Supplemental Appendix for

A Unifying Approach to the Empirical Evaluation of Asset Pricing Models

Francisco Peñaranda

SanFI, Paseo Menéndez Pelayo 94-96, E-39006 Santander, Spain

<fpenaranda@fundacion-uceif.org>

Enrique Sentana

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

<sentana@cemfi.es>

First version: July 2010

Revised: May 2014

C Another problematic case: An orthogonal factor

Theoretical discussion

Lemma G3 in appendix G studies the implications of $E(\mathbf{r}f) = \mathbf{0}$ when $E(\mathbf{r}) \neq \mathbf{0}$. In this case, the centred SDF moment conditions (2) asymmetrically normalized in terms of τ will have a full rank Jacobian at the true value $\tau = -1/E(f)$. The same applies to the centred regression moment conditions (3) asymmetrically normalized in the same way, in which case the true value of the price of risk will be $\lambda = -V(f)/E(f)$. Similarly, the asymmetric normalizations $(a/b, 1)$ in (1) and $(c/d, 1)$ in (3) will also be well-behaved.

In contrast, (1) asymmetrically normalized in terms of δ cannot be set to 0. In fact, the multistep GMM criterion function does not depend on δ at all, which is not surprising given that the model is linear in this parameter and the expected Jacobian of those moment conditions is precisely $-E(\mathbf{r}f)$. As a result, the distribution of the associated J test will be non-standard. Nevertheless, we show in lemma G3 that the CU criterion converges to 0 as $\delta^{-1} \to 0$, which simply reflects the fact that a must go to 0 for the moment conditions (1) to hold.

Graphically, the differential behavior of the asymmetric normalizations $(1, b/a)$ and $(1, b/c)$ can be understood as follows. The point (c, b) chosen in Figure 2b is pinned down by the intersection between the straight lines given by the pricing condition in (2) and $(1, b/c)$, which remains well defined even if the pricing factor is orthogonal to the excess returns. However, as $E(\mathbf{r}f) \to \mathbf{0}$ the moments (1) define a flatter and flatter straight line in space (a, b), whose intersection with $(1, b/a)$ happens at a higher and higher b. In the limit, the first line becomes $a = 0$, which is parallel to the normalization (see Figure 3b).

Like in the case of an uncorrelated factor, a SDF that is exactly proportional to an orthogonal factor is not very attractive from an economic point of view. This is confirmed by the fact that the uncentred mimicking portfolio r^+ is 0. Given that the J tests of the asset pricing conditions that do not impose the problematic asymmetric normalization $(1, b/a)$ will not reject their null, we propose another simple additional test to detect this special case.

The null hypothesis $E(\mathbf{r}f) = \mathbf{0}$ is equivalent to all valid SDFs affine in f having a 0 intercept. In the case of the centred SDF moment conditions (2), this restriction can be assessed by means of a DM test of the additional moment condition

$$E(c - bf) = 0$$

expressed in such a way that it is compatible with the asymmetric or symmetric normalization used. Intuitively, this additional moment condition defines the intercept of the SDF, which we then set to 0 under the null. As expected, this DM test will also follow an asymptotic χ_1^2 distribution under the null of $E(\mathbf{r}f) = \mathbf{0}$.

Monte Carlo

We can make the pricing factor orthogonal to the vector of excess returns by merely changing the mean of f in the baseline design, while we leave the rest of DGP parameters as in section 6.1.

As expected from the preceding theoretical discussion, Table C1 shows high rejection rates for the multistep implementations of the uncentred SDF moment conditions (1) asymmetrically normalized with $(1, b/a)$. This pattern is even stronger in Table G6 when $T = 500$. The rejection rates of the multistep implementations of the asymmetric centred SDF are also high, but they decrease for $T = 500$, thereby resembling their behavior in the baseline design.

(TABLE C1)

Table C1 also reports the DM tests of the null hypotheses of an uncorrelated factor and an orthogonal one. In this design, the first hypothesis is false while the second one is true. We find overrejection for the orthogonal factor test for $T = 50$, but the rejection rates get very close to the nominal size when $T = 500$. In turn, the rejection rates for the test of an uncorrelated factor are noticeably higher for both sample sizes.

The bicorne plots for the prices of risk in Figures C1 and G4 show that the biggest differences across GMM implementations correspond to the estimators of δ. Once again, the sampling distribution of the CU estimator reflects much better than the multistep estimators the lack of a finite true parameter value.

(FIGURE C1)

In contrast, the three GMM implementations behave similarly for υ and λ. Moreover, the CU estimator of ψ and τ does not show the bias of iterated and two-step estimators, although it leads to a higher dispersion in the case of τ.

D Monte Carlo Design

An unrestricted Gaussian data generating process (DGP) for $(f, \mathbf{r})$ is

$$f \sim N\left(\mu, \sigma^2\right),$$
$$\mathbf{r} = \boldsymbol{\mu}_r + \boldsymbol{\beta}_r\left(f - \mu\right) + \mathbf{u}_r, \qquad \mathbf{u}_r \sim N\left(\mathbf{0}, \boldsymbol{\Omega}_{rr}\right).$$

However, given that we use the simulated data to test that an affine function of f is orthogonal to $\mathbf{r}$, the only thing that matters is the linear span of $\mathbf{r}$. As a result, we can substantially reduce the number of parameters characterizing the conditional DGP for $\mathbf{r}$ by means of the following steps:

1. a Cholesky transformation $\mathbf{C}^{-1}\mathbf{r}$ to get a residual variance equal to the identity matrix,
2. a Householder transformation $\mathbf{Q}_1\mathbf{C}^{-1}\mathbf{r}$ that makes the second to the last entries of the vector of risk premia $\boldsymbol{\mu}_r$ equal to zero (see Householder (1964)), and
3. another Householder transformation $\mathbf{Q}_2\mathbf{Q}_1\mathbf{C}^{-1}\mathbf{r}$ that makes the third to the last entries of the vector of betas equal to zero.

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

$$\dot{\mathbf{r}} = \dot{\mu}_1\mathbf{e}_1 + \left(\dot{\beta}_1\mathbf{e}_1 + \dot{\beta}_2\mathbf{e}_2\right)\left(\dot{f} - \dot{\mu}\right) + \dot{\mathbf{u}}_r, \qquad \dot{\mathbf{u}}_r \sim N\left(\mathbf{0}, \mathbf{I}\right),$$

where $\dot{\mathbf{r}} = \mathbf{Q}_2\mathbf{Q}_1\mathbf{C}^{-1}\mathbf{r}$, $\dot{\mu}_1 = \left(\boldsymbol{\mu}_r'\boldsymbol{\Omega}_{rr}^{-1}\boldsymbol{\mu}_r\right)^{1/2}$, $(\mathbf{e}_1, \mathbf{e}_2)$ are the first and second columns of the identity matrix, $\dot{\beta}_1$ is the first entry of $\mathbf{Q}_1\mathbf{C}^{-1}\boldsymbol{\beta}_r\sigma$, $\dot{\beta}_2$ is the norm of the remaining entries of $\mathbf{Q}_1\mathbf{C}^{-1}\boldsymbol{\beta}_r\sigma$, and

$$\dot{f} \sim N\left(\dot{\mu}, 1\right), \qquad \dot{\mu} = \mu/\sigma.$$

The parameter $\dot{\mu}$ can be directly calibrated from data on f.

In turn, $\dot{\beta}_2$ can be calibrated from a Hansen-Jagannathan (HJ) distance. Specifically, let $y = c + b(\dot{f} - \dot{\mu})$ denote a potentially invalid SDF based on $\dot{f}$. Given that the scale of y does not matter for pricing excess returns and we are ruling out $E(\mathbf{r}) = \mathbf{0}$, we can simply normalize it as

$$\dot{y} = \dot{c} + (\dot{f} - \dot{\mu}), \qquad \dot{c} = c/b,$$

and the potential pricing errors are

$$E\left(\dot{y}\dot{\mathbf{r}}\right) = \left(\dot{c}\dot{\mu}_1 + \dot{\beta}_1\right)\mathbf{e}_1 + \dot{\beta}_2\mathbf{e}_2.$$

The null hypothesis of a valid SDF is equivalent to the existence of some $\dot{c}$ such that all pricing errors are zero. It is then easy to see that the null hypothesis is equivalent to $\dot{\beta}_2 = 0$, in which case the valid SDF satisfies $\dot{c}\dot{\mu}_1 + \dot{\beta}_1 = 0$.

On the other hand, if $\dot{\beta}_2 \neq 0$ then all the pricing errors cannot be equal to zero. But we can still choose some invalid SDF that minimizes pricing errors under some metric in order to select $\dot{c}$. In particular, we can minimize with respect to $\dot{c}$ the following criterion

$$E\left(\dot{y}\mathbf{r}\right)' Var^{-1}\left(\mathbf{r}\right) E\left(\dot{y}\mathbf{r}\right) = E\left(\dot{y}\dot{\mathbf{r}}\right)' Var^{-1}\left(\dot{\mathbf{r}}\right) E\left(\dot{y}\dot{\mathbf{r}}\right),$$

which we can interpret as a HJ distance. The corresponding invalid SDF satisfies $\dot{c}\dot{\mu}_1(1+\dot{\beta}_2^2) = -\dot{\beta}_1$ and the value of the HJ distance is

$$\frac{\dot{\beta}_2^2}{1+\dot{\beta}_2^2}$$

for the normalization $(c/b, 1)$. Therefore, we can calibrate $\dot{\beta}_2$ from a given value of this distance. Interestingly, we get the exactly same value of $\dot{c}$ and HJ distance if we use the criterion $E\left(\dot{y}\dot{\mathbf{r}}\right)' E^{-1}\left(\dot{\mathbf{r}}\dot{\mathbf{r}}'\right) E\left(\dot{y}\dot{\mathbf{r}}\right)$ instead, which is closer to the original definition in Hansen and Jagannathan (1997).

Having chosen $\dot{\beta}_2$, the remaining two parameters $\left(\dot{\mu}_1, \dot{\beta}_1\right)$ can be calibrated by means of the maximum Sharpe ratio S of $\mathbf{r}$ (or $\dot{\mathbf{r}}$), and the R^2 of regressing f onto a constant and $\mathbf{r}$ (or equivalently $\dot{f}$ onto a constant and $\dot{\mathbf{r}}$), which are given by

$$S^2 = E\left(\dot{\mathbf{r}}\right)' Var^{-1}\left(\dot{\mathbf{r}}\right) E\left(\dot{\mathbf{r}}\right) = \frac{\dot{\mu}_1^2\left(1+\dot{\beta}_2^2\right)}{1+\dot{\beta}_1^2+\dot{\beta}_2^2},$$
$$R^2 = Cov(\dot{\mathbf{r}}, \dot{f})' Var^{-1}\left(\dot{\mathbf{r}}\right) Cov(\dot{\mathbf{r}}, \dot{f}) = \frac{\dot{\beta}_1^2+\dot{\beta}_2^2}{1+\dot{\beta}_1^2+\dot{\beta}_2^2}.$$

Given the desired $\left(S^2, R^2\right)$, the corresponding $\left(\dot{\mu}_1, \dot{\beta}_1\right)$ must satisfy

$$\dot{\mu}_1^2 = \frac{S^2}{(1-R^2)\left(1+\dot{\beta}_2^2\right)}, \qquad \dot{\beta}_1^2 + \dot{\beta}_2^2 = \frac{R^2}{1-R^2}.$$

Note that the signs of $\dot{\beta}_2$, $\dot{\beta}_1$ and $\dot{\mu}_1$ do not matter for our purposes. Without loss of generality, we choose the positive root of the equations above.

In this context, the special case of an uncorrelated factor will be given by $Cov(\dot{\mathbf{r}}, \dot{f}) = \mathbf{0}$, or $Cov\left(\mathbf{r}, f\right) = \mathbf{0}$, and it is equivalent to $R^2 = 0$. Hence the corresponding DGP satisfies $\dot{\beta}_1 = 0$ and $\dot{\beta}_2 = 0$. As for the orthogonal factor in appendix C, which is equivalent to $E(\dot{f}\dot{\mathbf{r}}) = \mathbf{0}$, we choose the DGP so that $\dot{\mu}\dot{\mu}_1 + \dot{\beta}_1 = 0$ and $\dot{\beta}_2 = 0$.

E Multifactor models

In what follows we represent a set of k factors by the vector $\mathbf{f}$, and their mean vector, second moment matrix and covariance matrix by $\boldsymbol{\mu}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\mu}\boldsymbol{\mu}'$, respectively. In this multifactor context, the connection between the SDF and regression approaches is given by

$$\begin{aligned} E(\mathbf{r})a + E(\mathbf{r}\mathbf{f}')\mathbf{b} &= E(\mathbf{r})(a+\mathbf{b}'\boldsymbol{\mu}) + Cov(\mathbf{r},\mathbf{f})\mathbf{b} = \boldsymbol{\phi}(a+\mathbf{b}'\boldsymbol{\mu}) + \mathbf{B}\boldsymbol{\mu}(a+\mathbf{b}'\boldsymbol{\mu}) + \mathbf{B}\boldsymbol{\Sigma}\mathbf{b} \\ &= \boldsymbol{\phi}(a+\mathbf{b}'\boldsymbol{\mu}) + \mathbf{B}\left(a\boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{b}\right) = \boldsymbol{\phi}c + \mathbf{B}\mathbf{d} = \mathbf{0}, \end{aligned} \tag{E9}$$

where c is the mean of the SDF, $\mathbf{d}$ the shadow costs of $\mathbf{f}$ (or actual costs if it is a vector of traded payoffs), $\boldsymbol{\phi}$ the vector of regression intercepts and

$$\mathbf{B} = \begin{pmatrix} \boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_k \end{pmatrix}$$

the $n \times k$ matrix of regression slopes.

The existence of a unique (up to scale) affine SDF $a + \mathbf{f}'\mathbf{b}$ that correctly prices the vector of excess returns at hand is equivalent to the $n \times (k+1)$ matrix with columns $E(\mathbf{r})$ and $E(\mathbf{r}\mathbf{f}')$ having rank k. Such a condition is related to the uncentred SDF approach. We can transfer this rank k condition to a matrix constructed with $E(\mathbf{r})$ and $Cov(\mathbf{r},\mathbf{f})$, which is related to the centred SDF approach, and another matrix built from $\boldsymbol{\phi}$ and $\mathbf{B}$ in the case of the centred regression.

Below we define several moment conditions and parameters for the SDF and regression approaches. We focus on the case of excess returns and traded or nontraded factors. Extensions to mixed factors and the addition of a gross return are straightforward.

Traded factors

We can evaluate the asset pricing model by means of the uncentred SDF influence functions

$$\mathbf{h}_U(\mathbf{r},\mathbf{f};a,\mathbf{b}) = \begin{bmatrix} \mathbf{r}(a+\mathbf{f}'\mathbf{b}) \\ \mathbf{f}(a+\mathbf{f}'\mathbf{b}) \end{bmatrix}, \tag{E10}$$

the centred SDF influence functions

$$\mathbf{h}_C(\mathbf{r},\mathbf{f};c,\mathbf{b},\boldsymbol{\mu}) = \begin{bmatrix} \mathbf{r}\left(c+(\mathbf{f}-\boldsymbol{\mu})'\mathbf{b}\right) \\ \mathbf{f}\left(c+(\mathbf{f}-\boldsymbol{\mu})'\mathbf{b}\right) \\ \mathbf{f}-\boldsymbol{\mu} \end{bmatrix}, \tag{E11}$$

or the regression influence functions

$$\mathbf{h}_R(\mathbf{r},\mathbf{f};\mathbf{B}) = \begin{bmatrix} \mathbf{r}-\mathbf{B}\mathbf{f} \\ vec\left((\mathbf{r}-\mathbf{B}\mathbf{f})\,\mathbf{f}'\right) \end{bmatrix}, \tag{E12}$$

where we have used (E9) and the fact that $\mathbf{d}=\mathbf{0}$ when the factors are excess returns.

The SDF functions require some normalization in their implementation. A symmetrically normalized version of the SDF approach would use the normalization $a^2+\mathbf{b}'\mathbf{b}=1$ or $c^2+\mathbf{b}'\mathbf{b}=1$, but asymmetric normalizations are more common in empirical work. The uncentred SDF method is usually combined with the asymmetric normalization $(1,\mathbf{b}/a)$ in (E10), implemented in terms of the parameterization $\boldsymbol{\delta}=-\mathbf{b}/a$, which gives rise to the influence functions

$$\begin{bmatrix} \mathbf{r}\,(1-\mathbf{f}'\boldsymbol{\delta}) \\ \mathbf{f}\,(1-\mathbf{f}'\boldsymbol{\delta}) \end{bmatrix},$$

while the influence functions of the centred SDF method that imposes the asymmetric normalization $(1,\mathbf{b}/c)$ in (E11) become

$$\begin{bmatrix} \mathbf{r}\left(1-(\mathbf{f}-\boldsymbol{\mu})'\boldsymbol{\tau}\right) \\ \mathbf{f}\left(1-(\mathbf{f}-\boldsymbol{\mu})'\boldsymbol{\tau}\right) \\ \mathbf{f}-\boldsymbol{\mu} \end{bmatrix}$$

with $\boldsymbol{\tau}=-\mathbf{b}/c$.

