathbf{a}}\mathring{\mathbf{a}}' & (\kappa+1)(\mathring{\mathbf{a}}'\mathring{\mathbf{f}})\mathbf{I}_{N_2}+(\kappa+1)\mathring{\mathbf{f}}\mathring{\mathbf{a}}'+\kappa\mathring{\mathbf{a}}\mathring{\mathbf{f}}'\\ (\kappa+1)(\mathring{\mathbf{a}}'\mathring{\mathbf{f}})\mathbf{I}_{N_2}+(\kappa+1)\mathring{\mathbf{a}}\mathring{\mathbf{f}}'+\kappa\mathring{\mathbf{f}}\mathring{\mathbf{a}}' & (\kappa+1)(\mathring{\mathbf{f}}'\mathring{\mathbf{f}})\mathbf{I}_{N_2}+(2\kappa+1)\mathring{\mathbf{f}}\mathring{\mathbf{f}}'\end{bmatrix}$$
$$=\begin{pmatrix}\mathbf{N}_{11} & \mathbf{N}_{12}\\ \mathbf{N}_{12}' & \mathbf{N}_{22}\end{pmatrix},$$

where $\mathring{\mathbf{a}} = \mathbf{\Omega}^{-1/2}\mathbf{a}$ and $\mathring{\mathbf{f}} = \mathbf{\Omega}^{-1/2}\mathbf{f}$. Given that for elliptical distributions $\kappa \geq -2/(N+2)$ (see Appendix D), and that any spanning test involves at least two assets, then both diagonal blocks are positive definite because $2\kappa + 1 \geq 0$. Therefore, the whole matrix $\mathbf{N}$ will be positive semidefinite if and only if $\mathbf{N}_{22|1} = \mathbf{N}_{22} - \mathbf{N}'_{12}\mathbf{N}_{11}^{-1}\mathbf{N}_{12}$ is also positive semidefinite (We are grateful to Jan Magnus for suggesting this line of proof). But since

$$\mathbf{N}_{11}^{-1} = \frac{1}{(\kappa+1)(\mathring{\mathbf{a}}'\mathring{\mathbf{a}})}\mathbf{I}_{N_2} - \frac{(2\kappa+1)}{(\kappa+1)(3\kappa+2)(\mathring{\mathbf{a}}'\mathring{\mathbf{a}})^2}\mathring{\mathbf{a}}\mathring{\mathbf{a}}',$$

by virtue of the Woodbury formula, it follows that

$$\begin{aligned}\frac{(3\kappa+2)(\mathring{\mathbf{a}}'\mathring{\mathbf{a}})^2}{(\kappa+1)}\mathbf{N}_{22|1} &= (3\kappa+2)\left[(\mathring{\mathbf{a}}'\mathring{\mathbf{a}})^2(\mathring{\mathbf{f}}'\mathring{\mathbf{f}}) - (\mathring{\mathbf{a}}'\mathring{\mathbf{a}})(\mathring{\mathbf{a}}'\mathring{\mathbf{f}})^2\right]\mathbf{I}_{N_2} \\ &\quad -[(3\kappa+2)(\mathring{\mathbf{a}}'\mathring{\mathbf{a}})(\mathring{\mathbf{f}}'\mathring{\mathbf{f}}) - 4(2\kappa+1)(\mathring{\mathbf{a}}'\mathring{\mathbf{f}})^2]\mathring{\mathbf{a}}\mathring{\mathbf{a}}' \\ &\quad +(5\kappa+2)(\mathring{\mathbf{a}}'\mathring{\mathbf{a}})^2(\mathring{\mathbf{f}}\mathring{\mathbf{f}}') - (5\kappa+2)(\mathring{\mathbf{a}}'\mathring{\mathbf{a}})(\mathring{\mathbf{a}}'\mathring{\mathbf{f}})(\mathring{\mathbf{a}}\mathring{\mathbf{f}}' + \mathring{\mathbf{f}}\mathring{\mathbf{a}}').\end{aligned}$$

Let us study the eigenvalues of $\mathbf{N}_{22|1}$. One is clearly 0 since $\mathbf{N}_{22.1}\mathring{\mathbf{a}} = \mathbf{0}$, and another one

$$\frac{4(\kappa+1)(2\kappa+1)}{(3\kappa+2)\,(\mathring{\mathbf{a}}'\mathring{\mathbf{a}})}[(\mathring{\mathbf{a}}'\mathring{\mathbf{a}})(\mathring{\mathbf{f}}'\mathring{\mathbf{f}}) - (\mathring{\mathbf{a}}'\mathring{\mathbf{f}})^2],$$

since

$$\mathbf{N}_{22|1}\left(\mathring{\mathbf{f}} - \frac{\mathring{\mathbf{a}}'\mathring{\mathbf{f}}}{\mathring{\mathbf{a}}'\mathring{\mathbf{a}}}\mathring{\mathbf{a}}\right) = \frac{4(\kappa+1)(2\kappa+1)}{(3\kappa+2)}\left[\mathring{\mathbf{f}}'\mathring{\mathbf{f}} - \frac{(\mathring{\mathbf{a}}'\mathring{\mathbf{f}})^2}{\mathring{\mathbf{a}}'\mathring{\mathbf{a}}}\right]\left(\mathring{\mathbf{f}} - \frac{\mathring{\mathbf{a}}'\mathring{\mathbf{f}}}{\mathring{\mathbf{a}}'\mathring{\mathbf{a}}}\mathring{\mathbf{a}}\right),$$

