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

Abstract

Regression and SDF approaches with centred or uncentred moments and symmetric or asymmetric normalizations are commonly used to empirically evaluate linear factor pricing models. We show that unlike two-step or iterated GMM procedures, single-step estimators such as continuously updated GMM yield numerically identical risk prices, pricing errors and overidentifying restrictions tests irrespective of the model validity and regardless of the factors being traded, or the use of excess or gross returns. We illustrate our results with Lustig and Verdelhan's (2007) currency returns, propose tests to detect some problematic cases and provide Monte Carlo evidence on the reliability of asymptotic approximations.

Keywords: CU-GMM, Factor pricing models, Forward premium puzzle, Generalized Empirical Likelihood, Stochastic discount factor.

JEL: G12, G15, C12, C13.

*We would like to thank Abhay Abhyankar, Manuel Arellano, Craig Burnside, Antonio Díez de los Ríos, Prosper Dovonon, Lars Hansen, Raymond Kan, Craig MacKinlay, Cesare Robotti, Rosa Rodríguez, Amir Yaron, participants at the Finance Forum (Elche, 2010), SAEe (Madrid, 2010), ESEM (Oslo, 2011), FMG 25th Anniversary Conference (London, 2012), MFA (New Orleans, 2012), as well as audiences at the Atlanta Fed, Bank of Canada, Banque de France, Duke, Edinburgh, Fuqua, Geneva, Kenan-Flagler, Málaga, Montreal, Princeton, St.Andrews, UPF, Warwick and Wharton for helpful comments, suggestions and discussions. The comments from an associate editor and an anonymous referee have also led to a substantially improved paper. Felipe Carozzi and Luca Repetto provided able research assistance for the Monte Carlo simulations. Of course, the usual caveat applies. Financial support from the Spanish Ministry of Science and Innovation through grants ECO 2008-03066 and 2011-25607 (Peñaranda) and ECO 2008-00280 and 2011-26342 (Sentana) is gratefully acknowledged.

1 Introduction

Asset pricing theories are concerned with determining the expected returns of assets whose payoffs are risky. Specifically, these models analyze the relationship between risk and expected returns, and address the crucial question of how to value risk. The most popular empirically oriented asset pricing models effectively assume the existence of a common stochastic discount factor (SDF) that is linear in some risk factors, which discounts uncertain payoffs differently across different states of the world. Those factors can be either the excess returns on some traded securities, as in the traditional CAPM of Sharpe (1964), Lintner (1965) and Mossin (1966) or the so-called Fama and French (1993) model, non-traded economy wide sources of uncertainty related to macroeconomic variables, like in the Consumption CAPM (CCAPM) of Breeden (1979), Lucas (1978) or Rubinstein (1976), or a combination of the two, as in the exact version of Ross' (1976) APT.

There are two main approaches to formally evaluate linear factor pricing models from an empirical point of view using optimal inference procedures. The traditional method relies on regressions of excess returns on factors, and exploits the fact that an asset pricing model imposes certain testable constraints on the relationship between slopes and intercepts. More recent methods rely on the SDF representation of the model instead, and exploit the fact that the corresponding pricing errors should be zero. There are in fact two variants of the SDF method, one that demeans the factors (the "centred" version) and another one that does not (the "uncentred" one), and one can envisage analogous variants of the regression approach, although only the "centred" one has been used so far in empirical work.

The initial asset pricing tests tended to make the assumption that asset returns and factors were independently and identically distributed as a multivariate normal vector. Nowadays, empirical researchers rely on the generalized method of moments (GMM) of Hansen (1982), which has the advantage of yielding asymptotically valid inferences even if the assumptions of serial independence, conditional homoskedasticity or normality are not totally realistic in practice (see Campbell, Lo and MacKinlay (1996) or Cochrane (2001a) for textbook treatments).

Unfortunately, though, each approach (and their multiple variants) typically yields different estimates of prices of risk and pricing errors, and different values for the overidentifying restrictions test. This begs the question of which approach is best, and there has been some controversy surrounding the answer. For example, Kan and Zhou (1999) advocated the use of the regression method over the uncentred SDF method because the former provides more reliable risk premia estimators and more powerful pricing tests than the latter. However, Cochrane (2001b) and Jagannathan and Wang (2002) criticized their conclusions on the grounds that they did not con-

sider the estimation of factor means and variances. Specifically, Jagannathan and Wang (2002) showed that if the excess returns and the factor are jointly distributed as an *iid* multivariate normal random vector, in which case the regression approach is optimal, the (uncentred) SDF approach is asymptotically equivalent under the null. Kan and Zhou (2002) acknowledged this equivalence result, and extended it to compatible sequences of local alternatives under weaker distributional assumptions.

More recently, Burnside (2012) and Kan and Robotti (2008) have also pointed out that in certain cases there may be dramatic differences between the results obtained by applying standard two-step or iterated GMM procedures to the centred and uncentred versions of the SDF approach. Moreover, Kan and Robotti (2008, footnote 3) effectively exploit the invariance to coefficient normalizations of the continuously updated GMM estimator (CU-GMM) of Hansen, Heaton and Yaron (1996) to prove the numerical equivalence of the overidentification tests associated to the centred and uncentred versions of the SDF approach. As is well known, CU-GMM is a single-step method that integrates the heteroskedasticity and autocorrelation consistent (HAC) estimator of the long-run covariance matrix in the objective function.