In all these methods the number of degrees of freedom of the corresponding J tests is n regardless of the number of factors k. The Jensen's alphas and pricing errors of excess returns $\mathbf{r}$ are defined by $E(\mathbf{r})-\mathbf{B}E(\mathbf{f})$, $E(\mathbf{r})-E\left(\mathbf{r}\mathbf{f}'\right)\delta$, and $E(\mathbf{r})-E\left[\mathbf{r}\,(\mathbf{f}-\boldsymbol{\mu})'\right]\boldsymbol{\tau}$, respectively.

Non-traded factors

Condition (E9) with a unique (up to scale) valid SDF $a+\mathbf{f}'\mathbf{b}$ is equivalent to both $\boldsymbol{\phi}$ and $\mathbf{B}$ belonging to the span of some $n\times k$ matrix that we can denote as

$$\mathbf{P} = \begin{pmatrix} \boldsymbol{\varphi}_1 & \cdots & \boldsymbol{\varphi}_k \end{pmatrix},$$

which has full column rank. Assuming that the number of assets exceeds the number of factors $(n>k)$ to ensure that the linear factor pricing model imposes testable restrictions on asset returns, we can impose this implicit constraint on the intercepts and slopes of the regression of $\mathbf{r}$ on a constant and $\mathbf{f}$ as follows:

$$\boldsymbol{\phi} = -\mathbf{P}\mathbf{d} = -\left(\boldsymbol{\varphi}_1 d_1 + ... + \boldsymbol{\varphi}_k d_k\right),$$
$$\mathbf{B} = \begin{pmatrix} \boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_k \end{pmatrix} = \begin{pmatrix} c\boldsymbol{\varphi}_1 & \cdots & c\boldsymbol{\varphi}_k \end{pmatrix} = c\mathbf{P}.$$

Therefore, we can evaluate the corresponding asset pricing model by means of the uncentred SDF influence functions

$$\mathbf{g}_S(\mathbf{r},\mathbf{f};a,\mathbf{b}) = \left[\mathbf{r}\left(a+\mathbf{f}'\mathbf{b}\right)\right], \tag{E13}$$

the centred SDF influence functions

$$\mathbf{g}_C(\mathbf{r},\mathbf{f};c,\mathbf{b},\boldsymbol{\mu}) = \begin{bmatrix} \mathbf{r}\left(c+(\mathbf{f}-\boldsymbol{\mu})'\mathbf{b}\right) \\ \mathbf{f}-\boldsymbol{\mu} \end{bmatrix} \tag{E14}$$

or the centred regression influence functions

$$\mathbf{g}_R(\mathbf{r},\mathbf{f};\mathbf{P},c,\mathbf{d}) = \begin{bmatrix} \mathbf{r}-\mathbf{P}(c\mathbf{f}-\mathbf{d}) \\ vec\left((\mathbf{r}-\mathbf{P}(c\mathbf{f}-\mathbf{d}))\,\mathbf{f}'\right) \end{bmatrix}. \tag{E15}$$

The SDF functions require some normalization in their implementation. A symmetrically normalized version of the SDF approach would use the normalization $a^2+\mathbf{b}'\mathbf{b}=1$ or $c^2+\mathbf{b}'\mathbf{b}=1$, while the centred regression would rely on the normalization $c^2+\mathbf{d}'\mathbf{d}=1$. On the other hand, asymmetric normalizations are more common in empirical work. Specifically, the uncentred SDF method is usually combined with the asymmetric normalization $(1,\mathbf{b}/a)$ implemented in terms of the parameterization $\boldsymbol{\delta}=-\mathbf{b}/a$, which gives rise to the influence functions

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

In turn, the influence functions of the centred SDF method that imposes the asymmetric normalization $(1,\mathbf{b}/c)$ in (E11) become

$$\begin{bmatrix} \mathbf{r}\left(1-(\mathbf{f}-\boldsymbol{\mu})'\boldsymbol{\tau}\right) \\ \mathbf{f}-\boldsymbol{\mu} \end{bmatrix},$$

with $\boldsymbol{\tau}=-\mathbf{b}/c$. Finally, the usual centred regression imposes $(1,\mathbf{d}/c)$ and relies on the influence functions

$$\begin{bmatrix} \mathbf{r}-\mathbf{B}(\mathbf{f}+\varkappa) \\ vec\left((\mathbf{r}-\mathbf{B}(\mathbf{f}+\varkappa))\,\mathbf{f}'\right) \end{bmatrix},$$

where the parameters are $(\varkappa,\mathbf{B})$, with $\varkappa=-\mathbf{d}/c$. Alternatively, we can define the vector $\boldsymbol{\lambda}$ of factor risk premia as $E(\mathbf{r})=\mathbf{B}\boldsymbol{\lambda}$, so that $\boldsymbol{\lambda}=\varkappa+\boldsymbol{\mu}$, and add the estimation of $\boldsymbol{\mu}$. This yields

$$\begin{bmatrix} \mathbf{r}-\mathbf{B}(\mathbf{f}-\boldsymbol{\mu}+\boldsymbol{\lambda}) \\ vec\left((\mathbf{r}-\mathbf{B}(\mathbf{f}-\boldsymbol{\mu}+\boldsymbol{\lambda}))\,\mathbf{f}'\right) \\ \mathbf{f}-\boldsymbol{\mu} \end{bmatrix}.$$

In any case, the degrees of freedom of the J test will be $n-k$. As for Jensen's alphas and pricing errors, they are defined by $E(\mathbf{r})-\mathbf{B}\boldsymbol{\lambda}$, $E(\mathbf{r})-E\left(\mathbf{r}\mathbf{f}'\right)\delta$, and $E(\mathbf{r})-E\left[\mathbf{r}(\mathbf{f}-\boldsymbol{\mu})'\right]\boldsymbol{\tau}$, respectively.

F Single-step methods: Continuously Updated GMM

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 a vector of known functions of $\mathbf{x}_t$, where $\boldsymbol{\theta}$ is a vector of unknown parameters. The true parameter value, $\boldsymbol{\theta}^0$, which we assume belongs to the interior of the compact set $\boldsymbol{\Theta} \subseteq \mathbb{R}^{\dim(\boldsymbol{\theta})}$, is implicitly defined by the (population) moment conditions:

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

where the expectation is taken with respect to the stationary distribution of $\mathbf{x}_t$. In our context of asset pricing models, $\mathbf{x}_t = (\mathbf{f}_t', \mathbf{r}_t')'$ represents data on excess returns and factors, and $\boldsymbol{\theta}$ represents the parameters of the specific model under evaluation.

GMM estimators minimize a specific norm $\bar{\mathbf{h}}_T'(\boldsymbol{\theta})\boldsymbol{\Upsilon}_T\bar{\mathbf{h}}_T(\boldsymbol{\theta})$ of the sample moments $\bar{\mathbf{h}}_T(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^T \mathbf{h}(\mathbf{x}_t;\boldsymbol{\theta})$ defined by some weighting matrix $\boldsymbol{\Upsilon}_T$. In overidentified cases such as ours, Hansen (1982) showed that if the long-run covariance matrix of the moment conditions $\mathbf{S}(\boldsymbol{\theta}^0) = avar[\sqrt{T}\bar{\mathbf{h}}_T(\boldsymbol{\theta}^0)]$ has full rank, then $\mathbf{S}^{-1}(\boldsymbol{\theta}^0)$ will be the "optimal" weighting matrix, in the sense that the difference between the asymptotic covariance matrix of the resulting GMM estimator and a GMM estimator based on any other norm of the same moment conditions is positive semidefinite. Therefore, the optimal GMM estimator of $\boldsymbol{\theta}$ will be

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

where

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

This optimal estimator is infeasible unless we know $\mathbf{S}(\boldsymbol{\theta}^0)$, but under additional regularity conditions, we can define an asymptotically equivalent but feasible two-step optimal GMM estimator by replacing $\mathbf{S}(\boldsymbol{\theta}^0)$ with an estimator $\mathbf{S}_T(\boldsymbol{\theta})$ evaluated at some initial consistent estimator of $\boldsymbol{\theta}^0$, $\dot{\boldsymbol{\theta}}_T$ say. There is an extensive literature on heteroskedasticity and autocorrelation consistent (HAC) estimators of long-run covariance matrices (see for example DeJong and Davidson (2000) and the references therein). In practice, we can repeat this two-step procedure many times to obtain iterated GMM estimators, although there is no guarantee that such a procedure will converge, and in fact it may cycle around several values instead.

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

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

where

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

Peñaranda and Sentana (2012) discuss how to express the CU-GMM criterion in terms of OLS output, which facilitates its optimization. Although this estimator is often more difficult to compute than two-step and iterated estimators, particularly in linear models, an important advantage is that it is numerically invariant to normalization, bijective reparameterizations and parameter-dependent linear transformations of the moment conditions, which will again prove useful in our context. In contrast, these properties do not necessarily hold for two-step or iterated GMM.

Newey and Smith (2004) highlight other important advantages of CU- over two-step GMM by going beyond the usual first-order asymptotic equivalence results. They also discuss alternative generalized empirical likelihood (GEL) estimators, such as empirical likelihood or exponentially-tilted methods. In fact, Antoine, Bonnal and Renault (2006) study the Euclidean empirical likelihood estimator, which is numerically equivalent to CU-GMM as far as $\boldsymbol{\theta}$ is concerned. Importantly, it is straightforward to show that these GEL methods share the numerical invariance properties of CU-GMM.

Our empirical application and simulation experiments will consider two-step, iterated and CU-GMM. Under standard regularity conditions (see Hansen (1982) and Newey and MacFadden (1994)), $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^0)$ and $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^0)$ will be asymptotically distributed up to first-order as the same normal random vector with zero mean and variance

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

where $\mathbf{D}(\boldsymbol{\theta}^0)$ denotes the probability limit of the Jacobian of $\bar{\mathbf{h}}_T(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta}^0$. In our empirical application, we replace $\mathbf{D}(\boldsymbol{\theta}^0)$ by $\partial\bar{\mathbf{h}}_T(\hat{\boldsymbol{\theta}}_T)/\partial\boldsymbol{\theta}'$ in the case of two-step and iterated GMM estimators. In contrast, for the CU-GMM estimator $\tilde{\boldsymbol{\theta}}_T$ we compute a consistent estimator of $\mathbf{D}(\boldsymbol{\theta}^0)$ that takes into account that the weighting matrix $\mathbf{S}_T^{-1}(\boldsymbol{\theta})$ is not fixed in the criterion function. Specifically, we estimate the asymptotic variance of $\tilde{\boldsymbol{\theta}}_T$ as

$$\left[\mathfrak{D}_T'(\tilde{\boldsymbol{\theta}}_T)\mathbf{S}_T^{-1}(\tilde{\boldsymbol{\theta}}_T)\mathfrak{D}_T(\tilde{\boldsymbol{\theta}}_T)\right]^{-1},$$

where

$$\mathfrak{D}_T(\tilde{\boldsymbol{\theta}}_T) = \frac{\partial \bar{\mathbf{h}}_T(\tilde{\boldsymbol{\theta}}_T)}{\partial \boldsymbol{\theta}'} - \frac{1}{2}\left[\bar{\mathbf{h}}_T'(\tilde{\boldsymbol{\theta}}_T)\mathbf{S}_T^{-1}(\tilde{\boldsymbol{\theta}}_T) \otimes \mathbf{I}_{\dim(\mathbf{h})}\right] \frac{\partial vec\left(\mathbf{S}_T(\tilde{\boldsymbol{\theta}}_T)\right)}{\partial \boldsymbol{\theta}'},$$

whose second term is nonzero in finite samples but asymptotically negligible.

Finally, $T \cdot J_T(\hat{\boldsymbol{\theta}}_T)$ and $T \cdot \tilde{J}_T(\tilde{\boldsymbol{\theta}}_T)$ will be asymptotically distributed as the same chi-square with $\dim(\mathbf{h}) - \dim(\boldsymbol{\theta})$ degrees of freedom if $E[\mathbf{h}(\mathbf{x};\boldsymbol{\theta})] = \mathbf{0}$ holds, so that we can use those statistics to compute overidentifying restrictions (J) tests. One could also use the GMM criterion to obtain alternative *t-ratios* by computing distance metric tests of the null hypotheses that every single element of $\boldsymbol{\theta}$ is 0 at a time.

G Additional Results

Lemma G1 *Assume that the asset pricing models holds for* $\mathbf{r}$ *with a traded factor* f *such that* $V(f) > 0$ *and* $E(\mathbf{r}) \neq \mathbf{0}$*. Then the Jacobians of all the following moment conditions have full column rank at the true parameter values:*
1) The uncentred SDF moment conditions that rely on a symmetric normalization.
2) The centred SDF moment conditions that rely on a symmetric normalization.
3) The uncentred SDF moment conditions that rely on the asymmetric normalization $(1, b/a)$*.*
4) The centred SDF moment conditions that rely on the asymmetric normalization $(1, b/c)$*.*
5) The regression moment conditions.

Proof.

1) When we work with (1), (5) and a symmetric normalization, we need to study the behavior of

$$E\left[\mathbf{x}\left(\sin\psi + f\cos\psi\right)\right],$$

where we have defined $\mathbf{x} = (f, \mathbf{r}')'$ to simplify the expressions.

In this case, the true value of the parameter is defined by

$$\begin{pmatrix} E(\mathbf{x}) & E(\mathbf{x}f) \end{pmatrix} \begin{pmatrix} \sin\psi \\ \cos\psi \end{pmatrix} = \mathbf{0},$$

while the Jacobian of the moment conditions with respect to ψ is

$$\begin{pmatrix} E(\mathbf{x}) & E(\mathbf{x}f) \end{pmatrix} \begin{pmatrix} \cos\psi \\ -\sin\psi \end{pmatrix}.$$

These two linear combinations are orthogonal at the same ψ. Therefore, the Jacobian has full column rank at the true value.

2) When we work with (2), (7) and a symmetric normalization, we need to study

$$E\begin{bmatrix} \mathbf{x}\left(\sin\upsilon + (f-\mu)\cos\upsilon\right) \\ f-\mu \end{bmatrix}.$$

Here, the true values of the parameters are $\mu = E(f)$ and υ given by

$$\begin{pmatrix} E(\mathbf{x}) & Cov(\mathbf{x},f) \end{pmatrix} \begin{pmatrix} \sin\upsilon \\ \cos\upsilon \end{pmatrix} = \mathbf{0},$$

which make those moments equal to zero.

The Jacobian of the moment conditions with respect to (υ, μ) is

$$\begin{bmatrix} E(\mathbf{x})(\cos\upsilon + \mu\sin\upsilon) - E(\mathbf{x}f)\sin\upsilon & -E(\mathbf{x})\cos\upsilon \\ 0 & -1 \end{bmatrix}.$$

As $\mu = E(f)$ at the true values, and the linear combination that defines the true υ is orthogonal to the combination that defines the first entry of the Jacobian

$$\begin{pmatrix} E(\mathbf{x}) & Cov(\mathbf{x},f) \end{pmatrix} \begin{pmatrix} \cos\upsilon \\ -\sin\upsilon \end{pmatrix},$$

the Jacobian must have full column rank because its first column cannot be zero.

3) When we work with (1), (5) and the asymmetric normalization $(1, b/a)$, we need to study

$$E[\mathbf{x}(1-\delta f)].$$

The Jacobian of the moment conditions with respect to δ is $-E(\mathbf{x}f)$, which has full column rank because $E(\mathbf{x}f) \neq \mathbf{0}$ since its last entry is $E(f^2)$.

4) When we work with (2), (7) and the asymmetric normalization $(1, b/c)$, we need to study

$$E\begin{bmatrix} \mathbf{x}(1-\tau(f-\mu)) \\ f-\mu \end{bmatrix}$$

Note that we can make the last moment condition equal to zero with $\mu = E(f)$, while simultaneously satisfying the pricing conditions with $\tau = -\cot\upsilon$, as we saw in the proof of point 2).

The Jacobian of the moment conditions with respect to (τ, μ) is

$$\begin{pmatrix} E(\mathbf{x})\mu - E(\mathbf{x}f) & E(\mathbf{x})\tau \\ 0 & -1 \end{pmatrix},$$

which has full column rank if and only if

$$E(\mathbf{x})\mu - E(\mathbf{x}f) \neq \mathbf{0}.$$

At the true values, this condition becomes

$$Cov(\mathbf{x}, f) \neq \mathbf{0},$$

which must hold since its last entry is $V(f)$.

5) When we work with (9), we need to study the behavior of

$$E\left[\begin{pmatrix} 1 \\ f \end{pmatrix} \otimes (\mathbf{r} - \boldsymbol{\beta} f)\right].$$

The Jacobian with respect to $\boldsymbol{\beta}$ is

$$\begin{pmatrix} -E(f)\,\mathbf{I} \\ -E(f^2)\,\mathbf{I} \end{pmatrix},$$

which has full column rank. $\square$

Lemma G2 *Assume that the asset pricing model holds for* $\mathbf{r}$ *with a nontraded factor* f *such that* $V(f) > 0$ *and* $E(\mathbf{r}) \neq \mathbf{0}$. *Then the Jacobians of all the following moment conditions have full column rank at the true parameter values:*
1) The uncentred SDF moment conditions that rely on a symmetric normalization.
2) The centred SDF moment conditions that rely on a symmetric normalization.
3) The centred regression moment conditions that rely on a symmetric normalization.