where the associated eigenvector is the residual of the Euclidean projection of $\mathring{\mathbf{f}}$ onto the linear span of $\mathring{\mathbf{a}}$. A third eigenvalue is

$$\frac{(\kappa+1)}{(\mathring{\mathbf{a}}'\mathring{\mathbf{a}})}\left[(\mathring{\mathbf{a}}'\mathring{\mathbf{a}})(\mathring{\mathbf{f}}'\mathring{\mathbf{f}}) - (\mathring{\mathbf{a}}'\mathring{\mathbf{f}})^2\right]$$

because for any vector $\mathbf{y}$ orthogonal to the linear span of $\mathring{\mathbf{a}}$ and $\mathring{\mathbf{f}}$

$$\mathbf{N}_{22|1}\mathbf{y} = (\kappa+1)\left[\mathring{\mathbf{f}}'\mathring{\mathbf{f}} - \frac{(\mathring{\mathbf{a}}'\mathring{\mathbf{f}})^2}{\mathring{\mathbf{a}}'\mathring{\mathbf{a}}}\right]\mathbf{y}.$$

But since the orthogonal complement to $\left\langle \mathring{\mathbf{a}}, \mathring{\mathbf{f}} \right\rangle$ is of dimension $N_2 - 2$, then there are no further eigenvalues. Finally, since $(\mathring{\mathbf{a}}'\mathring{\mathbf{a}})\,(\mathring{\mathbf{f}}'\mathring{\mathbf{f}}) - (\mathring{\mathbf{a}}'\mathring{\mathbf{f}})^2 \geq 0$ by virtue of the Cauchy-Schwartz inequality, all the eigenvalues of $\mathbf{N}_{22|1}$ are non-negative, which implies that the matrix $\mathbf{N}$ will be positive semidefinite. $\square$

C A useful GMM result:

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).

Lemma C1 *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*

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

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

$$\begin{bmatrix} \mathbf{I}_{n_1} & \mathbf{0} \\ -\mathbf{S}_{21}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{S}^-_{11}(\boldsymbol{\theta}^0) & \mathbf{I}_{n_2} \end{bmatrix} \begin{bmatrix} \mathbf{S}_{11}(\boldsymbol{\theta}^0) & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22|1}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0) \end{bmatrix} \begin{bmatrix} \mathbf{I}_{n_1} & -\mathbf{S}^{-\prime}_{11}(\boldsymbol{\theta}^0)\mathbf{S}'_{21}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0) \\ \mathbf{0} & \mathbf{I}_{n_2} \end{bmatrix},$$

where $\mathbf{S}^-_{11}(\boldsymbol{\theta}^0)$ is some generalised inverse of $\mathbf{S}_{11}(\boldsymbol{\theta}^0)$, and

$$\mathbf{S}_{22|1}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0) = \mathbf{S}_{22}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0) - \mathbf{S}_{21}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{S}^-_{11}(\boldsymbol{\theta}^0)\mathbf{S}'_{21}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)$$

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

$$\mathbf{g}_2(\mathbf{x}_t;\boldsymbol{\theta},\boldsymbol{\rho}) = \mathbf{h}_2(\mathbf{x}_t;\boldsymbol{\theta},\boldsymbol{\rho}) - \mathbf{S}_{21}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{S}^-_{11}(\boldsymbol{\theta}^0)\mathbf{h}_1(\mathbf{x}_t;\boldsymbol{\theta}),$$

are some transformed estimating functions which are invariant to the choice $\mathbf{S}^-_{11}(\boldsymbol{\theta}^0)$. 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 $(n_1+n_2)\times(k_{1\ominus}+k_{2\ominus})$ matrix of continuously differentiable functions of $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$, with $0 \leq k_{1\ominus} \leq k_1$ and $0 \leq k_{2\ominus} \leq k_2$, such that

$$\begin{bmatrix} \boldsymbol{\Pi}'_{11}(\boldsymbol{\theta}) & \mathbf{0} \\ \boldsymbol{\Pi}'_{12}(\boldsymbol{\theta},\boldsymbol{\rho}) & \boldsymbol{\Pi}'_{22}(\boldsymbol{\theta},\boldsymbol{\rho}) \end{bmatrix} \begin{bmatrix} \mathbf{h}_1(\mathbf{x}_t;\boldsymbol{\theta}^0) \\ \mathbf{g}_2(\mathbf{x}_t;\boldsymbol{\theta},\boldsymbol{\rho}) \end{bmatrix} = \mathbf{0}\ \forall \mathbf{x}_t \Leftrightarrow \begin{bmatrix} \mathbf{m}_{1\ominus}(\boldsymbol{\theta}) \\ \mathbf{m}_{2\ominus}(\boldsymbol{\theta},\boldsymbol{\rho}) \end{bmatrix} = \mathbf{m}_{\ominus}(\boldsymbol{\theta},\boldsymbol{\rho}) = \mathbf{0},$$

where $\mathbf{m}_{\ominus}(\boldsymbol{\theta},\boldsymbol{\rho})$ is a $(k_{1\ominus}+k_{2\ominus})$ continuously differentiable function of $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$ with

$$\text{rank}\left[\frac{\partial \mathbf{m}_{\ominus}(\boldsymbol{\theta})}{\partial(\boldsymbol{\theta}',\boldsymbol{\rho}')}\right] = (k_{1\ominus}+k_{2\ominus})$$

in an open neighbourhood of $(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)$. Moreover, assume that $\mathbf{m}_{\ominus}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0) = \mathbf{0}$ if $k_{1\ominus}+k_{2\ominus} > 0$, and that $\text{rank}[\mathbf{S}_{11}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)] = n_1 - k_{1\ominus}$, $\text{rank}[\mathbf{S}_{22|1}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)] = n_2 - k_{2\ominus}$, so that $\text{rank}[\mathbf{S}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)] = (n_1+n_2)-(k_{1\ominus}+k_{2\ominus})$. Then, subject to the required regularity conditions, the optimal GMM estimators of $\boldsymbol{\theta}$ based on