Proof.

1) When we work with (1) and a symmetric normalization, we need to study the behavior of

$$E\left[\mathbf{r}(\sin\psi + f\cos\psi)\right].$$

In this case, the true value of the parameter is defined by

$$\begin{pmatrix} E(\mathbf{r}) & E(\mathbf{r}f) \end{pmatrix} \begin{pmatrix} \sin\psi \\ \cos\psi \end{pmatrix} = \mathbf{0},$$

while the Jacobian of the moment conditions with respect to ψ is

$$\begin{pmatrix} E(\mathbf{r}) & E(\mathbf{r}f) \end{pmatrix} \begin{pmatrix} \cos\psi \\ -\sin\psi \end{pmatrix}.$$

These two linear combinations are orthogonal at the same ψ. Therefore, the Jacobian has full column rank at the true value.

2) When we work with (2) and a symmetric normalization, we need to study

$$E\begin{bmatrix} \mathbf{r}(\sin\upsilon + (f-\mu)\cos\upsilon) \\ f-\mu \end{bmatrix}.$$

Thus, the true value of the parameters of the symmetrically normalized centred SDF are defined by

$$\begin{pmatrix} E(\mathbf{r}) & Cov(\mathbf{r}, f) \end{pmatrix} \begin{pmatrix} \sin\upsilon \\ \cos\upsilon \end{pmatrix} = \mathbf{0},$$

and $\mu = E(f)$, which make those moments equal to zero.

The Jacobian of the moment conditions with respect to (υ, μ) is

$$\begin{bmatrix} E(\mathbf{r})(\cos\upsilon + \mu\sin\upsilon) - E(\mathbf{r}f)\sin\upsilon & -E(\mathbf{r})\cos\upsilon \\ 0 & -1 \end{bmatrix}.$$

As $\mu = E(f)$ at the true values, the linear combination that defines the true υ is orthogonal to the combination that defines the first entry of the Jacobian at the true values

$$\begin{pmatrix} E(\mathbf{r}) & Cov(\mathbf{r},f) \end{pmatrix} \begin{pmatrix} \cos\upsilon \\ -\sin\upsilon \end{pmatrix}.$$

Therefore, the Jacobian must have full column rank because it first column cannot be zero.

3) When we work with (3) and a symmetric normalization, we need to study the behavior of

$$E\left[\begin{pmatrix} 1 \\ f \end{pmatrix} \otimes (\mathbf{r} + \boldsymbol{\varphi}(\cos\vartheta - \sin\vartheta f))\right].$$

The Jacobian of the moment conditions with respect to $(\boldsymbol{\varphi}, \vartheta)$ is

$$\begin{pmatrix} (\cos\vartheta - \sin\vartheta E(f))\mathbf{I} & -\boldsymbol{\varphi}(\sin\vartheta + \cos\vartheta E(f)) \\ \left(\cos\vartheta E(f) - \sin\vartheta E\left(f^2\right)\right)\mathbf{I} & -\boldsymbol{\varphi}\left(\sin\vartheta E(f) + \cos\vartheta E\left(f^2\right)\right) \end{pmatrix},$$

which has full column rank because $\boldsymbol{\varphi} \neq \mathbf{0}$ and

$$\left|\begin{pmatrix} \cos\vartheta - \sin\vartheta E(f) & \sin\vartheta + \cos\vartheta E(f) \\ \cos\vartheta E(f) - \sin\vartheta E\left(f^2\right) & \sin\vartheta E(f) + \cos\vartheta E\left(f^2\right) \end{pmatrix}\right| = V(f) > 0.$$

□

Lemma G3 *Assume that the asset pricing models holds for* $\mathbf{r}$ *with a nontraded factor* f *such that* $V(f) > 0$ *and* $E(\mathbf{r}) \neq \mathbf{0}$*. Then*
1) The Jacobians of the uncentred and centred SDF moment conditions that rely on the asymmetric normalizations $(a/b, 1)$ *and* $(c/b, 1)$*, respectively, have full column rank at the true parameter values.*
2) The Jacobian of the uncentred SDF moment conditions that rely on the asymmetric normalization $(1, b/a)$ *has full column rank at the true parameter values if and only if the factor is not orthogonal to the vector of excess returns. When* $E(\mathbf{r}f) = \mathbf{0}$*, the iterated GMM criterion function fails to identify* δ *in the population, while the CU criterion function goes to 0 as* $\delta^{-1} \to 0$*.*
3) The Jacobian of the centred SDF moment conditions that rely on the asymmetric normalization $(1, b/c)$ *has full column rank at the true parameter values if and only if the factor is not uncorrelated with the vector of excess returns. When* $Cov(\mathbf{r},f) = \mathbf{0}$*, both the CU and iterated GMM criterion functions converge to 0 in the population as* $\mu \to E(f)$ *and* $\tau(E(f) - \mu) \to 1$*.*
4) The Jacobian of the centred regression moment conditions that rely on the asymmetric normalization $(1, d/c)$ *has full column rank at the true parameter values if and only if the factor is not uncorrelated with the vector of excess returns. When* $Cov(\mathbf{r},f) = \mathbf{0}$*, both the CU and iterated GMM criterion functions converge to 0 in the population as* $\boldsymbol{\beta} \to 0$ *and* $\boldsymbol{\beta}\varkappa \to E(\mathbf{r})$*.*
5) The Jacobian of the centred regression moment conditions that rely on the asymmetric normalization $(c/d, 1)$ *has full column rank at the true parameter values if and only if the least squares projection of* $\mathbf{r}$ *onto the span of* $(1, f)$ *is not proportional to* f*. When* $E(\mathbf{r}) - Cov(\mathbf{r},f)E(f)/V(f) = \mathbf{0}$*, both the CU and iterated GMM criterion functions converge to 0 in the population as* $\boldsymbol{\phi} \to 0$ *and* $\boldsymbol{\phi}\nu \to Cov(\mathbf{r},f)/V(f)$*.*

Proof.

1) When we work (1) and the asymmetric normalization $(a/b, 1)$, we need to study the behavior of

$$E\left[\mathbf{r}\left(\mathfrak{a}+f\right)\right].$$

The Jacobian of the moment conditions with respect to $\mathfrak{a}$ is $E\left(\mathbf{r}\right)$, which is nonzero by assumption.

For the centred SDF (2) and the asymmetric normalization $(c/b, 1)$, the relevant moments are

$$E\left[\begin{array}{c}\mathbf{r}\left(\mathfrak{c}+\left(f-\mu\right)\right)\\ f-\mu\end{array}\right].$$

The Jacobian of these moment conditions with respect to $(\mathfrak{c}, \mu)$ is

$$\left(\begin{array}{cc}E\left(\mathbf{r}\right) & -E\left(\mathbf{r}\right)\\ 0 & -1\end{array}\right),$$

which has full column rank because $E\left(\mathbf{r}\right) \neq \mathbf{0}$ by assumption.

2) When we work with (1) and the asymmetric normalization $(1, b/a)$, we need to study the behavior of

$$E\left[\mathbf{r}\left(1-\delta f\right)\right].$$

The true value of δ is $-b/a$ from (1) whenever the true a is not zero. In this regard, $a = 0$ if and only if $E\left(\mathbf{r}f\right) = \mathbf{0}$. In that case, the standard GMM criterion becomes

$$E\left(\mathbf{r}\right)'\mathbf{W}E\left(\mathbf{r}\right)$$

in the population for any positive definite weighting matrix $\mathbf{W}$. Since this expression does not depend on δ, it cannot be zero. In contrast, assuming *i.i.d.* data for simplicity, the CU criterion becomes

$$E\left(\mathbf{r}\right)'\left[V\left[\mathbf{r}\left(1-\delta f\right)\right]\right]^{-1}E\left(\mathbf{r}\right)=\frac{1}{\delta^{2}}E\left(\mathbf{r}\right)'\left[V\left[\mathbf{r}\left(\frac{1}{\delta}-f\right)\right]\right]^{-1}E\left(\mathbf{r}\right),$$

which converges to zero as $\delta^{-1} \to 0$.

Finally, the Jacobian of the moment conditions with respect to δ is $-E\left(\mathbf{r}f\right)$, which has full column rank if and only if $E\left(\mathbf{r}f\right) \neq \mathbf{0}$.

3) When we work with (2) and the asymmetric normalization $(1, b/c)$, we need to study the behavior of

$$E\left[\begin{array}{c}\mathbf{r}\left(1-\tau\left(f-\mu\right)\right)\\ f-\mu\end{array}\right].$$

We can make the last moment condition equal to zero with $\mu = E(f)$, and simultaneously satisfy the pricing conditions with $-\tau^{-1} = a/b + E(f)$ whenever $a/b + E(f) \neq 0$.

Let us now study the case of $a/b + E(f) = 0$, which is equivalent to $Cov(\mathbf{r}, f) = \mathbf{0}$. In this case, the moments become

$$\begin{bmatrix} E(\mathbf{r})(1 - \tau(E(f) - \mu)) \\ E(f) - \mu \end{bmatrix},$$

which can be made arbitrarily close to zero by choosing

$$\mu \to E(f), \qquad \tau(E(f) - \mu) \to 1.$$

This would make the standard GMM criterion close to zero in the population for any positive definite weighting matrix. Nevertheless, note the discontinuity of the criterion function, which becomes strictly positive at $\mu = E(f)$ because the moments evaluated at that value are $\left(E(\mathbf{r})', 0\right)'$.

Assuming $i.i.d.$data for simplicity, the CU criterion can expressed as

$$\begin{pmatrix} E(\mathbf{r}m) \\ E(u) \end{pmatrix}' \begin{bmatrix} V(\mathbf{r}m) & Cov(\mathbf{r}m, u) \\ Cov(u, \mathbf{r}m) & V(u) \end{bmatrix}^{-1} \begin{pmatrix} E(\mathbf{r}m) \\ E(u) \end{pmatrix},$$

with $u = f - \mu$, or equivalently as

$$E(\mathbf{r}m)' [V(\mathbf{r}m)]^{-1} E(\mathbf{r}m) + \frac{\left[E(u) - Cov(u, \mathbf{r}m)[V(\mathbf{r}m)]^{-1} E(\mathbf{r}m)\right]^2}{V(u) - Cov(u, \mathbf{r}m)[V(\mathbf{r}m)]^{-1} Cov(\mathbf{r}m, u)}.$$

If we write $m = \tau\mathfrak{m}$ with $\mathfrak{m} = (1/\tau) - (f - \mu)$, we can express the CU criterion as

$$E(\mathbf{r}\mathfrak{m})' [V(\mathbf{r}\mathfrak{m})]^{-1} E(\mathbf{r}\mathfrak{m}) + \frac{\left[E(u) - Cov(u, \mathbf{r}\mathfrak{m})[V(\mathbf{r}\mathfrak{m})]^{-1} E(\mathbf{r}\mathfrak{m})\right]^2}{V(u) - Cov(u, \mathbf{r}\mathfrak{m})[V(\mathbf{r}\mathfrak{m})]^{-1} Cov(\mathbf{r}\mathfrak{m}, u)}.$$

If we take the limits $\tau E[u] \to 1$ and $\mu \to E(f)$ then the above expressions are well defined, and the criterion converges to zero because $E(\mathbf{r}\mathfrak{m}) = E(\mathbf{r}) E(\mathfrak{m}) \to \mathbf{0}$ and $E(u) \to 0$.

Finally, the Jacobian of the moment conditions with respect to (τ, μ) is

$$\begin{pmatrix} E(\mathbf{r})\mu - E(\mathbf{r}f) & E(\mathbf{r})\tau \\ 0 & -1 \end{pmatrix},$$

which has full column rank if and only if

$$E(\mathbf{r})\mu - E(\mathbf{r}f) \neq \mathbf{0}.$$

In the special case $Cov(\mathbf{r}, f) = \mathbf{0}$, the first column of the Jacobian would converge to zero, and the second one would have unbounded entries.

4) When we work with (3) and the asymmetric normalization $(1, d/c)$, we need to study the behavior of

$$E\left[\begin{pmatrix} 1 \\ f \end{pmatrix} \otimes (\mathbf{r} - \boldsymbol{\beta}(f + \varkappa))\right].$$

The true parameters must satisfy

$$E(\mathbf{r}) = \boldsymbol{\beta}(\varkappa + E(f)),$$

$$E(\mathbf{r}f) = \boldsymbol{\beta}\left(\varkappa E(f) + E\left(f^2\right)\right),$$

which means that $\varkappa = -\left(E(f)\, a/b + E\left(f^2\right)\right) / (a/b + E(f))$ and $\boldsymbol{\beta} \neq \mathbf{0}$ whenever $a/b + E(f) \neq 0$.

Now let us study the case of $a/b + E(f) = 0$, which is equivalent to $Cov(\mathbf{r}, f) = \mathbf{0}$. In that case, the moments

$$E(\mathbf{r}) - \boldsymbol{\beta}(\varkappa + E(f)),$$

$$E(\mathbf{r})\, E(f) - \boldsymbol{\beta}\left(\varkappa E(f) + E\left(f^2\right)\right)$$

can be made arbitrarily close to zero by choosing

$$\boldsymbol{\beta} \to \mathbf{0}, \qquad \boldsymbol{\beta}\varkappa \to E(\mathbf{r}).$$

This would make the standard GMM criterion close to zero in the population for any positive definite weighting matrix. Again, note the discontinuity of the criterion function, which becomes strictly positive at $\boldsymbol{\beta} = \mathbf{0}$ because the moments evaluated at that value are $\left(E(\mathbf{r})', E(\mathbf{r})' E(f)\right)'$.

Assuming *i.i.d.* data for simplicity, the CU criterion converges to

$$E\left[\begin{pmatrix} 1 \\ f \end{pmatrix} \otimes \boldsymbol{\varepsilon}\right]' \left[V\left[\begin{pmatrix} 1 \\ f \end{pmatrix} \otimes \boldsymbol{\varepsilon}\right]\right]^{-1} E\left[\begin{pmatrix} 1 \\ f \end{pmatrix} \otimes \boldsymbol{\varepsilon}\right].$$

where $\boldsymbol{\varepsilon} = \mathbf{r} - E(\mathbf{r})$. This is a quadratic form in a zero vector with a definite positive weighting matrix, which implies that the CU criterion also converges to zero.

The Jacobian of the moment conditions with respect to $(\boldsymbol{\beta}, \varkappa)$ is

$$\begin{pmatrix} -(\varkappa + E(f))\,\mathbf{I} & -\boldsymbol{\beta} \\ -\left(\varkappa E(f) + E\left(f^2\right)\right)\mathbf{I} & -\boldsymbol{\beta} E(f) \end{pmatrix},$$

which, given that

$$\left|\begin{pmatrix} \varkappa + E(f) & 1 \\ \varkappa E(f) + E\left(f^2\right) & E(f) \end{pmatrix}\right| = -V(f),$$

has full column rank if and only if $\boldsymbol{\beta} \neq \mathbf{0}$. Therefore, the Jacobian is ill-behaved when $Cov(\mathbf{r},f) = \mathbf{0}$.

5) When we work with (3) and the asymmetric normalization $(c/d, 1)$, we need to study the behavior of

$$E\left[\begin{pmatrix} 1 \\ f \end{pmatrix} \otimes (\mathbf{r} - \boldsymbol{\phi}(1 + \nu f))\right],$$

where $\nu = -c/d$. The true parameters must satisfy

$$E(\mathbf{r}) = \boldsymbol{\phi}(1 + \nu E(f)),$$

$$E(\mathbf{r}f) = \boldsymbol{\phi}\left(E(f) + \nu E\left(f^2\right)\right),$$

which means that $\nu = -(a/b + E(f)) / \left(E(f)\, a/b + E\left(f^2\right)\right)$ and $\boldsymbol{\phi} \neq \mathbf{0}$ whenever $E(f)\, a/b + E\left(f^2\right) \neq 0$.

Let us study now the case of $E(f)\, a/b + E\left(f^2\right) = 0$. The pricing condition (1) becomes

$$E(\mathbf{r}) = E(\mathbf{r}f)\frac{E(f)}{E(f^2)} = Cov(\mathbf{r},f)\frac{E(f)}{V(f)},$$

which means that the nontraded factor satisfies the same pricing condition as a traded factor. In this special case, the moments

$$E(\mathbf{r}) - \boldsymbol{\phi}(1 + \nu E(f)),$$

$$E(\mathbf{r}f) - \boldsymbol{\phi}\left(E(f) + \nu E\left(f^2\right)\right),$$

can be made arbitrarily close to zero by choosing

$$\boldsymbol{\phi} \to \mathbf{0}, \qquad \boldsymbol{\phi}\nu \to Cov(\mathbf{r},f)/V(f).$$

This would make the standard GMM criterion close to zero in the population for any positive definite weighting matrix. Assuming *i.i.d.* data for simplicity, the CU criterion converges to

$$E\left[\begin{pmatrix} 1 \\ f \end{pmatrix} \otimes \boldsymbol{\epsilon}\right]' \left[V\left[\begin{pmatrix} 1 \\ f \end{pmatrix} \otimes \boldsymbol{\epsilon}\right]\right]^{-1} E\left[\begin{pmatrix} 1 \\ f \end{pmatrix} \otimes \boldsymbol{\epsilon}\right],$$

where $\boldsymbol{\epsilon} = \mathbf{r} - f\,Cov(\mathbf{r},f)/V(f)$. This is a quadratic form in a zero vector with a definite positive weighting matrix, which means that the CU criterion also converges to zero.

The Jacobian with respect to $(\boldsymbol{\phi}, \nu)$ is

$$\begin{pmatrix} -(1 + \nu E(f))\,\mathbf{I} & -\boldsymbol{\phi} E(f) \\ -\left(E(f) + \nu E\left(f^2\right)\right)\mathbf{I} & -\boldsymbol{\phi} E\left(f^2\right) \end{pmatrix},$$

which has full column rank if and only if $\boldsymbol{\phi} \neq \mathbf{0}$ in view of the fact that

$$\left| \begin{pmatrix} 1+\nu E(f) & E(f) \\ E(f)+\nu E\left(f^2\right) & E\left(f^2\right) \end{pmatrix} \right| = V(f).$$

Therefore, the Jacobian is ill-behaved when $E(\mathbf{r}) = E(\mathbf{r}f)\,E(f)\,/E\left(f^2\right)$. □

Lemma G4 *Assume that the asset pricing model holds for* $\mathbf{r}$ *with a traded factor* f_1 *and a nontraded factor* f_2 *such that the linear span of* $(1, f_1, f_2)$ *is of dimension 3,* $E(\mathbf{r}) \neq \mathbf{0}$ *and* $E(\mathbf{r}) \neq E(\mathbf{r}f_1)\,E(f_1)\,/E\left(f_1^2\right)$*. Then the Jacobians of all the following moment conditions have full column rank at the true parameter values:*
1) The uncentred SDF moment conditions that rely on a symmetric normalization.
2) The centred SDF moment conditions that rely on a symmetric normalization.
3) The centred regression moment conditions that rely on a symmetric normalization.

Proof.

1) When we work with (A1) and a symmetric normalization, the true parameters are defined by the linear combination

$$\mathbf{M} \begin{pmatrix} \sin\psi_1 \\ \cos\psi_1 \sin\psi_2 \\ \cos\psi_1 \cos\psi_2 \end{pmatrix} = \mathbf{0},$$

where

$$\mathbf{M} = \begin{pmatrix} E(\mathbf{r}) & E(\mathbf{r}f_1) & E(\mathbf{r}f_2) \\ E(f_1) & E\left(f_1^2\right) & E(f_1 f_2) \end{pmatrix}.$$

The matrix $\mathbf{M}$ has rank 2 under the assumptions of the lemma, and hence the moment conditions above identify a unique (ψ_1, ψ_2).

The Jacobian is given by two different linear combinations

$$\mathbf{M} \begin{pmatrix} \cos\psi_1 & 0 \\ -\sin\psi_1 \sin\psi_2 & \cos\psi_1 \cos\psi_2 \\ -\sin\psi_1 \cos\psi_2 & -\cos\psi_1 \sin\psi_2 \end{pmatrix}.$$

Since these linear combinations are orthogonal to the ones that define the true parameters, the Jacobian will have rank 2 at those values.

2) Let us define the matrix

$$\mathbf{N}(\boldsymbol{\mu}) = \begin{pmatrix} E(\mathbf{r}) & E(\mathbf{r}u_1) & E(\mathbf{r}u_2) \\ E(f_1) & E(f_1 u_1) & E(f_1 u_2) \end{pmatrix},$$

where $u_i = f_i - \mu_i$ for $i = 1, 2$. When we work with (A2) and a symmetric normalization, the true parameters are defined by $\boldsymbol{\mu} = E(\mathbf{f})$ and the linear combination

$$\mathbf{N}[E(\mathbf{f})] \begin{pmatrix} \sin v_1 \\ \cos v_1 \sin v_2 \\ \cos v_1 \cos v_2 \end{pmatrix} = \mathbf{0}.$$

The true values are unique because $\mathbf{N}[E(\mathbf{f})]$ has rank 2 under the assumptions of the lemma.

The Jacobian is given by

$$\begin{pmatrix} \mathbf{N}[E(\mathbf{f})]\mathbf{A}(v_1, v_2) & \mathbf{N}[E(\mathbf{f})]\mathbf{B}(v_1, v_2) \\ \mathbf{0} & -\mathbf{I} \end{pmatrix},$$

where

$$\mathbf{A}(v_1, v_2) = \begin{pmatrix} \cos v_1 & 0 \\ -\sin v_1 \sin v_2 & \cos v_1 \cos v_2 \\ -\sin v_1 \cos v_2 & -\cos v_1 \sin v_2 \end{pmatrix},$$

$$\mathbf{B}(v_1, v_2) = \begin{pmatrix} -\cos v_1 \sin v_2 & -\cos v_1 \cos v_2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

The Jacobian has rank 4 if and only if $\mathbf{N}[E(\mathbf{f})]\mathbf{A}(v_1, v_2)$ has in turn rank 2. But since this matrix is given by two linear combinations of $\mathbf{N}[E(\mathbf{f})]$ that are orthogonal to each other, and orthogonal to the linear combination that defines the true parameters, it has indeed has rank 2.

3) When we work with (A3) and the symmetric normalization

$$(c, d, \boldsymbol{\varphi}, \boldsymbol{\beta}_1) = (\sin\vartheta, \cos\vartheta, \boldsymbol{\varphi}, \boldsymbol{\beta}_1),$$

the true values must satisfy

$$E(\mathbf{r}) - \boldsymbol{\beta}_1 E(f_1) + \boldsymbol{\varphi}[\cos\vartheta - E(f_2)\sin\vartheta] = \mathbf{0},$$

$$E(\mathbf{r}f_1) - \boldsymbol{\beta}_1 E\left(f_1^2\right) + \boldsymbol{\varphi}[E(f_1)\cos\vartheta - E(f_1 f_2)\sin\vartheta] = \mathbf{0},$$

$$E(\mathbf{r}f_2) - \boldsymbol{\beta}_1 E(f_1 f_2) + \boldsymbol{\varphi}\left[E(f_2)\cos\vartheta - E\left(f_2^2\right)\sin\vartheta\right] = \mathbf{0}.$$

The Jacobian of the moments with respect to $(\boldsymbol{\beta}_1, \boldsymbol{\varphi}, \vartheta)$ is

$$\begin{pmatrix} -E(f_1)\mathbf{I} & (\cos\vartheta - E(f_2)\sin\vartheta)\mathbf{I} & -(\sin\vartheta + E(f_2)\cos\vartheta)\boldsymbol{\varphi} \\ -E\left(f_1^2\right)\mathbf{I} & (E(f_1)\cos\vartheta - E(f_1 f_2)\sin\vartheta)\mathbf{I} & -(E(f_1)\sin\vartheta + E(f_1 f_2)\cos\vartheta)\boldsymbol{\varphi} \\ -E(f_1 f_2)\mathbf{I} & \left(E(f_2)\cos\vartheta - E\left(f_2^2\right)\sin\vartheta\right)\mathbf{I} & -\left(E(f_2)\sin\vartheta + E\left(f_2^2\right)\cos\vartheta\right)\boldsymbol{\varphi} \end{pmatrix},$$

which has full column rank under the assumptions of the lemma because $\boldsymbol{\varphi} \neq \mathbf{0}$ and

$$\left| \begin{pmatrix} -E(f_1) & \cos\vartheta - E(f_2)\sin\vartheta & -(\sin\vartheta + E(f_2)\cos\vartheta) \\ -E\left(f_1^2\right) & E(f_1)\cos\vartheta - E(f_1 f_2)\sin\vartheta & -(E(f_1)\sin\vartheta + E(f_1 f_2)\cos\vartheta) \\ -E(f_1 f_2) & E(f_2)\cos\vartheta - E\left(f_2^2\right)\sin\vartheta & -\left(E(f_2)\sin\vartheta + E\left(f_2^2\right)\cos\vartheta\right) \end{pmatrix} \right|$$

$$= \left[E^2(f_1 f_2) - E\left(f_1^2\right)E\left(f_2^2\right)\right] + \left[E\left(f_2^2\right)E^2(f_1) + E\left(f_1^2\right)E^2(f_2) - 2E(f_1 f_2)E(f_1)E(f_2)\right]$$

$$= -\left|E\left(\mathbf{y}\mathbf{y}'\right)\right|, \qquad \mathbf{y} = (1, f_1, f_2).$$

□

Lemma G5 *Assume that the asset pricing model holds for* $\mathbf{r}$ *with a traded factor* f_1 *and a nontraded factor* f_2 *such that the linear span of* $(1, f_1, f_2)$ *is of dimension 3,* $E(\mathbf{r}) \neq \mathbf{0}$ *and* $E(\mathbf{r}) \neq E(\mathbf{r}f_1)E(f_1)/E\left(f_1^2\right)$*. Then*
1) The Jacobian of the uncentred SDF moment conditions that rely on the asymmetric normalization $(1, b_1/a, b_2/a)$ *has full column rank at the true parameter values if and only if the residual of the uncentred regression of* f_2 *onto* f_1 *is not orthogonal to the vector of excess returns. When* $E(\mathbf{r}f_2) = E(\mathbf{r}f_1)E(f_1 f_2)/E\left(f_1^2\right)$*, the iterated GMM criterion function fails to identify* (δ_1, δ_2) *in the population, while the CU criterion function goes to 0 as* $\delta_2^{-1} \to 0$ *and* $\delta_1/\delta_2 \to -E(f_1 f_2)/E\left(f_1^2\right)$*.*
2) The Jacobian of the centred SDF moment conditions that rely on the asymmetric normalization $(1, b_1/c, b_2/c)$ *has full column rank at the true parameter values if and only if the residual of the centred regression of* f_2 *onto* f_1 *is not uncorrelated with the vector of excess returns. When* $Cov(\mathbf{r}, f_2) = Cov(\mathbf{r}, f_1)Cov(f_1, f_2)/V(f_1)$*, both the CU and iterated GMM criterion functions converge to 0 in the population as* $\boldsymbol{\mu} \to E(\mathbf{f})$*, while* $\boldsymbol{\tau}'(E(\mathbf{f}) - \boldsymbol{\mu}) \to 1$ *and* $\tau_1/\tau_2 \to -Cov(f_1, f_2)/V(f_1)$*.*
3) The Jacobian of the centred regression moment conditions that rely on the asymmetric normalization $(1, d/c)$ *has full column rank at the true parameter values if and only if the residual of the centred regression of* f_2 *onto* f_1 *is not uncorrelated with the vector of excess returns. When* $Cov(\mathbf{r}, f_2) = Cov(\mathbf{r}, f_1)Cov(f_1, f_2)/V(f_1)$*, both the CU and iterated GMM criterion functions converge to 0 in the population as* $\boldsymbol{\beta}_2 \to \mathbf{0}$ *and* $\boldsymbol{\beta}_2\varkappa \to E(\mathbf{r}) - \boldsymbol{\beta}_1 E(f_1)$*, jointly with* $\boldsymbol{\beta}_1 \to Cov(\mathbf{r}, f_1)/V(f_1)$*.*

Proof.

1) When we work with (A1) and the asymmetric normalization $(1, b_1/a, b_2/a)$, we need to study the behavior of

$$\begin{pmatrix} E(\mathbf{r}) & E(\mathbf{r}f_1) & E(\mathbf{r}f_2) \\ E(f_1) & E\left(f_1^2\right) & E(f_1 f_2) \end{pmatrix} \begin{pmatrix} 1 \\ -\delta_1 \\ -\delta_2 \end{pmatrix}.$$

The Jacobian of these moments with respect to (δ_1, δ_2) is equal to

$$-\begin{pmatrix} E(\mathbf{r}f_1) & E(\mathbf{r}f_2) \\ E\left(f_1^2\right) & E(f_1 f_2) \end{pmatrix},$$

which does not necessarily have rank 2 under the assumptions of the lemma. The necessary and sufficient condition for rank 2 is

$$E(\mathbf{r}f_2) \neq E(\mathbf{r}f_1)\frac{E(f_1f_2)}{E(f_1^2)}.$$

When $E(\mathbf{r}f_2) = E(\mathbf{r}f_1)E(f_1f_2)/E(f_1^2)$, the moment conditions (A1) collapse to

$$\begin{aligned} E(\mathbf{r})a + E(\mathbf{r}f_1)\left(b_1 + \frac{E(f_1f_2)}{E(f_1^2)}b_2\right) &= \mathbf{0}, \\ E(f_1)a + E(f_1^2)\left(b_1 + \frac{E(f_1f_2)}{E(f_1^2)}b_2\right) &= 0, \end{aligned}$$

which cannot be zero at any nonzero vector $\left(a, b_1 + b_2E(f_1f_2)/E(f_1^2)\right)$ due to the assumptions of the lemma. Hence the true parameter values must satisfy

$$a = 0, \qquad b_1 + \frac{E(f_1f_2)}{E(f_1^2)}b_2 = 0.$$

In contrast, if we turn to the asymmetric normalization, the moments collapse to

$$\begin{bmatrix} E(\mathbf{r}) - E(\mathbf{r}f_1)\left(\delta_1 + \frac{E(f_1f_2)}{E(f_1^2)}\delta_2\right) \\ E(f_1) - E(f_1^2)\left(\delta_1 + \frac{E(f_1f_2)}{E(f_1^2)}\delta_2\right) \end{bmatrix}$$

and hence a standard GMM criterion with a fixed weighting matrix cannot be zero in the population. Moreover, any values for the pair (δ_1, δ_2) which give rise to the same $\delta_1 + \delta_2E(f_1f_2)/E(f_1^2)$ yield the same value of the criterion.

Assuming *i.i.d.* data for simplicity, the CU criterion has a weighting matrix equal to the inverse of

$$V\left[\begin{pmatrix} \mathbf{r} \\ f_1 \end{pmatrix}(1 - \delta_1f_1 - \delta_2f_2)\right] = \delta_2^2V\left[\begin{pmatrix} \mathbf{r} \\ f_1 \end{pmatrix}\left(\frac{1}{\delta_2} - \frac{\delta_1}{\delta_2}f_1 - f_2\right)\right].$$

Similarly we can express the moments as

$$\delta_2\begin{bmatrix} E(\mathbf{r})\frac{1}{\delta_2} - E(\mathbf{r}f_1)\left(\frac{\delta_1}{\delta_2} + \frac{E(f_1f_2)}{E(f_1^2)}\right) \\ E(f_1)\frac{1}{\delta_2} - E(f_1^2)\left(\frac{\delta_1}{\delta_2} + \frac{E(f_1f_2)}{E(f_1^2)}\right) \end{bmatrix}.$$

Therefore, the CU criterion will converge to zero if we take the limits

$$\frac{1}{\delta_2} \to 0, \qquad \frac{\delta_1}{\delta_2} \to -\frac{E(f_1f_2)}{E(f_1^2)}.$$

2) When we work with (A2) and the asymmetric normalization $(1, b_1/c, b_2/c)$, we need to study

$$\left(\begin{array}{c}\begin{pmatrix} E(\mathbf{r}) & E(\mathbf{r}u_1) & E(\mathbf{r}u_2) \\ E(f_1) & E(f_1u_1) & E(f_1u_2) \end{pmatrix}\begin{pmatrix} 1 \\ -\tau_1 \\ -\tau_2 \end{pmatrix} \\ E(\mathbf{f}) - \boldsymbol{\mu} \end{array}\right),$$

where $u_i = f_i - \mu_i$ for $i = 1, 2$.

There is a well-behaved mapping from these asymmetrically normalized parameters to the ones in (A2) with a symmetric normalization whenever $Cov(\mathbf{r},f_2) \neq Cov(\mathbf{r},f_1)\, Cov(f_1, f_2)/V(f_1)$.

Let us now study the problematic case of $Cov(\mathbf{r},f_2) = Cov(\mathbf{r},f_1)\, Cov(f_1, f_2)/V(f_1)$. If we evaluate the moments (A2) at $\boldsymbol{\mu} = E(\mathbf{f})$, then we can make the last two moments equal to zero, so that we can focus on the pricing conditions, which collapse to

$$\begin{aligned} E(\mathbf{r})\, c + Cov(\mathbf{r},f_1)\left(b_1 + \frac{Cov(f_1, f_2)}{V(f_1)} b_2\right) &= \mathbf{0}, \\ E(f_1)\, c + V(f_1)\left(b_1 + \frac{Cov(f_1, f_2)}{V(f_1)} b_2\right) &= 0. \end{aligned}$$

These conditions cannot be zero at any nonzero vector $(c, b_1 + b_2 Cov(f_1, f_2)/V(f_1))$ due to the assumptions of the lemma. Hence the true parameter values must satisfy

$$c = 0, \qquad b_1 + \frac{Cov(f_1, f_2)}{V(f_1)} b_2 = 0.$$

However, if we turn to the asymmetric normalization, its moments collapse to

$$\begin{bmatrix} E(\mathbf{r})\left[1 - \tau_1 (E(f_1) - \mu_1) - \tau_2 (E(f_2) - \mu_2)\right] - Cov(\mathbf{r},f_1)\left[\tau_1 + \frac{Cov(f_1,f_2)}{V(f_1)}\tau_2\right] \\ E(f_1)\left[1 - \tau_1 (E(f_1) - \mu_1) - \tau_2 (E(f_2) - \mu_2)\right] - V(f_1)\left[\tau_1 + \frac{Cov(f_1,f_2)}{V(f_1)}\tau_2\right] \\ E(f_1) - \mu_1 \\ E(f_2) - \mu_2 \end{bmatrix}$$

in this special case. These moments can be made arbitrarily close to zero by choosing

$$\boldsymbol{\mu} \to E(\mathbf{f}), \qquad \boldsymbol{\tau}'(E(\mathbf{f}) - \boldsymbol{\mu}) \to 1, \qquad \frac{\tau_1}{\tau_2} \to -\frac{Cov(f_1, f_2)}{V(f_1)}.$$

This would make the standard GMM criterion close to zero in the population for any positive definite weighting matrix. Once again, note the discontinuity of the criterion at $\boldsymbol{\mu} = E(\mathbf{f})$ because the pricing conditions cannot be zero at that value of $\boldsymbol{\mu}$, which renders the criterion strictly positive.