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

alone are asymptotically as efficient as the optimal GMM estimators that additionally use

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

if and only if

$$\mathbf{D}_{2\oplus\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\in\left\langle\mathbf{D}_{2\oplus\boldsymbol{\rho}_\oplus}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\right\rangle, \tag{C1}$$

where

$$\begin{aligned}\mathbf{D}_{2\oplus\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus,\boldsymbol{\rho}^0_\oplus) &= \mathbf{P}'_{2\oplus}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{D}_{2\boldsymbol{\theta}}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{L}_{1\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}^0_\oplus,\mathbf{0}),\\ \mathbf{D}_{2\oplus\boldsymbol{\rho}_\oplus}(\boldsymbol{\theta}^0_\oplus,\boldsymbol{\rho}^0_\oplus) &= \mathbf{P}'_{2\oplus}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{D}_{2\boldsymbol{\rho}}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{L}_{2\boldsymbol{\rho}_\oplus}(\boldsymbol{\theta}^0_\oplus,\mathbf{0},\boldsymbol{\rho}^0_\oplus,\mathbf{0}),\end{aligned}$$

$\mathbf{P}_{2\oplus}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\boldsymbol{\Delta}_{2\oplus}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{P}'_{2\oplus}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)$ *provides the spectral decompositions of* $\mathbf{S}_{22|1}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)$*,*

$$\begin{aligned}\mathbf{D}_{2\boldsymbol{\theta}}(\boldsymbol{\theta},\boldsymbol{\rho}) &= E\left[\frac{\partial\mathbf{g}_2(\mathbf{x}_t;\boldsymbol{\theta},\boldsymbol{\rho})}{\partial\boldsymbol{\theta}'}\right]=E\left[\frac{\partial\mathbf{h}_2(\mathbf{x}_t;\boldsymbol{\theta},\boldsymbol{\rho})}{\partial\boldsymbol{\theta}'}\right]-\mathbf{S}_{21}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{S}^{-}_{11}(\boldsymbol{\theta}^0)E\left[\frac{\partial\mathbf{h}_1(\mathbf{x}_t;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right],\\ \mathbf{D}_{2\boldsymbol{\rho}}(\boldsymbol{\theta},\boldsymbol{\rho}) &= E\left[\frac{\partial\mathbf{g}_2(\mathbf{x}_t;\boldsymbol{\theta},\boldsymbol{\rho})}{\partial\boldsymbol{\rho}'}\right]=E\left[\frac{\partial\mathbf{h}_2(\mathbf{x}_t;\boldsymbol{\theta},\boldsymbol{\rho})}{\partial\boldsymbol{\rho}'}\right],\end{aligned}$$

$$\begin{pmatrix}\boldsymbol{\theta}_\oplus\\ \boldsymbol{\theta}_\ominus\\ \boldsymbol{\rho}_\oplus\\ \boldsymbol{\rho}_\ominus\end{pmatrix}=\begin{bmatrix}\mathbf{m}_{1\oplus}(\boldsymbol{\theta})\\ \mathbf{m}_{1\ominus}(\boldsymbol{\theta})\\ \mathbf{m}_{2\oplus}(\boldsymbol{\theta},\boldsymbol{\rho})\\ \mathbf{m}_{2\ominus}(\boldsymbol{\theta},\boldsymbol{\rho})\end{bmatrix}=\begin{bmatrix}\mathbf{m}_1(\boldsymbol{\theta})\\ \mathbf{m}_2(\boldsymbol{\theta},\boldsymbol{\rho})\end{bmatrix}=\mathbf{m}(\boldsymbol{\theta},\boldsymbol{\rho}),$$

$$\begin{aligned}\mathbf{L}_{1\boldsymbol{\theta}_\oplus}(\boldsymbol{\theta}_\oplus,\boldsymbol{\theta}_\ominus) &= \frac{\partial\mathbf{l}_1(\boldsymbol{\theta}_\oplus,\boldsymbol{\theta}_\ominus)}{\partial\boldsymbol{\theta}'_\oplus},\\ \mathbf{L}_{2\boldsymbol{\rho}_\oplus}(\boldsymbol{\theta}_\oplus,\boldsymbol{\theta}_\ominus,\boldsymbol{\rho}_\oplus,\boldsymbol{\rho}_\ominus) &= \frac{\partial\mathbf{l}_2(\boldsymbol{\theta}_\oplus,\boldsymbol{\theta}_\ominus,\boldsymbol{\rho}_\oplus,\boldsymbol{\rho}_\ominus)}{\partial\boldsymbol{\rho}'_\oplus},\end{aligned}$$