Assuming *i.i.d.* data for simplicity, the CU criterion can expressed as

$$\begin{pmatrix} E(\mathbf{x}m) \\ E(\mathbf{u}) \end{pmatrix}' \begin{bmatrix} V(\mathbf{x}m) & Cov(\mathbf{x}m, \mathbf{u}) \\ Cov(\mathbf{u}, \mathbf{x}m) & V(\mathbf{u}) \end{bmatrix}^{-1} \begin{pmatrix} E(\mathbf{x}m) \\ E(\mathbf{u}) \end{pmatrix},$$

where $\mathbf{x} = (f_1, \mathbf{r}')'$ and $\mathbf{u} = E(\mathbf{f}) - \boldsymbol{\mu}$, or equivalently as

$$\begin{gathered} E(\mathbf{x}m)' [V(\mathbf{x}m)]^{-1} E(\mathbf{x}m) \\ + \left[E(\mathbf{u}) - Cov(\mathbf{u}, \mathbf{x}m) [V(\mathbf{x}m)]^{-1} E(\mathbf{x}m)\right]' \\ \times \left[V(\mathbf{u}) - Cov(\mathbf{u}, \mathbf{x}m) [V(\mathbf{x}m)]^{-1} Cov(\mathbf{x}m, \mathbf{u})\right]^{-1} \left[E(\mathbf{u}) - Cov(\mathbf{u}, \mathbf{x}m) [V(\mathbf{x}m)]^{-1} E(\mathbf{x}m)\right]. \end{gathered}$$

If we write $m = \tau_2 \mathfrak{m}$ with $\mathfrak{m} = (1/\tau_2) - (\tau_1/\tau_2)(f_1 - \mu_1) - (f_2 - \mu_2)$, we can express the CU criterion as

$$\begin{gathered} E(\mathbf{x}\mathfrak{m})' [V(\mathbf{x}\mathfrak{m})]^{-1} E(\mathbf{x}\mathfrak{m}) \\ + \left[E(\mathbf{u}) - Cov(\mathbf{u}, \mathbf{x}\mathfrak{m}) [V(\mathbf{x}\mathfrak{m})]^{-1} E(\mathbf{x}\mathfrak{m})\right]' \\ \times \left[V(\mathbf{u}) - Cov(\mathbf{u}, \mathbf{x}\mathfrak{m}) [V(\mathbf{x}\mathfrak{m})]^{-1} Cov(\mathbf{x}\mathfrak{m}, \mathbf{u})\right]^{-1} \left[E(\mathbf{u}) - Cov(\mathbf{u}, \mathbf{x}\mathfrak{m}) [V(\mathbf{x}\mathfrak{m})]^{-1} E(\mathbf{x}\mathfrak{m})\right]. \end{gathered}$$

Now, if we take again the limits

$$\boldsymbol{\mu} \to E(\mathbf{f}), \qquad \boldsymbol{\tau}'(E(\mathbf{f}) - \boldsymbol{\mu}) \to 1, \qquad \frac{\tau_1}{\tau_2} \to -\frac{Cov(f_1, f_2)}{V(f_1)}.$$

then the above expressions are well defined and the CU criterion converges to zero as $E(\mathbf{x}\mathfrak{m}) \to \mathbf{0}$ and $E(\mathbf{u}) \to \mathbf{0}$.

Finally, the Jacobian of the moments with respect to $(\tau_1, \tau_2, \mu_1, \mu_2)$ is

$$\begin{pmatrix} -\begin{pmatrix} E(\mathbf{r}u_1) & E(\mathbf{r}u_2) \\ E(f_1 u_1) & E(f_1 u_2) \end{pmatrix} & \begin{pmatrix} \tau_1 E(\mathbf{r}) & \tau_2 E(\mathbf{r}) \\ \tau_1 E(f_1) & \tau_2 E(f_1) \end{pmatrix} \\ \mathbf{0} & -\mathbf{I} \end{pmatrix}.$$

This matrix is ill-behaved in the previous limit because there is a rank failure in the first two columns, while the last two columns would have unbounded entries.

3) When we work with (A3) and the asymmetric normalization

$$(1, d/c, \boldsymbol{\varphi}, \boldsymbol{\beta}_1) = (1, -\varkappa, \boldsymbol{\beta}_2, \boldsymbol{\beta}_1),$$

the true parameters must satisfy

$$\begin{aligned} E(\mathbf{r}) &= \boldsymbol{\beta}_1 E(f_1) + \boldsymbol{\beta}_2 (\varkappa + E(f_2)), \\ E(\mathbf{r} f_1) &= \boldsymbol{\beta}_1 E\left(f_1^2\right) + \boldsymbol{\beta}_2 (\varkappa E(f_1) + E(f_2 f_1)), \\ E(\mathbf{r} f_2) &= \boldsymbol{\beta}_1 E(f_2 f_1) + \boldsymbol{\beta}_2 \left(\varkappa E(f_2) + E\left(f_2^2\right)\right). \end{aligned}$$

These true parameters are related to (A1) by

$$\varkappa = -\left(aE\left(f_2\right) + b_1E\left(f_2f_1\right) + b_2E\left(f_2^2\right)\right) / \left(a + b_1E\left(f_1\right) + b_2E\left(f_2\right)\right)$$

whenever $a + b_1E(f_1) + b_2E(f_2) \neq 0$. We will also have $\boldsymbol{\beta}_2 \neq \mathbf{0}$ in those circumstances.

Let us now study the problematic case $a + b_1E(f_1) + b_2E(f_2) = 0$, which we know is equivalent to $Cov(\mathbf{r}, f_2) = Cov(\mathbf{r}, f_1)\,Cov(f_1, f_2)/V(f_1)$. In that case, the moments (A3) with the asymmetric normalization become

$$\begin{gathered}
E(\mathbf{r}) - \boldsymbol{\beta}_1 E(f_1) - \boldsymbol{\beta}_2(\varkappa + E(f_2)), \\
[Cov(\mathbf{r}, f_1) + E(\mathbf{r})E(f_1)] \\
-\boldsymbol{\beta}_1\left[V(f_1) + E^2(f_1)\right] - \boldsymbol{\beta}_2(\varkappa E(f_1) + [Cov(f_2, f_1) + E(f_2)E(f_1)]), \\
[Cov(\mathbf{r}, f_1)\,Cov(f_1, f_2)/V(f_1) + E(\mathbf{r})E(f_2)] \\
-\boldsymbol{\beta}_1[Cov(f_2, f_1) + E(f_2)E(f_1)] - \boldsymbol{\beta}_2\left(\varkappa E(f_2) + \left[V(f_2) + E^2(f_2)\right]\right),
\end{gathered}$$

which can be made arbitrarily close to zero by choosing

$$\boldsymbol{\beta}_2 \to \mathbf{0}, \qquad \boldsymbol{\beta}_2\varkappa \to E(\mathbf{r}) - \boldsymbol{\beta}_1 E(f_1), \qquad \boldsymbol{\beta}_1 \to Cov(\mathbf{r}, f_1)/V(f_1).$$

This limit would make the standard GMM criterion close to zero in the population for any positive definite weighting matrix. But since the moment conditions cannot be zero when $\boldsymbol{\beta}_2 = \mathbf{0}$, the criterion function will be strictly positive at that value, which results once more in a discontinuity.

Assuming *i.i.d.* data for simplicity, the CU criterion converges to

$$E\left[\begin{pmatrix} 1 \\ \mathbf{f} \end{pmatrix} \otimes \boldsymbol{\varepsilon}\right]' \left[V\left[\begin{pmatrix} 1 \\ \mathbf{f} \end{pmatrix} \otimes \boldsymbol{\varepsilon}\right]\right]^{-1} E\left[\begin{pmatrix} 1 \\ \mathbf{f} \end{pmatrix} \otimes \boldsymbol{\varepsilon}\right].$$

where $\boldsymbol{\varepsilon} = (\mathbf{r} - E(\mathbf{r})) - (f_1 - E(f_1))\,Cov(\mathbf{r}, f_1)/V(f_1)$. This criterion is a quadratic form in a zero vector with a definite positive weighting matrix. Hence the CU criterion also goes to zero along the previously defined limits.

Nevertheless, the Jacobian of the moments with respect to $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \varkappa)$

$$\begin{pmatrix}
-E(f_1)\mathbf{I} & -(E(f_2) + \varkappa)\mathbf{I} & -\boldsymbol{\beta}_2 \\
-E\left(f_1^2\right)\mathbf{I} & -(E(f_1f_2) + \varkappa E(f_1))\mathbf{I} & -\boldsymbol{\beta}_2 E(f_1) \\
-E(f_1f_2)\mathbf{I} & -\left(E\left(f_2^2\right) + \varkappa E(f_2)\right)\mathbf{I} & -\boldsymbol{\beta}_2 E(f_2)
\end{pmatrix}$$

does not necessarily have rank $2n+1$ under the assumptions of the lemma. The necessary and sufficient condition for rank $2n+1$ is $\boldsymbol{\beta}_2 \neq \mathbf{0}$ because the determinant

$$\left| \begin{pmatrix} -E(f_1) & -(E(f_2)+\varkappa) & -1 \\ -E(f_1^2) & -(E(f_1 f_2)+\varkappa E(f_1)) & -E(f_1) \\ -E(f_1 f_2) & -(E(f_2^2)+\varkappa E(f_2)) & -E(f_2) \end{pmatrix} \right|$$

$$= \left[E^2(f_1 f_2) - E(f_1^2)E(f_2^2)\right] + \left[E(f_2^2)E^2(f_1) + E(f_1^2)E^2(f_2) - 2E(f_1 f_2)E(f_1)E(f_2)\right]$$

$$= -\left|E(\mathbf{y}\mathbf{y}')\right|, \qquad \mathbf{y} = (1, f_1, f_2).$$

is different from zero under the assumptions of the lemma. Therefore, this Jacobian is ill-behaved in the special case $Cov(\mathbf{r}, f_2) = Cov(\mathbf{r}, f_1)\, Cov(f_1, f_2)/V(f_1)$. □

Lemma G6 *Assume that the asset pricing models holds for* $\mathbf{r}$ *with a traded or nontraded factor* f *such that* $V(f) > 0$ *and* $E(\mathbf{r}) \neq \mathbf{0}$. *Then, when we add a gross return, the Jacobians of all the following moment conditions have full column rank at the true parameter values:*
1) The uncentred SDF moment conditions that rely on a symmetric normalization.
2) The centred SDF moment conditions that rely on a symmetric normalization.
3) The regression moment conditions when f *is traded, and the centred regression moment conditions that rely on a symmetric normalization when* f *is nontraded.*

Proof.

1) If the factor is nontraded then we need to study

$$E\begin{bmatrix} \mathbf{r}(\sin\psi + f\cos\psi) \\ R(\sin\psi + f\cos\psi) - q, \end{bmatrix}$$

when we work with (1), (B7) and a symmetric normalization. Here the true value of the parameters are defined by

$$\begin{pmatrix} E(\mathbf{r}) & E(\mathbf{r}f) \end{pmatrix} \begin{pmatrix} \sin\psi \\ \cos\psi \end{pmatrix} = \mathbf{0},$$

and $q = E[R(\sin\psi + f\cos\psi)]$.

The Jacobian of the moment conditions with respect to (ψ, q) is

$$\begin{pmatrix} \begin{pmatrix} E(\mathbf{r}) & E(\mathbf{r}f) \end{pmatrix} \begin{pmatrix} \cos\psi \\ -\sin\psi \end{pmatrix} & \mathbf{0} \\ \begin{pmatrix} E(R) & E(Rf) \end{pmatrix} \begin{pmatrix} \cos\psi \\ -\sin\psi \end{pmatrix} & -1 \end{pmatrix}.$$

This matrix has full column rank if and only if the first column is not proportional to the second one. This is the case at the true values because the upper block of the first column cannot be zero following the argument of the proof of point 1) in Lemma G2.

If the factor is a traded excess return then similar arguments apply with $\mathbf{x} = (f, \mathbf{r}')'$ instead of $\mathbf{r}$.

2) If the factor is nontraded then we need to study

$$E\begin{bmatrix} \mathbf{r}\left(\sin \upsilon + (f-\mu)\cos \upsilon\right) \\ f-\mu \\ R\left(\sin \upsilon + (f-\mu)\cos \upsilon\right) - q \end{bmatrix}$$

when we work with (2), (B8) and a symmetric normalization. In this case, the true value of the parameters of the symmetrically normalized version of the centred SDF are defined by the same υ as in the proof of point 2) in Lemma G2, $\mu = E(f)$ and $q = E\left[R\left(\sin \upsilon + (f-\mu)\cos \upsilon\right)\right]$.

The Jacobian of the moment conditions with respect to (υ, μ, q) is

$$\begin{bmatrix} E(\mathbf{r})(\cos \upsilon + \mu \sin \upsilon) - E(\mathbf{r}f)\sin \upsilon & -E(\mathbf{r})\cos \upsilon & \mathbf{0} \\ 0 & -1 & 0 \\ E(R)(\cos \upsilon + \mu \sin \upsilon) - E(Rf)\sin \upsilon & -E(R)\cos \upsilon & -1 \end{bmatrix}.$$

Thus, the Jacobian has full column rank at the true values because the upper block of the first column cannot be zero following the argument of the proof of point 2) in Lemma G2.

If the factor is a traded excess return then similar arguments apply with $\mathbf{x} = (f, \mathbf{r}')'$ instead of $\mathbf{r}$.

3) If the factor is nontraded then we need to study

$$E\begin{bmatrix} \begin{pmatrix} 1 \\ f \end{pmatrix} \otimes \left(\mathbf{r} + \boldsymbol{\varphi}\left(\cos \vartheta - \sin \vartheta f\right)\right) \\ \begin{pmatrix} 1 \\ f \end{pmatrix} \otimes \left(R - \phi_R - \beta_R f\right) \end{bmatrix}$$

when we work with the normal equations of the two projections (3) and (A6), and (3) is symmetrically normalized.

In this case, the Jacobian is block diagonal because we have a separation of parameters between the excess returns conditions and the gross return conditions. We prove in point 3) of Lemma G2 that the block corresponding to the excess returns has full column rank. This is also the case for the block corresponding to the gross return because it is derived from an unrestricted least squares projection with $V(f) > 0$.

If the factor is traded then the residual of the first projection simplifies to $\mathbf{r} - \boldsymbol{\beta} f$ and we do not need a normalization. Therefore, the previous arguments are even easier to apply. $\square$

Lemma G7 *Assume that the asset pricing models holds for* $\mathbf{r}$ *with a nontraded factor* f *such that* $V(f) > 0$ *and* $E(\mathbf{r}) \neq \mathbf{0}$*. Then, when we add a gross return*
1) The Jacobian of the uncentred SDF moment conditions that rely on the asymmetric normalization $(1, b/a)$ *and (B7) has full column rank at the true parameter values if and only if the factor is not orthogonal to the vector of excess returns. When* $E(\mathbf{r}f) = \mathbf{0}$*, the iterated GMM criterion function fails to identify* (δ, q) *in the population, while the CU criterion function goes to 0 as* $\delta^{-1} \to 0$ *and* $q/\delta \to -E(Rf)$*.*
2) The Jacobian of the centred SDF moment conditions that rely on the asymmetric normalization $(1, b/c)$ *and (B8) has full column rank at the true parameter values if and only if the factor is not uncorrelated with the vector of excess returns. When* $Cov(\mathbf{r}, f) = \mathbf{0}$*, both the CU and iterated GMM criterion functions converge to 0 in the population as* $\mu \to E(f)$*,* $\tau(E(f) - \mu) \to 1$ *and* $q/\tau \to -Cov(R, f)$*.*
3) The Jacobian of the centred regression moment conditions that rely on the estimation of $(\boldsymbol{\beta}, \varkappa, \beta_R, \varkappa_R)$ *has full column rank at the true parameter values if and only if the factor is not uncorrelated with the vector of excess returns. When* $Cov(\mathbf{r}, f) = \mathbf{0}$*, both the CU and iterated GMM criterion functions converge to 0 in the population as* $\boldsymbol{\beta} \to 0$ *and* $\boldsymbol{\beta}\varkappa \to E(\mathbf{r})$*, jointly with* $\beta_R \to Cov(R, f)/V(f)$ *and* $\varkappa_R + \beta_R\varkappa \to E(R) - \beta_R E(f)$*.*
4) The Jacobian of the uncentred SDF moment conditions that rely on (A4) and estimates (a, b) *directly has full column rank at the true parameter values if and only if the SDFs that price* $\mathbf{r}$ *assign a nonzero cost to* R*. When* $aE(R) + bE(Rf) = 0$ *along the same direction in* (a, b) *space as* $aE(\mathbf{r}) + bE(\mathbf{r}f) = \mathbf{0}$*, the iterated GMM criterion function cannot be equal to zero in the population, while the CU criterion function goes to 0 as* $(a^2 + b^2)^{-1/2} \to 0$ *along the common direction.*
5) The Jacobian of the centred SDF moment conditions that rely on (A5) and estimates (c, b, μ) *directly has full column rank at the true parameter values if and only if the SDFs that price* $\mathbf{r}$ *assign a nonzero cost to* R*. When* $cE(R) + bCov(R, f) = 0$ *along the same direction in* (c, b) *space as* $cE(\mathbf{r}) + bCov(\mathbf{r}, f) = \mathbf{0}$*, the iterated GMM criterion function cannot be equal to zero in the population, while the CU criterion function goes to 0 as* $(c^2 + b^2)^{-1/2} \to 0$ *along the common direction.*

Proof.

1) When we work with (1), (B7) and the asymmetric normalization $(1, b/a)$, we need to study

$$E\begin{pmatrix} \mathbf{r}(1 - \delta f) \\ R(1 - \delta f) - q \end{pmatrix},$$

with parameters (δ, q). In this context, we find that the true parameter values in (1) and (B7) are $\delta = -b/a$ and $q = E(R(1 + (a/b)f))$ for both standard and CU-GMM whenever $a \neq 0$.

On the other hand, $a = 0$ if and only if $E(\mathbf{r}f) = \mathbf{0}$. In that special case, the moments collapse to

$$\begin{pmatrix} E(\mathbf{r}) \\ E(R) - E(Rf)\delta - q \end{pmatrix}$$

and we have a problem of identification with standard GMM. However, CU is invariant to dividing the moments by δ

$$\begin{pmatrix} \frac{1}{\delta}E(\mathbf{r}) \\ \frac{1}{\delta}E(R) - E(Rf) - \frac{q}{\delta} \end{pmatrix}$$

and we can see that we can make the CU criterion arbitrarily close to 0 by choosing $\delta^{-1} \to 0$ and $q/\delta \to -E(Rf)$.

The Jacobian of the moment conditions with respect to (δ, q)

$$\begin{pmatrix} -E(\mathbf{r}f) & \mathbf{0} \\ -E(Rf) & -1 \end{pmatrix}$$

has full column rank if and only if $E(\mathbf{r}f) \neq \mathbf{0}$, like in the full rank case without a gross return.

2) When we work with (2), (B8) and the asymmetric normalization $(1, b/c)$, we need to study

$$E\begin{pmatrix} \mathbf{r}(1-\tau(f-\mu)) \\ f-\mu \\ R(1-\tau(f-\mu))-q \end{pmatrix},$$

with parameters (τ, q). We can make the second moment condition equal to zero with $\mu = E(f)$, and simultaneously satisfy the pricing conditions with $-\tau^{-1} = a/b + E(f)$ and $q = E[R(1+(f-E(f))/(a/b+E(f)))]$ whenever $a/b + E(f) \neq 0$ in (2) and (B8).

Now let us study the case of $a/b + E(f) = 0$, which is equivalent to $Cov(\mathbf{r},f) = \mathbf{0}$. In this case, the moments become

$$\begin{bmatrix} E(\mathbf{r})(1-\tau(E(f)-\mu)) \\ E(f)-\mu \\ E(R)-\tau E(R(f-\mu))-q \end{bmatrix},$$

and they can be made arbitrarily close to zero by choosing

$$\mu \to E(f), \qquad \tau(E(f)-\mu) \to 1, \qquad q/\tau \to -Cov(R,f)$$

This would make the standard GMM criterion arbitrarily close to zero in the population for any positive definite weighting matrix. Once again note the discontinuity of the criterion at $\mu = E(f)$.

Assuming *i.i.d.* data for simplicity, the CU criterion can expressed as

$$\begin{pmatrix} E(\mathbf{y}) \\ E(u) \end{pmatrix}' \begin{bmatrix} V(\mathbf{y}) & Cov(\mathbf{y},u) \\ Cov(u,\mathbf{y}) & V(u) \end{bmatrix}^{-1} \begin{pmatrix} E(\mathbf{y}) \\ E(u) \end{pmatrix},$$

where $\mathbf{y} = \left((\mathbf{r}(1-\tau(f-\mu)))', R(1-\tau(f-\mu))-q\right)'$ and $u = E(f)-\mu$, or equivalently as

$$E(\mathbf{y})'[V(\mathbf{y})]^{-1}E(\mathbf{y}) + \frac{\left[E(u) - Cov(u,\mathbf{y})[V(\mathbf{y})]^{-1}E(\mathbf{y})\right]^2}{\left[V(u) - Cov(u,\mathbf{y})[V(\mathbf{y})]^{-1}Cov(\mathbf{y},u)\right]}.$$

We can write $\mathbf{y} = \tau \mathbf{z}$ with

$$\mathbf{z} = \left(\begin{array}{c} \mathbf{r}\left(\frac{1}{\tau} - (f - \mu)\right) \\ R\left(\frac{1}{\tau} - (f - \mu)\right) - \frac{q}{\tau} \end{array} \right)$$

and express the CU criterion as

$$E(\mathbf{z})'[V(\mathbf{z})]^{-1} E(\mathbf{z}) + \frac{\left[E(u) - Cov(u, \mathbf{z})[V(\mathbf{z})]^{-1} E(\mathbf{z})\right]^2}{\left[V(u) - Cov(u, \mathbf{z})[V(\mathbf{z})]^{-1} Cov(\mathbf{z}, u)\right]}.$$

If we take again the limits

$$\mu \to E(f), \qquad \tau(E(f) - \mu) \to 1, \qquad q/\tau \to -Cov(R, f)$$

then the above expressions are well defined and the criterion converges to zero as $E(\mathbf{z}) \to \mathbf{0}$ and $E(u) \to 0$. Therefore, the CU criterion also goes to zero.

The Jacobian of the moment conditions with respect to (τ, μ, q) is

$$\left(\begin{array}{ccc} E(\mathbf{r})\mu - E(\mathbf{r}f) & E(\mathbf{r})\tau & 0 \\ 0 & -1 & 0 \\ E(R)\mu - E(Rf) & E(R)\tau & -1 \end{array} \right),$$

which has full column rank if and only if

$$E(\mathbf{r})\mu - E(\mathbf{r}f) \neq \mathbf{0}.$$

The Jacobian of the moment conditions is ill-behaved in the special case $Cov(\mathbf{r}, f) = \mathbf{0}$ because the first column of the Jacobian would converge to zero, and the second would have unbounded entries.

3) When we work with the normal equations of the two relevant projections (3), (A6) and the symmetric normalization, we need to study

$$E\left[\begin{array}{c} \left(\begin{array}{c} 1 \\ f \end{array} \right) \otimes (\mathbf{r} - \boldsymbol{\beta}(f + \varkappa)) \\ \left(\begin{array}{c} 1 \\ f \end{array} \right) \otimes (R - \varkappa_R - \beta_R(f + \varkappa)) \end{array} \right].$$

In this context, the true parameters must satisfy

$$E(\mathbf{r}) = \boldsymbol{\beta}(\varkappa + E(f)),$$

$$E(\mathbf{r}f) = \boldsymbol{\beta}\left(\varkappa E(f) + E\left(f^2\right)\right),$$

$$E(R) = \varkappa_R + \beta_R(\varkappa + E(f)),$$

$$E(Rf) = \varkappa_R E(f) + \beta_R\left(\varkappa E(f) + E\left(f^2\right)\right).$$

The last two equations are exactly identified for $(\varkappa_R, \beta_R)$ at a given $\varkappa$ and trivially satisfied with $V(f) > 0$. Therefore, we can focus on the first two blocks, which were studied in point 4) of Lemma G3. There we showed that there are finite values of $(\boldsymbol{\beta}, \varkappa)$ that make the moments equal to zero whenever $Cov(\mathbf{r},f) \neq \mathbf{0}$. In contrast,, when $Cov(\mathbf{r},f) = \mathbf{0}$ then the moments converge to zero if we choose

$$\boldsymbol{\beta} \to \mathbf{0}, \qquad \boldsymbol{\beta}\varkappa \to E(\mathbf{r}).$$

Accordingly, in that case we need to choose

$$\beta_R \to Cov(R, f)/V(f), \qquad \varkappa_R + \beta_R \varkappa \to E(R) - \beta_R E(f)$$

to make the last two moments close to zero simultaneously. This would make the standard GMM criterion close to zero in the population for any positive definite weighting matrix. Once again note the discontinuity of the criterion at $\boldsymbol{\beta} = \mathbf{0}$.

Assuming *i.i.d.* data for simplicity, the CU criterion converges to

$$E\left[\begin{pmatrix} 1 \\ f \end{pmatrix} \otimes \boldsymbol{\varepsilon}\right]' \left[V\left[\begin{pmatrix} 1 \\ f \end{pmatrix} \otimes \boldsymbol{\varepsilon}\right]\right]^{-1} E\left[\begin{pmatrix} 1 \\ f \end{pmatrix} \otimes \boldsymbol{\varepsilon}\right].$$

where

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \mathbf{r} - E(\mathbf{r}) \\ (R - E(R)) - (f - E(f))\, Cov(R, f)/V(f) \end{pmatrix}.$$

This criterion is a quadratic form in a zero vector with a definite positive weighting matrix. Hence the CU criterion also goes to zero.

Finally, the Jacobian with respect to $(\boldsymbol{\beta}, \varkappa, \beta_R, \varkappa_R)$ is

$$\begin{pmatrix} -(\varkappa + E(f))\mathbf{I} & -\boldsymbol{\beta} & \mathbf{0} & \mathbf{0} \\ -\left(\varkappa E(f) + E\left(f^2\right)\right)\mathbf{I} & -\boldsymbol{\beta}E(f) & \mathbf{0} & \mathbf{0} \\ 0 & -\beta_R & -(\varkappa + E(f)) & -1 \\ 0 & -\beta_R E(f) & -\left(\varkappa E(f) + E\left(f^2\right)\right) & -E(f) \end{pmatrix},$$

which has full column rank if and only if $\boldsymbol{\beta} \neq \mathbf{0}$, the same condition as in point 4) of Lemma G3.

4) When we work with (1) and (A4), we need to study

$$E\begin{bmatrix} \mathbf{r}(a + bf) \\ R(a + bf) - 1 \end{bmatrix}.$$

We know that (1) holds along a particular direction in (a, b) space, which means that there is a vector (a, b) that makes the moments above equal to zero if and only if $aE(R) + bE(Rf) \neq 0$ along the same direction. Moreover, it is straightforward to prove that if this last condition is satisfied, the Jacobian with respect to (a, b)

$$\begin{pmatrix} E(\mathbf{r}) & E(\mathbf{r}f) \\ E(R) & E(Rf) \end{pmatrix}$$

will have full column rank.

On the other hand, when $aE(R) + bE(Rf) = 0$ along the direction in (a, b) space which satisfies (1), then the standard GMM criterion cannot be equal to zero in the population. However, the CU criterion is invariant to reparameterizing and rescaling the influence functions. Specifically, we can divide the influence functions by the norm of (a, b), which is always positive, and thus obtain the symmetrically normalized influence functions in point 1) of Lemma G6

$$\begin{bmatrix} \mathbf{r}(\sin\psi + f\cos\psi) \\ R(\sin\psi + f\cos\psi) - q, \end{bmatrix},$$

with $q = \left(a^2 + b^2\right)^{-1/2}$. The corresponding criterion is zero at $q = 0$ but its Jacobian is not singular at that value. Of course, this is also true for multi-step GMM based on those influence functions parameterized in terms of (ψ, q), but standard GMM is not invariant to rescaling the influence functions, and hence the connection to the original influence functions in terms of (a, b) is lost.

5) When we work with (2) and (A5), we need to study

$$E\begin{bmatrix} \mathbf{r}(c + b(f - \mu)) \\ f - \mu \\ R(a + b(f - \mu)) - 1 \end{bmatrix}.$$

The second moment is trivially zero at $\mu = E(f)$. In addition, we know that (2) holds along a particular direction in (c, b) space because the linear factor pricing model holds for $\mathbf{r}$. Therefore, there is a vector (c, b) that satisfies the pricing conditions if and only if $cE(R) + bCov(R,f) \neq 0$, which is equivalent to the condition $aE(R) + bE(Rf) \neq 0$ in point 4) above.

In fact, the Jacobian with respect to (c, b, μ) is

$$\begin{pmatrix} E(\mathbf{r}) & E(\mathbf{r}(f - \mu)) & -bE(\mathbf{r}) \\ 0 & 0 & -1 \\ E(R) & E(R(f - \mu)) & -bE(R) \end{pmatrix},$$

and hence the only possible rank failure is that the first two columns have rank one. That rank depends on the last row since the remaining rows have rank one when (2) holds. Therefore, the Jacobian has full column rank if and only if the condition $cE(R)+bCov(R,f)\neq 0$ holds.

On the other hand, when $cE(R)+bCov(R,f)=0$ along the same direction in (c,b) space as (2) holds, then a standard GMM criterion cannot be equal to zero in the population. However, the CU criterion is invariant to reparameterizations and rescaling of the influence functions. Hence, we can divide the influence functions by the norm of (c,b), which must be different from zero, and thus obtain the symmetrically normalized influence functions in point 2) of Lemma G6

$$\begin{bmatrix} \mathbf{r}(\sin\upsilon+(f-\mu)\cos\upsilon) \\ f-\mu \\ R(\sin\upsilon+(f-\mu)\cos\upsilon)-q \end{bmatrix}$$

with $q=\left(c^2+b^2\right)^{-1/2}$. The corresponding criterion is zero at $q=0$, and its Jacobian is not singular. Of course, this is also true for the standard GMM based on those influence functions if we work in terms of (υ,μ,q), but standard GMM is not invariant to rescaling the influence functions, and hence the connection to the original influence functions in terms of (c,b,μ) is lost.

□

References

Antoine, A., H. Bonnal and E. Renault, “On the efficient use of the informational content of estimating equations: Implied probabilities and Euclidean empirical likelihood”, *Journal of Econometrics* 138 (2007), 461–487.

de Jong, R.M. and J. Davidson, “Consistency of kernel estimators of heteroskedasticity and autocorrelation covariance matrices”, *Econometrica* 68 (2000), 407–423.

Hansen, L.P., “Large sample properties of generalized method of moments estimators”, *Econometrica* 50 (1982), 1029–1054.

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

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

Householder, A.S., *The Theory of Matrices in Numerical Analysis*, (London: Blaisdell Publishing Co., 1964).

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

Newey, W.K. and R.J. Smith, “Higher order properties of GMM and generalized empirical likelihood estimators”, *Econometrica* 72 (2004), 219–255.

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

Table C1: Rejection rates in the orthogonal factor design (T=50)

	Nominal size		
	10	5	1
<u>J tests</u>			
CU	11.64	5.12	0.64
Uncentred SDF - Symmetric normalization			
Iterated	27.90	18.94	7.53
2S	21.89	13.57	4.88
Centred SDF - Symmetric normalization			
Iterated	15.62	8.42	1.96
2S	22.51	14.04	4.68
Uncentred SDF - Asymmetric normalization			
Iterated	70.61	60.56	39.59
2S	68.35	58.84	39.16
Centred SDF - Asymmetric normalization			
Iterated	50.08	40.88	25.93
2S	57.07	47.09	28.73
Centred regression - Asymmetric normalization			
Iterated	13.26	6.58	1.33
2S	15.93	8.23	1.84
<u>CU DM tests of problematic cases</u>			
Uncorrelated f	66.97	57.81	37.54
Orthogonal f	29.30	20.63	8.87

Note: This table displays the rejection rates of the J tests of each method by continuously updated (CU), iterated and two-step (2S) GMM. The rates are shown in percentage for the asymptotic critical values at 10, 5 and 1%. The table also displays the CU DM tests of an uncorrelated factor and an orthogonal factor. 10000 samples of 8 excess returns are simulated under the uncorrelated factor design. The only change with respect to the baseline design is in the mean of f to obtain an orthogonal factor.