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

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

Proof. : We know from Lemma 1 that $E[\mathbf{g}_2(\mathbf{x}_t;\boldsymbol{\theta},\boldsymbol{\rho})]=\mathbf{0}$ 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 3.2 implies that the right way to exploit the potential singularities in both sets of moment conditions is to estimate the parameters $\boldsymbol{\theta}_\oplus$ and $\boldsymbol{\rho}_\oplus$ from the transformed moment conditions:

$$\begin{aligned}\mathbf{h}_{1\oplus}(\mathbf{x}_t;\boldsymbol{\theta}_\oplus,\mathbf{0}|\boldsymbol{\theta}^0) &= \mathbf{P}'_{1\oplus}(\boldsymbol{\theta}^0)\mathbf{h}_1[\mathbf{x}_t;\mathbf{l}_1(\boldsymbol{\theta}_\oplus,\mathbf{0})],\\ \mathbf{g}_{2\oplus}(\mathbf{x}_t;\boldsymbol{\theta}_\oplus,\mathbf{0},\boldsymbol{\rho}_\oplus,\mathbf{0}|\boldsymbol{\theta}^0,\boldsymbol{\rho}^0) &= \mathbf{P}'_{2\oplus}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{h}_2[\mathbf{x}_t;\mathbf{l}_1(\boldsymbol{\theta}_\oplus,\mathbf{0}),\mathbf{l}_2(\boldsymbol{\theta}_\oplus,\mathbf{0},\boldsymbol{\rho}_\oplus,\mathbf{0})],\end{aligned}$$

where $\mathbf{P}_{1\oplus}(\boldsymbol{\theta}^0)\boldsymbol{\Delta}_{1\oplus}(\boldsymbol{\theta}^0)\mathbf{P}'_{1\oplus}(\boldsymbol{\theta}^0)$ provides the spectral decomposition of $\mathbf{S}_{11}(\boldsymbol{\theta}^0)$.

Let $\hat{\boldsymbol{\theta}}_{\oplus T}$ and $\hat{\boldsymbol{\rho}}_{\oplus T}$ denote the optimal GMM estimators of $\boldsymbol{\theta}_{\oplus}$ and $\boldsymbol{\rho}_{\oplus}$ based on both subsets of moment conditions. Similarly, let $\bar{\boldsymbol{\theta}}_{\oplus T}$ 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:

$$\lim_{T\to\infty} V\begin{bmatrix} \sqrt{T}(\hat{\boldsymbol{\theta}}_{\oplus T}-\boldsymbol{\theta}^0_{\oplus}) \\ \sqrt{T}(\hat{\boldsymbol{\rho}}_{\oplus T}-\boldsymbol{\rho}^0_{\oplus}) \end{bmatrix} = \left\{ \begin{bmatrix} \mathbf{D}'_{1\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus}) & \mathbf{D}'_{2\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus}) \\ \mathbf{0} & \mathbf{D}'_{2\oplus\boldsymbol{\rho}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus}) \end{bmatrix}\right.$$
$$\left.\times\begin{bmatrix} \boldsymbol{\Delta}^{-1}_{1\oplus}(\boldsymbol{\theta}^0) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}^{-1}_{2\oplus}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)) \end{bmatrix} \begin{bmatrix} \mathbf{D}_{1\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus}) & \mathbf{0} \\ \mathbf{D}_{2\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus}) & \mathbf{D}_{2\oplus\boldsymbol{\rho}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus}) \end{bmatrix}\right\}^{-1},$$

and

$$\lim_{T\to\infty} V[\sqrt{T}(\bar{\boldsymbol{\theta}}_{\oplus T}-\boldsymbol{\theta}^0_{\oplus})] = [\mathbf{D}'_{1\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus})\boldsymbol{\Delta}^{-1}_{1\oplus}(\boldsymbol{\theta}^0)\mathbf{D}_{1\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus})]^{-1}.$$

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

$$\{\mathbf{D}'_{1\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus})\boldsymbol{\Delta}^{-1}_{1\oplus}(\boldsymbol{\theta}^0)\mathbf{D}_{1\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus}) + \mathbf{D}'_{2\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus})\boldsymbol{\Delta}^{-1}_{2\oplus}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{D}_{2\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus})$$
$$-\mathbf{D}'_{2\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus})\boldsymbol{\Delta}^{-1}_{2\oplus}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{D}_{2\oplus\boldsymbol{\rho}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus})[\mathbf{D}'_{2\oplus\boldsymbol{\rho}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus})\boldsymbol{\Delta}^{-1}_{2\oplus}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{D}_{2\oplus\boldsymbol{\rho}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus})]^{-1}$$
$$\times\mathbf{D}'_{2\oplus\boldsymbol{\rho}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus})\boldsymbol{\Delta}^{-1}_{2\oplus}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{D}_{2\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus})\}^{-1}.$$

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

$$\mathbf{D}'_{2\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus})\boldsymbol{\Delta}^{-1}_{2\oplus}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{D}_{2\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus}) - \mathbf{D}'_{2\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus})\boldsymbol{\Delta}^{-1}_{2\oplus}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{D}_{2\oplus\boldsymbol{\rho}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus})$$
$$\times[\mathbf{D}'_{2\oplus\boldsymbol{\rho}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus})\boldsymbol{\Delta}^{-1}_{2\oplus}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{D}_{2\oplus\boldsymbol{\rho}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus})]^{-1}\mathbf{D}'_{2\oplus\boldsymbol{\rho}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus})\boldsymbol{\Delta}^{-1}_{2\oplus}(\boldsymbol{\theta}^0,\boldsymbol{\rho}^0)\mathbf{D}_{2\oplus\boldsymbol{\theta}_{\oplus}}(\boldsymbol{\theta}^0_{\oplus},\boldsymbol{\rho}^0_{\oplus})$$