Table G1: Implied estimates of the CAPM

	CU	Iterated			2S		
		U. SDF	C. SDF	Reg	U. SDF	C. SDF	Reg
<u>Uncentred SDF - Asymmetric normalization</u> (δ)							
δ	4.826	4.534	2.606	4.330	4.455	2.290	3.616
Pricing error E(**r**)-E(**r**f)δ							
1	-0.568	-0.675	-1.381	-0.750	-0.704	-1.497	-1.011
2	-0.172	-0.214	-0.495	-0.244	-0.226	-0.540	-0.348
3	-0.325	-0.350	-0.519	-0.368	-0.357	-0.547	-0.431
4	1.991	1.890	1.226	1.820	1.863	1.117	1.574
5	-0.099	-0.102	-0.123	-0.104	-0.103	-0.126	-0.112
6	0.039	0.024	-0.077	0.013	0.020	-0.093	-0.024
7	2.827	2.836	2.901	2.843	2.839	2.911	2.867
8	0.931	0.997	1.437	1.044	1.015	1.509	1.207
f	-10.292	-9.245	-2.356	-8.519	-8.964	-1.226	-5.965
<u>Centred SDF - Asymmetric normalization</u> (τ)							
τ	16.945	14.258	3.290	11.240	11.633	2.724	6.292
Pricing error E(**r**)-E(**r**(f-μ))τ							
1	-1.994	-2.124	-1.744	-1.946	-1.839	-1.781	-1.760
2	-0.604	-0.675	-0.624	-0.633	-0.590	-0.643	-0.605
3	-1.14	-1.102	-0.655	-0.956	-0.933	-0.651	-0.750
4	6.989	5.944	1.548	4.724	4.864	1.329	2.739
5	-0.347	-0.321	-0.155	-0.271	-0.269	-0.150	-0.195
6	0.137	0.075	-0.097	0.034	0.051	-0.111	-0.042
7	9.924	8.921	3.662	7.380	7.414	3.464	4.989
8	3.267	3.137	1.814	2.709	2.651	1.796	2.100
f	-36.133	-29.077	-2.974	-22.113	-23.409	-1.459	-10.38
<u>Regression</u> (μ)							
μ	0.148	0.150	0.080	0.142	0.139	0.070	0.118
Pricing error E(**r**)-**β**E(f)							
1	-1.283	-1.347	-2.157	-1.686	-1.666	-2.196	-1.394
2	-0.434	-0.442	-0.658	-0.545	-0.522	-0.687	-0.498
3	-0.728	-0.759	-0.939	-0.941	-0.833	-0.928	-0.618
4	0.437	0.428	1.138	0.380	0.507	1.225	0.548
5	-0.309	-0.308	-0.400	-0.447	-0.411	-0.414	-0.230
6	-0.347	-0.375	-0.182	-0.490	-0.404	-0.111	-0.309
7	2.096	2.159	3.170	2.364	2.382	3.272	2.283
8	0.956	0.865	1.753	0.467	0.766	1.956	0.911
f	-0.079	-0.081	-0.010	-0.072	-0.069	0.000	-0.048

Note: This table displays the implied parameter estimates (δ, τ or μ) and pricing errors in percentage for the 8 Lustig-Verdelhan currency portfolios and the market factor. For the regression method, we define the last pricing error as E(f)-μ. We implement each method by continuously updated (CU), iterated and two-step (2S) GMM. The implied computations are as follows:

1) The uncentred SDF moments (1) and (5) in terms of δ are extended with the estimation of γ, which is exactly identified by the moment condition $E(f^2-\gamma)=0$. The implied parameters are computed as $\mu=\gamma\delta$ and $\tau=\mu/(\gamma-\mu^2)$. We also need an estimate of **β** to compute pricing errors like in the regression approach, which requires the addition of the exactly identified moment conditions $E[(\mathbf{r}-\boldsymbol{\beta}f)f]=\mathbf{0}$.

2) The centred SDF moments (2) and (7) in terms of τ and μ can be used to compute the implied $\delta=\tau/(1+\tau\mu)$. We also need an estimate of **β** to compute pricing errors like in the regression approach, and we can proceed as in 1).

3) The centred regression moments (9), after adding the estimation of μ to the estimation of **β**, are extended with the estimation of γ, which is estimated as before. The implied parameters are computed as $\delta=\mu/\gamma$ and $\tau=\mu/(\gamma-\mu^2)$.

Table G2: Implied estimates of the CCAPM

	CU	Iterated			2S		
		U. SDF	C. SDF	Reg	U. SDF	C. SDF	Reg
<u>Uncentred SDF - Asymmetric normalization</u> (δ)							
δ	49.507	48.835	41.129	49.610	48.850	52.999	43.938
Pricing error E(**r**)-E(**r**f)δ							
1	-0.645	-0.668	-0.931	-0.641	-0.667	-0.526	-0.835
2	-0.994	-0.992	-0.974	-0.994	-0.992	-1.003	-0.980
3	-0.446	-0.450	-0.497	-0.445	-0.450	-0.425	-0.480
4	-0.114	-0.108	-0.039	-0.115	-0.108	-0.145	-0.064
5	-0.696	-0.688	-0.604	-0.697	-0.689	-0.734	-0.635
6	-0.320	-0.319	-0.302	-0.320	-0.319	-0.328	-0.308
7	-0.462	-0.415	0.122	-0.469	-0.416	-0.705	-0.074
8	0.377	0.399	0.657	0.373	0.399	0.260	0.563
<u>Centred SDF - Asymmetric normalization</u> (τ)							
τ	438.769	409.159	115.428	447.377	409.908	120.114	183.445
Pricing error E(**r**)-E(**r**(f-μ))τ							
1	-5.716	-5.596	-2.613	-5.784	-5.600	-1.191	-3.487
2	-8.810	-8.315	-2.732	-8.966	-8.328	-2.272	-4.093
3	-3.953	-3.771	-1.395	-4.016	-3.776	-0.963	-2.004
4	-1.012	-0.907	-0.110	-1.038	-0.909	-0.330	-0.269
5	-6.167	-5.768	-1.694	-6.285	-5.778	-1.664	-2.649
6	-2.839	-2.671	-0.848	-2.890	-2.675	-0.743	-1.287
7	-4.091	-3.476	0.343	-4.228	-3.490	-1.598	-0.307
8	3.340	3.346	1.843	3.367	3.347	0.590	2.350
<u>Centred regression - Asymmetric normalization</u> (λ)							
λ	0.056	0.053	0.028	0.056	0.053	0.025	0.024
Pricing error E(**r**)-**β**λ							
1	-0.368	-0.389	-1.575	-0.324	-0.260	-1.212	-0.799
2	-0.900	-0.921	-1.650	-0.828	-0.927	-1.804	-0.972
3	-0.935	-0.935	-1.270	-0.894	-0.970	-2.085	-1.153
4	-0.568	-0.577	-0.909	-0.526	-0.657	-1.414	-0.773
5	-0.440	-0.443	-0.887	-0.402	-0.463	-1.075	-0.709
6	-0.732	-0.731	-0.676	-0.707	-0.779	-1.450	-0.750
7	-0.743	-0.738	0.043	-0.675	-1.002	0.135	-0.828
8	-1.170	-1.114	0.100	-1.208	-1.362	-1.883	-0.755

Note: This table displays the implied parameter estimates (δ, τ or λ) and pricing errors in percentage for the 8 Lustig-Verdelhan currency portfolios. We implement each method by continuously updated (CU), iterated and two-step (2S) GMM. The implied computations are as follows:

1) The uncentred SDF moments (1) in terms of δ are extended with the estimation of μ and γ, which are exactly identified by the moment conditions $E(f-\mu)=0$ and $E(f^2-\gamma)=0$ respectively. The implied parameters are computed as $\tau=\delta/(1-\delta\mu)$ and $\lambda=\tau(\gamma-\mu^2)$.) We also need an estimate of **β** to compute pricing errors like in the regression approach, which requires the addition of the exactly identified moment conditions $E[(\mathbf{r}-\boldsymbol{\beta}(f-\mu+\lambda))f]=\mathbf{0}$.

2) The centred SDF moments (2) in terms of τ and μ are extended with the estimation γ, which is exactly identified by the moment condition $E(f^2-\gamma)=0$. The implied parameters are computed as $\lambda=\tau(\gamma-\mu^2)$ and $\delta=\lambda/(\gamma-\mu^2+\lambda\mu)$. We also need an estimate of **β** to compute pricing errors like in the regression approach, and we can proceed as in 1).

3) The centred regression moments (3) in terms of λ, μ and **β** are extended with the estimation of γ, which is exactly identified by the moment condition $E(f^2-\gamma)=0$. The implied parameters are computed as $\delta=\lambda/(\gamma-\mu^2+\lambda\mu)$ and $\tau=\lambda/(\gamma-\mu^2)$.

Table G3: Rejection rates in the baseline design (T=500)

	Nominal size		
	10	5	1
J tests			
CU	10.88	5.76	1.08
Uncentred SDF - Symmetric normalization			
Iterated	10.94	5.76	1.08
2S	10.91	5.76	1.08
Centred SDF - Symmetric normalization			
Iterated	11.08	5.84	1.11
2S	11.71	6.38	1.33
Uncentred SDF - Asymmetric normalization			
Iterated	10.99	5.85	1.11
2S	11.04	5.84	1.15
Centred SDF - Asymmetric normalization			
Iterated	18.01	11.55	4.47
2S	26.07	18.17	8.17
Centred regression - Asymmetric normalization			
Iterated	10.88	5.76	1.08
2S	10.87	5.75	1.08
CU DM tests of problematic cases			
Uncorrelated f	100	100	100

Note: This table displays the rejection rates of the J tests of each method by continuously updated (CU), iterated and two-step (2S) GMM. The rates are shown in percentage for the asymptotic critical values at 10, 5 and 1%. The table also displays the CU DM test of an uncorrelated factor. 10000 samples of 8 excess returns are simulated under the baseline design. The mean and the standard deviation of f are 1; the maximum Sharpe ratio achievable with **r** is 0.5; the R^2 of the regression of f on **r** is 0.1; all the underlying random variables are independent and identically distributed over time as multivariate Gaussian vectors.

Table G4: Rejection rates in the uncorrelated factor design (T=500)

	Nominal size		
	10	5	1
J tests			
CU	11.05	5.39	0.95
Uncentred SDF - Symmetric normalization			
Iterated	12.05	6.33	1.41
2S	11.64	5.98	1.25
Centred SDF - Symmetric normalization			
Iterated	11.07	5.43	0.96
2S	11.14	5.50	1.00
Uncentred SDF - Asymmetric normalization			
Iterated	15.11	8.82	2.42
2S	13.37	7.48	1.88
Centred SDF - Asymmetric normalization			
Iterated	84.71	79.98	69.18
2S	84.84	77.23	57.49
Centred regression - Asymmetric normalization			
Iterated	11.05	5.40	0.95
2S	10.78	5.42	0.96
CU DM tests of problematic cases			
Uncorrelated f	11.61	6.00	1.28

Note: This table displays the rejection rates of the J tests of each method by continuously updated (CU), iterated and two-step (2S) GMM. The rates are shown in percentage for the asymptotic critical values at 10, 5 and 1%. The table also displays the CU DM tests of an uncorrelated factor. 10000 samples of 8 excess returns are simulated under the uncorrelated factor design. The only change with respect to the baseline design is a reduction of the R^2 of the regression of f on **r** to 0.

Table G5: Rejection rates in the missing factor design (T=500)

	Asymptotic critical values			Monte Carlo critical values		
	10	5	1	10	5	1
CU	90.25	83.81	62.95	89.62	82.13	61.99
Uncentred SDF - Symmetric normalization						
Iterated	90.26	83.84	63.03	89.62	82.16	62.05
2S	90.27	83.90	63.17	89.63	82.17	62.10
Centred SDF - Symmetric normalization						
Iterated	90.35	84.06	63.74	89.59	82.11	61.93
2S	90.98	84.93	65.99	89.68	82.15	61.82
Uncentred SDF - Asymmetric normalization						
Iterated	90.45	84.18	63.78	89.73	82.33	62.43
2S	90.50	84.31	63.98	89.71	82.39	62.09
Centred SDF - Asymmetric normalization						
Iterated	94.97	91.79	82.83	90.80	83.85	64.58
2S	97.03	95.08	88.93	90.83	83.80	64.92
Centred regression - Asymmetric normalization						
Iterated	90.25	83.82	62.96	89.62	82.14	61.99
2S	90.23	83.80	62.99	89.63	82.15	62.09

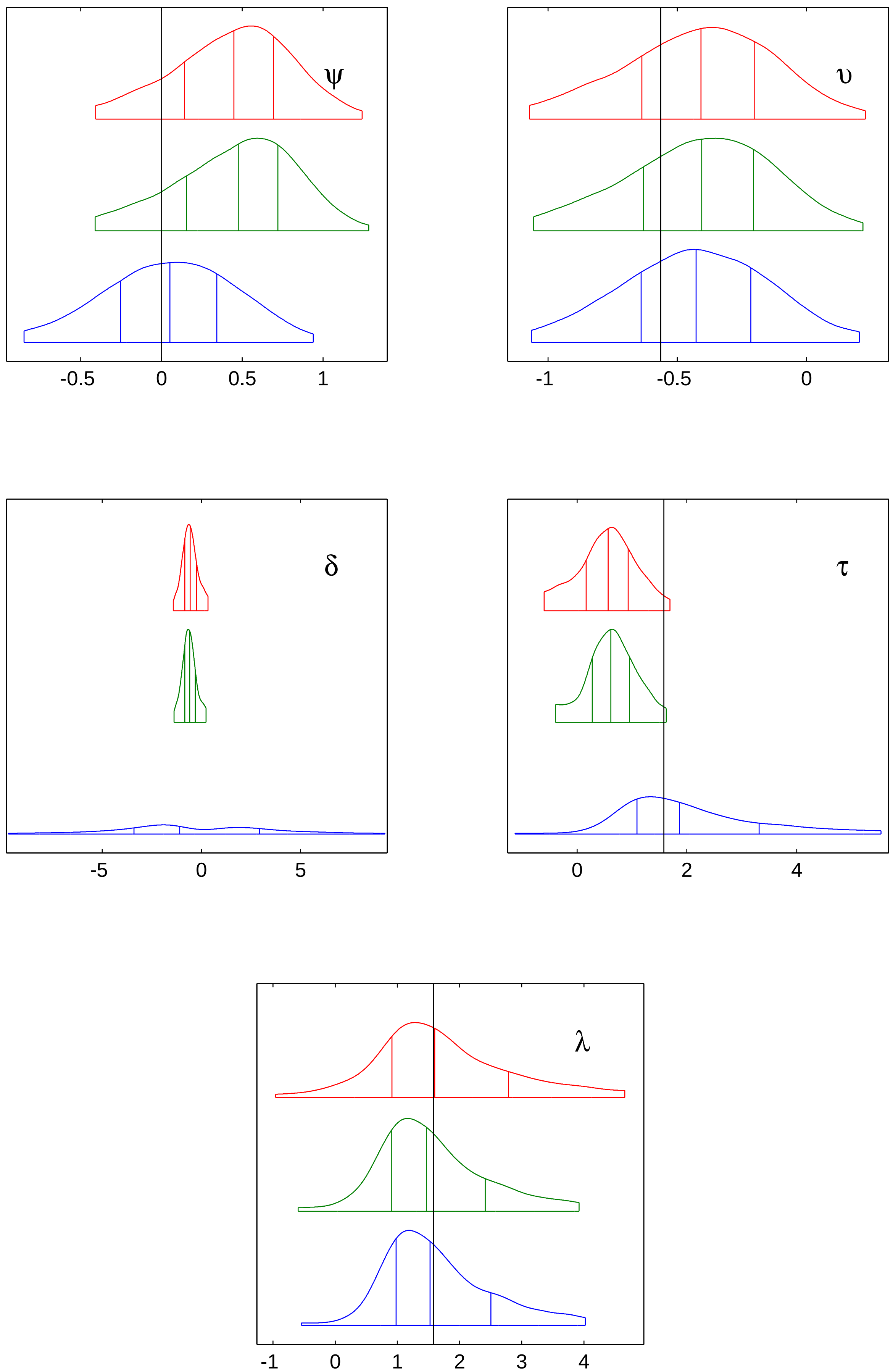
Note: This table displays the rejection rates of the J tests of each method by continuously updated (CU), iterated and two-step (2S) GMM. The rates are shown in percentage for the asymptotic and Monte Carlo critical values at 10, 5 and 1%. 10000 samples of 8 excess returns are simulated under the missing factor design. The only change with respect to the baseline design is an increase in the Hansen-Jagannathan distance to 0.2.

Table G6: Rejection rates in the orthogonal factor design (T=500)

	Nominal size		
	10	5	1
<u>J tests</u>			
CU	10.88	5.76	1.08
Uncentred SDF - Symmetric normalization			
Iterated	12.23	6.85	1.56
2S	11.32	6.17	1.25
Centred SDF - Symmetric normalization			
Iterated	11.08	5.84	1.11
2S	11.71	6.38	1.33
Uncentred SDF - Asymmetric normalization			
Iterated	89.35	87.01	82.10
2S	85.43	82.52	76.65
Centred SDF - Asymmetric normalization			
Iterated	18.01	11.55	4.47
2S	26.07	18.17	8.17
Centred regression - Asymmetric normalization			
Iterated	10.88	5.76	1.08
2S	10.31	5.36	1.03
<u>CU DM tests of problematic cases</u>			
Uncorrelated f	100	100	100
Orthogonal f	11.21	5.90	1.20

Note: This table displays the rejection rates of the J tests of each method by continuously updated (CU), iterated and two-step (2S) GMM. The rates are shown in percentage for the asymptotic critical values at 10, 5 and 1%. The table also displays the CU DM test s of an uncorrelated factor and an orthogonal factor. 10000 samples of 8 excess returns are simulated under the uncorrelated factor design. The only change with respect to the baseline design is in the mean of f to obtain an orthogonal factor.