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

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 $\mathbf{R}_t = \boldsymbol{\nu}^0 + (\boldsymbol{\Sigma}^0)^{1/2}\boldsymbol{\varepsilon}^{\circ}_t$, where $\boldsymbol{\varepsilon}^{\circ}_t$ 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 $E(\boldsymbol{\varepsilon}_t^\circ) = \mathbf{0}$, $V(\boldsymbol{\varepsilon}_t^\circ) = \mathbf{I}_N$, $E(\mathbf{R}_t) = \boldsymbol{\nu}$ and $V(\mathbf{R}_t) = \boldsymbol{\Sigma}$. For instance, if $e_t = \sqrt{(\upsilon^0 - 2)\zeta_t/\xi_t}$, ζ_t is a chi-square random variable with N degrees of freedom, and ξ_t is an independent Gamma variate with mean $\upsilon^0 > 2$ and variance $2\upsilon^0$, then $\boldsymbol{\varepsilon}_t^\circ$ will be distributed as a standardised multivariate Student t random vector of dimension N with υ^0 degrees of freedom, which converges to a standardised multivariate normal as $\upsilon^0 \to \infty$. 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, $\kappa = 2/(\upsilon^0 - 4)$ 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-Schwarzt 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

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

and

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

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 $\varrho(.)$ is a scalar function whose form depends on the member of the elliptical class (see again Fang, Kotz and Ng (1990)). For instance, $\varrho(.)$ 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:

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 $\sqrt{T}\bar{\mathbf{h}}_{UT}(\boldsymbol{\phi}^0)$, where $\mathbf{h}_U(\mathbf{R}_t;\boldsymbol{\phi})$ is defined in (5), will be $\mathbf{S}_U(\boldsymbol{\phi}^0)$, with:*

$$\mathbf{S}_U(\boldsymbol{\phi}) = \begin{Bmatrix} (1+\mathrm{B})^{-1}\left[1+\kappa\mathrm{B}(1+\mathrm{B})^{-1}\right] & \mathrm{A}(1+\mathrm{B})^{-2}\kappa \\ \mathrm{A}(1+\mathrm{B})^{-1}\kappa & [\mathrm{C}-\mathrm{A}^2(1+\mathrm{B})^{-1}]+\kappa\left([\mathrm{C}-\mathrm{A}^2(1+\mathrm{B})^{-1}]-\mathrm{A}^2(1+\mathrm{B})^{-2}\right) \end{Bmatrix}\otimes\boldsymbol{\Gamma}$$
$$+\begin{Bmatrix} -2(1+\mathrm{B})^{-2}+\left(3\mathrm{B}^2(1+\mathrm{B})^{-2}-5\mathrm{B}(1+\mathrm{B})^{-1}+2\right)\kappa & \mathrm{A}(1+\mathrm{B})^{-2}(2-3\kappa) \\ \mathrm{A}(1+\mathrm{B})^{-2}(2-3\kappa) & -2\mathrm{A}^2(1+\mathrm{B})^{-2}+\{3\mathrm{A}^2(1+\mathrm{B})^{-2}-[\mathrm{C}-\mathrm{A}^2(1+\mathrm{B})^{-1}]\}\kappa \end{Bmatrix}\otimes\boldsymbol{\nu}\boldsymbol{\nu}'$$
$$+\begin{pmatrix} 0 & 0 \\ 0 & 1+2\kappa \end{pmatrix}\otimes\ell_N\ell_N' + \begin{bmatrix} 0 & (1+\mathrm{B})^{-1}\kappa \\ (1+\mathrm{B})^{-1}\kappa & -2\mathrm{A}(1+\mathrm{B})^{-1}\kappa \end{bmatrix}\otimes(\boldsymbol{\nu}\ell_N'+\ell_N\boldsymbol{\nu}').$$

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

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 $\sqrt{T}\bar{\mathbf{h}}_{ET}(\boldsymbol{\varphi}^0,\boldsymbol{\nu}^0)$, where $\mathbf{h}_E(\mathbf{R}_t;\boldsymbol{\varphi},\boldsymbol{\nu})$ is defined in (7), will be $\mathbf{S}_E(\boldsymbol{\varphi}^0,\boldsymbol{\nu}^0)$, with:*

$$\mathbf{S}_E(\boldsymbol{\varphi},\boldsymbol{\nu}) = \begin{bmatrix} \mathbf{I} & -\mathbf{I} & \mathbf{0} \\ -\mathbf{I} & 1+(\kappa+1)\mathrm{B} & (\kappa+1)\mathrm{A} \\ \mathbf{0} & (\kappa+1)\mathrm{A} & (\kappa+1)\mathrm{C} \end{bmatrix}\otimes\boldsymbol{\Sigma}$$
$$+\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (2\kappa+1)\boldsymbol{\nu}\boldsymbol{\nu}' & (\kappa+1)\ell_N\boldsymbol{\nu}'+\kappa\boldsymbol{\nu}\ell_N' \\ \mathbf{0} & (\kappa+1)\boldsymbol{\nu}\ell_N'+\kappa\ell_N\boldsymbol{\nu}' & (2\kappa+1)\ell_N\ell_N' \end{pmatrix},$$

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

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 $\sqrt{T}\bar{\mathbf{h}}_{HT}(\mathbf{a}^0,\mathbf{b}^0)$, where $\mathbf{h}_H(\mathbf{R}_t;\mathbf{a},\mathbf{b})$ is defined in (11), will be $\mathbf{S}_H(\mathbf{a}^0,\mathbf{b}^0)$, with:*

$$\mathbf{S}_H(\mathbf{a},\mathbf{b}) = \begin{pmatrix} 1 & \boldsymbol{\nu}_1' \\ \boldsymbol{\nu}_1 & (\kappa+1)\boldsymbol{\Sigma}_{11}+\boldsymbol{\nu}_1\boldsymbol{\nu}_1' \end{pmatrix}\otimes\boldsymbol{\Omega}.$$

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

$$\mathbf{S}_H(\mathbf{a},\mathbf{b}) = E\{E[\mathbf{h}_H(\mathbf{R}_t;\mathbf{a},\mathbf{b})|\mathbf{R}_{1t}]\}$$
$$=\begin{bmatrix} E\{\varrho[(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)]\} & E\{\varrho[(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)]\mathbf{R}_{1t}'\} \\ E\{\varrho[(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)]\mathbf{R}_{1t}\} & E\{o\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)]\mathbf{R}_{1t}\mathbf{R}_{1t}'\} \end{bmatrix}\otimes\boldsymbol{\Omega}.$$

But since

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

then it must be the case that $E\{\varrho[(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)]\} = 1$. Similarly, since

$$\begin{gathered} E\left\{(\mathbf{R}_{1t}\text{-}\boldsymbol{\nu}_1)\cdot vec'\left[(\mathbf{R}_{2t}\text{-}\boldsymbol{\nu}_2)(\mathbf{R}_{2t}\text{-}\boldsymbol{\nu}_2)'\right]\right\} = E\left[(\mathbf{R}_{1t}\text{-}\boldsymbol{\nu}_1)\cdot E\left\{vec'\left[(\mathbf{R}_{2t}\text{-}\boldsymbol{\nu}_2)(\mathbf{R}_{2t}\text{-}\boldsymbol{\nu}_2)'\right]|\mathbf{R}_{1t}\right\}\right] \\ = E[(\mathbf{R}_{1t}\text{-}\boldsymbol{\nu}_1)\cdot vec\{\varrho[(\mathbf{R}_{1t}\text{-}\boldsymbol{\nu}_1)'\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_{1t}\text{-}\boldsymbol{\nu}_1)]\cdot\boldsymbol{\Omega}+\mathbf{B}(\mathbf{R}_{1t}\text{-}\boldsymbol{\nu}_1)(\mathbf{R}_{1t}\text{-}\boldsymbol{\nu}_1)'\mathbf{B}'\}] = \mathbf{0} \end{gathered}$$

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

$$E\{\varrho[(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)](\mathbf{R}_{1t}-\boldsymbol{\nu}_1)\} = \mathbf{0},$$

and consequently, that $E[\varrho[(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)]\mathbf{R}_{1t}\} = \boldsymbol{\nu}_1$.

Finally, since

$$\begin{gathered} E\left[(\mathbf{R}_{2t}-\boldsymbol{\nu}_2)(\mathbf{R}_{2t}-\boldsymbol{\nu}_2)'\otimes(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\right] \\ = (\kappa+1)\left[(\boldsymbol{\Sigma}_{22}\otimes\boldsymbol{\Sigma}_{11}) + (\boldsymbol{\Sigma}_{21}\otimes\boldsymbol{\Sigma}_{12})\mathbf{K}_{N_1,N_2} + vec(\boldsymbol{\Sigma}_{12})\,vec(\boldsymbol{\Sigma}_{12})'\right] \\ = E\left\{E\left[(\mathbf{R}_{2t}-\boldsymbol{\nu}_2)(\mathbf{R}_{2t}-\boldsymbol{\nu}_2)'|\mathbf{R}_{1t}\right]\otimes(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\right\} \\ = E[\{\varrho[(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)]\cdot\boldsymbol{\Omega} \\ +\mathbf{B}(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\mathbf{B}'\}\otimes(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'] \\ = \boldsymbol{\Omega}\otimes E\{\varrho[(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)]\cdot(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\} \\ +(\mathbf{B}\otimes\mathbf{I}_{N_1})\,E\left[(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\otimes(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\right](\mathbf{B}'\otimes\mathbf{I}_{N_1}) \\ = \boldsymbol{\Omega}\otimes E\{\varrho[(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)]\cdot(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\} \\ +(\kappa+1)\left[(\mathbf{B}\boldsymbol{\Sigma}_{12}\otimes\boldsymbol{\Sigma}_{11}) + (\boldsymbol{\Sigma}_{21}\otimes\boldsymbol{\Sigma}_{12})\mathbf{K}_{N_1,N_2} + vec(\boldsymbol{\Sigma}_{12})\,vec(\boldsymbol{\Sigma}_{12})'\right], \end{gathered}$$

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 $E\{\varrho[(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)]\cdot(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\} = (\kappa+1)\boldsymbol{\Sigma}_{11}$, and consequently, that $E\{\varrho[(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)'\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{R}_{1t}-\boldsymbol{\nu}_1)]\mathbf{R}_{1t}\mathbf{R}_{1t}'\} = (\kappa+1)\boldsymbol{\Sigma}_{11}+\boldsymbol{\nu}_1\boldsymbol{\nu}_1'$. $\square$

References

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.