Figure C1: Parameter estimates in orthogonal factor design (T=50)



Note: These bicorne plots combine a kernel density estimate on top of a box plot. The vertical lines describe the median and the first and third quartiles, while the length of the tails is one interquartile range. The common vertical line, if any, indicates the true parameter value. Two step, iterated and continuously updated GMM are presented in the top, middle and bottom, respectively, of each plot.

Figure G1: CU-GMM criterion functions (scaled by T)

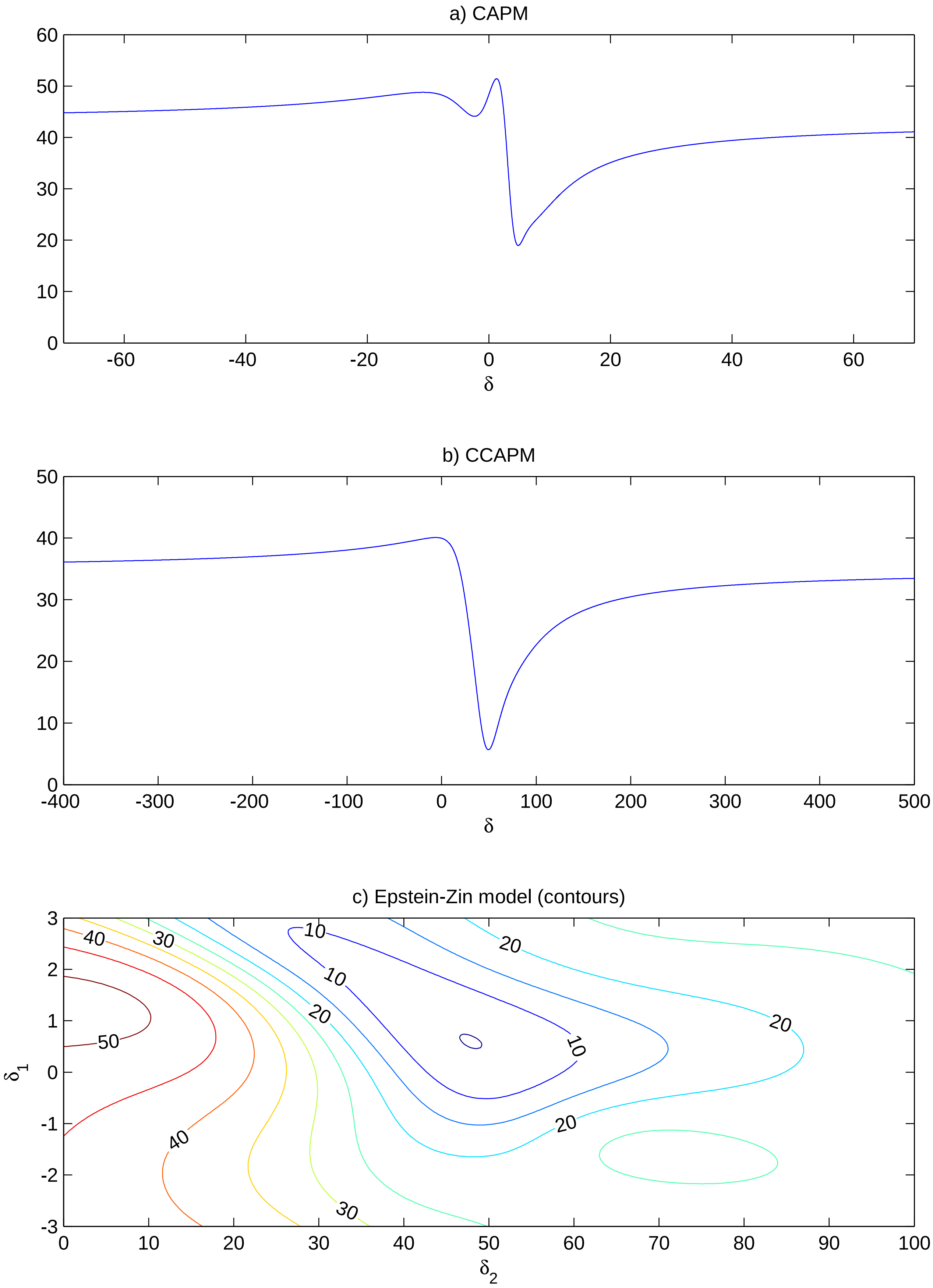
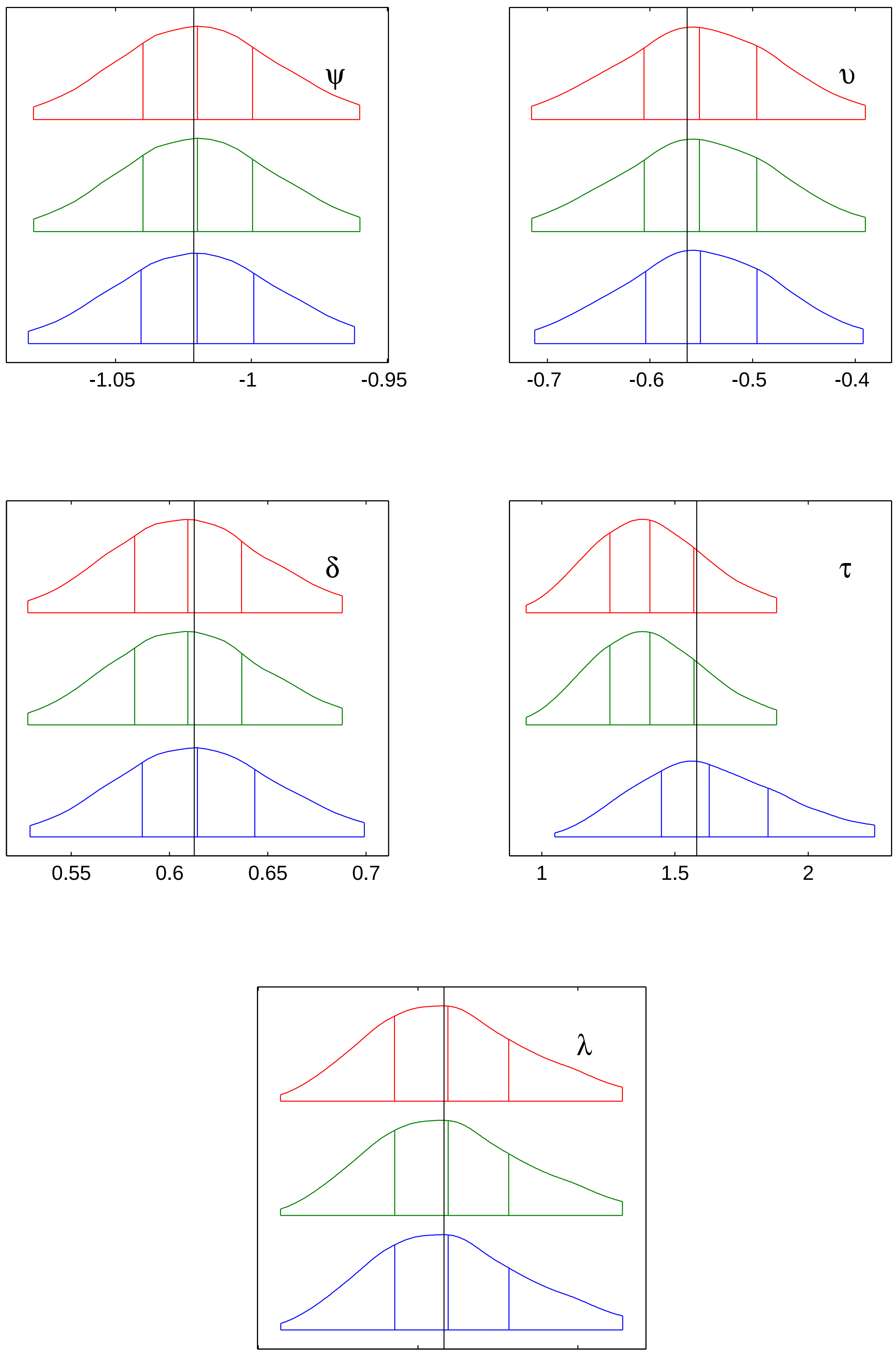
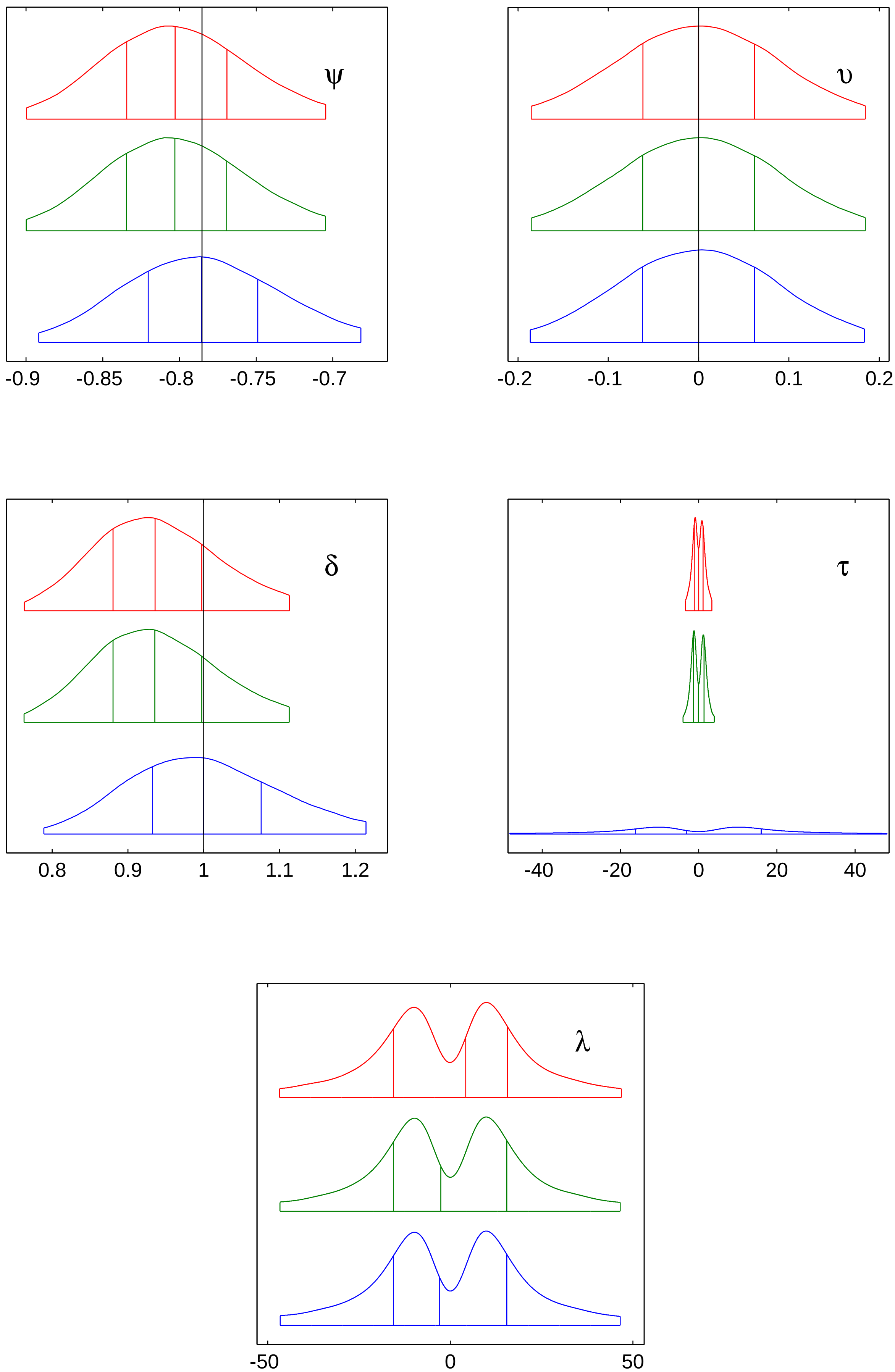


Figure G2: Parameter estimates in baseline design (T=500)



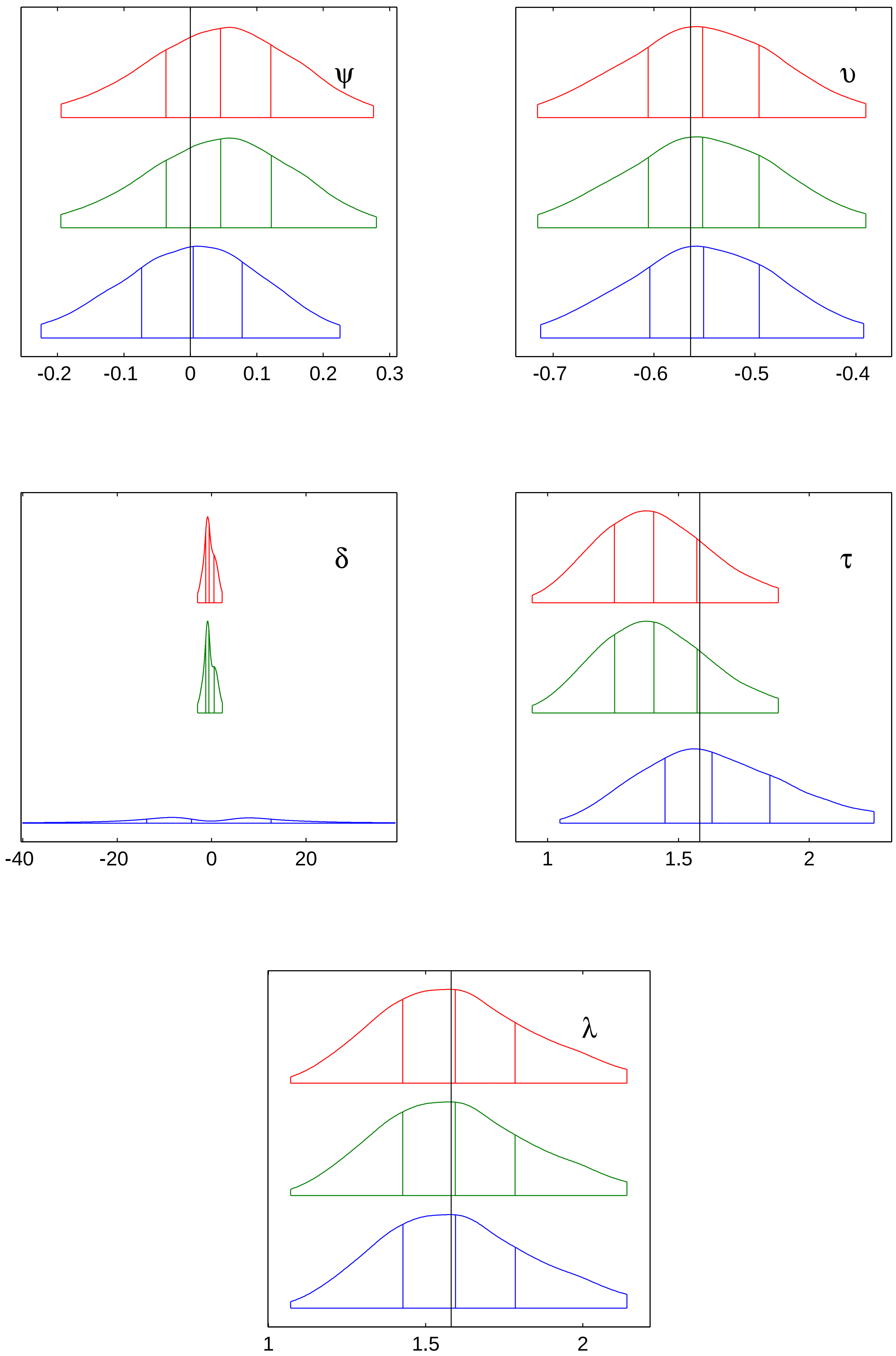
Note: These bicorne plots combine a kernel density estimate on top of a box plot. The vertical lines describe the median and the first and third quartiles, while the length of the tails is one interquartile range. The common vertical line, if any, indicates the true parameter value. Two step, iterated and continuously updated GMM are presented in the top, middle and bottom, respectively, of each plot.

Figure G3: Parameter estimates in uncorrelated factor design (T=500)



Note: These bicorne plots combine a kernel density estimate on top of a box plot. The vertical lines describe the median and the first and third quartiles, while the length of the tails is one interquartile range. The common vertical line, if any, indicates the true parameter value. Two step, iterated and continuously updated GMM are presented in the top, middle and bottom, respectively, of each plot.

Figure G4: Parameter estimates in orthogonal factor design (T=500)



Note: These bicorne plots combine a kernel density estimate on top of a box plot. The vertical lines describe the median and the first and third quartiles, while the length of the tails is one interquartile range. The common vertical line, if any, indicates the true parameter value. Two step, iterated and continuously updated GMM are presented in the top, middle and bottom, respectively, of each plot.