Berndt, E.R. and N.E. Savin. (1977): "Conflict among criteria for testing hypotheses in the multivariate linear regression model", *Econometrica* 45, 1263-1278.

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): "Funds, factors, and diversification in arbitrage pricing models", *Econometrica* 51, 1305-1323.

Chamberlain, G. and M. Rothschild (1983): "Arbitrage, factor structure, and mean-variance analysis on large asset markets", *Econometrica* 51, 1281-1304.

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.

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.

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.

Geweke, J. (1981): "The approximate slopes of econometric tests", *Econometrica* 49, 1427-1442.

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. and Monfort, A. (1995): *Statistics and Econometric models*, vols. I and II, Cambridge.

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 (2001): "The econometrics of efficient portfolios", mimeo, CREST-INSEE.

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.

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

Ogaki, M. (1993): “Generalized method of moments: econometric applications”, in Maddala, G.S. , C.R. Rao and H.D. Vinod, *Handbook of Statistics* vol. 11, 455-488, Elsevier.

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.

Rao, C.R. and S.K. Mitra (1971): *Generalized Inverse of Matrices and its Applications*, Wiley.

Sentana, E. (2004): “Least squares predictions and mean-variance analysis”, mimeo, CEMFI.

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 1a

Spanning Tests

All countries

Gross returns: 1 week and 3 months

p-values of the Uncentred Representing Portfolio (URP),

Centred Representing Portfolio (CRP), and Regression (Reg) versions

Weighting Matrix	Lags	URP	CRP	Reg
S	0	0.000	0.000	0.000
	8	0.000	0.000	0.000
S_0	0	0.000	0.000	0.000
	8	0.000	0.000	0.001

Table 1b

Spanning Tests

Belgium (BE), France (FR) and Italy (IT)

Gross returns: 1 week and 3 months

p-values of the Uncentred Representing Portfolio (URP) version

Weighting Matrix	Lags	BE	FR	IT
S	0	0.000	0.399	0.000
	8	0.000	0.460	0.000
S_0	0	0.000	0.482	0.000
	8	0.000	0.544	0.000

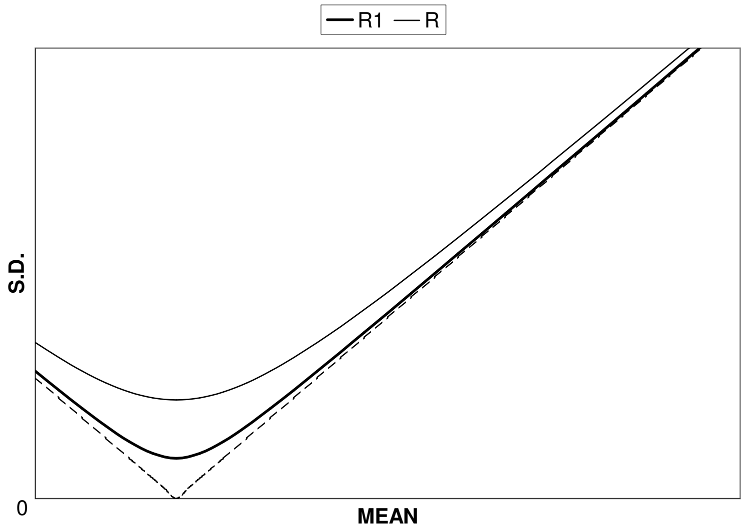

Figure 1a: Stochastic discount factor mean-variance frontiers that share the mean representing portfolio

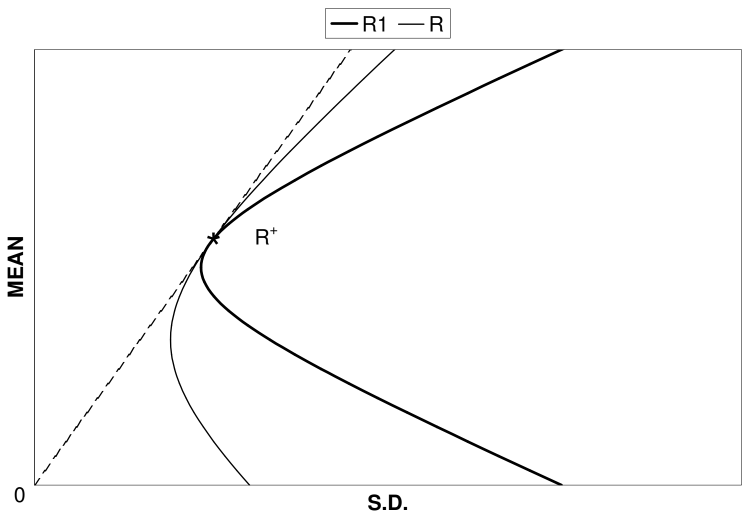

Figure 1b: Return mean-variance frontiers that share the mean representing portfolio

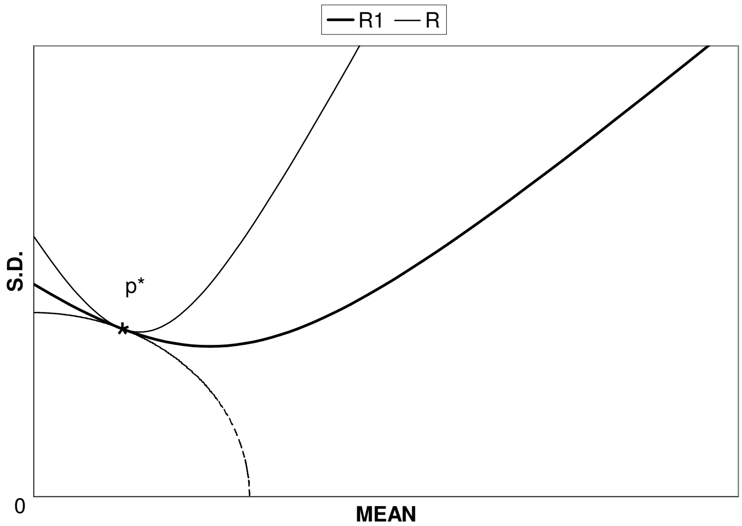

Figure 2a: Stochastic discount factor mean-variance frontiers that share the uncentred cost representing portfolio

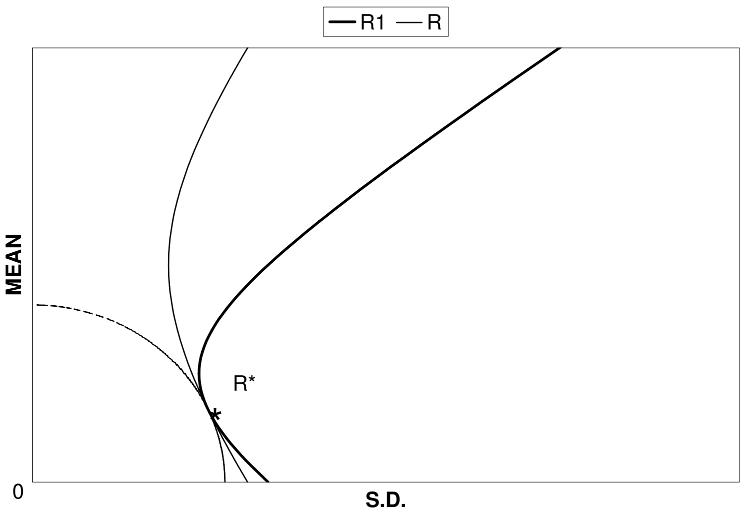

Figure 2b: Return mean-variance frontiers that share the uncentred cost representing portfolio

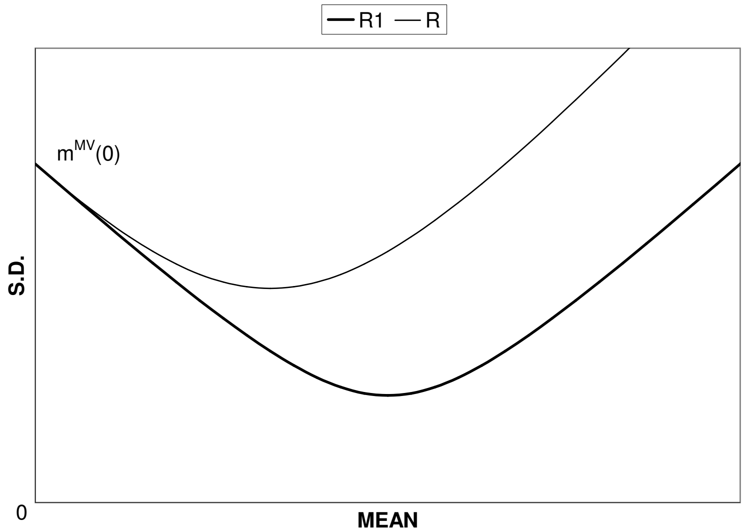

Figure 3a: Stochastic discount factor mean-variance frontiers that share the centred cost representing portfolio

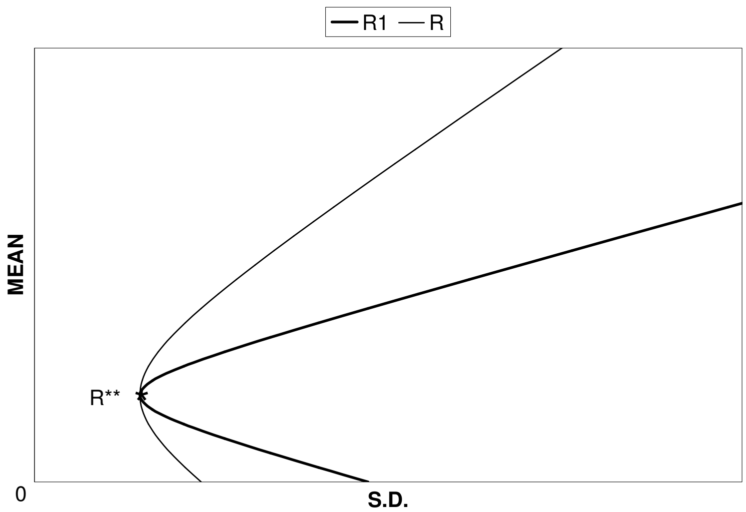

Figure 3b: Return mean-variance frontiers that share the centred cost representing portfolio