In this context, the main contribution of our paper is to show the more subtle result that in finite samples the application to both the regression and SDF approaches of single-step GMM methods, including CU-GMM, gives rise to numerically identical estimates of prices of risk, pricing errors and overidentifying restrictions tests irrespective of the validity of the asset pricing model and regardless of whether one uses centred or uncentred moments and symmetric or asymmetric normalizations. We also show that the empirical evidence in favour or against a pricing model is not affected by the addition of an asset with non-zero cost that pins down the scale of the SDF if one uses single step methods, unlike what may happen with multistep methods.

Therefore, one could argue that in effect, there is only one optimal GMM procedure to empirically evaluate asset-pricing models. Although the rationale for our results is the well-known functional invariance of maximum likelihood estimators, their validity does not depend on any distributional assumption, the number of assets, the specific combination of traded and non-traded factors, and remain true regardless of whether or not the researcher works with excess returns or gross returns. For ease of exposition, we centre most of our discussion on models with a single priced factor. Nevertheless, our numerical equivalence results do not depend in any way on this simplification. In fact, the proofs of our main results explicitly consider the general multifactor case.

Another relevant issue that arises with asset pricing tests is that the moment conditions are

sometimes compatible with SDFs which are affine functions of risk factors that are uncorrelated or orthogonal to the vector of excess returns. To detect such cases, which are unattractive from an economic point of view, we provide a battery of distance metric tests that empirical researchers should systematically report in addition to the J test.

We would like to emphasize that our results apply to optimal GMM inference procedures. In particular, we do not consider sequential GMM methods that fix the factor means to their sample counterparts. We do not consider either procedures that use alternative weighting matrices such as the uncentred second moment of returns chosen by Hansen and Jagannathan (1997) or the popular two-pass regressions. Those generally suboptimal GMM estimators fall outside the realm of single-step methods, and therefore they would typically give rise to numerically different statistics.

While single-step methods are not widespread in empirical finance applications, this situation is likely to change in the future, as the recent papers by Almeida and Garcia (2012), Bansal, Kiku and Yaron (2012), Campbell, Gilgio and Polk (2012) or Julliard and Gosh (2012) attest. There are several reasons for their increasing popularity. First, like traditional likelihood methods, these modern GMM variants substantially reduce the leeway of the empirical researcher to choose among the surprisingly large number of different ways of writing, parameterizing and normalizing the asset pricing moment conditions, which also avoids problematic cases.

More importantly, single step GMM implementations often yield more reliable inferences in finite samples than two step or iterated methods (see Hansen, Heaton and Yaron (1996)). Such Monte Carlo evidence is confirmed by Newey and Smith (2004), who highlight the finite sample advantages of CU and other generalized empirical likelihood estimators over two-step GMM by going beyond the usual first-order asymptotic equivalence results. As we shall see below, our own simulation evidence reinforces those conclusions.

However, the CU-GMM estimator and other single-step, generalized empirical likelihood (GEL) estimators, such as empirical likelihood or exponentially-tilted methods, are often more difficult to compute than two-step estimators, particularly in linear models, and they may sometimes give rise to multiple local minima and extreme results. Although we explain in Peñaranda and Sentana (2012) how to compute CU-GMM estimators by means of a sequence of OLS regressions, here we derive simple, intuitive consistent parameter estimators that can be used to obtain good initial values, and which will be efficient for elliptically distributed returns and factors. Interestingly, we can also show that these consistent estimators coincide with the GMM estimators recommended by Hansen and Jagannathan (1997), which use the second moment of returns as weighting matrix. In addition, we suggest the imposition of good deal restrictions

(see Cochrane and Saa-Requejo (2000)) that rule out implausible results.

We illustrate our results by using the currency portfolios constructed by Lustig and Verdelhan (2007) to assess some popular linear factor pricing models: the CAPM and linearized versions of the Consumption CAPM, including the Epstein and Zin (1989) model in appendix A. Our findings confirm that the conflict among criteria for testing asset pricing models that we have previously mentioned is not only a theoretical possibility, but a hard reality. Nevertheless, such a conflict disappears when one uses single-step methods. At the same time, our results confirm Burnside's (2011) findings that US consumption growth seems to be poorly correlated to currency returns. This fact could explain the discrepancies between the different two-step and iterated procedures that we find because non-traded factors that are uncorrelated with excess returns will automatically price those returns with a SDF whose mean is 0. Such a SDF is not very satisfactory, but strictly speaking, the vector of risk premia and the covariances between excess returns and factors belong to the same one-dimensional linear space. On the other hand, lack of correlation between factors and returns is not an issue when all the factors are traded, as long as they are part of the set of returns to be priced. In this sense, our empirical results indicate that the rejection of the CAPM that we find disappears when we do not attempt to price the market.

The rest of the paper is organized as follows. Section 2 provides the theoretical background for the centred and uncentred variants of the SDF and regression approaches that only consider excess returns. We then study in more detail SDFs with traded and non-traded factors in sections 3 and 4, respectively. We report the results of the empirical application to currency returns in section 5 and the simulation evidence in section 6. Finally, we summarize our conclusions and discuss some avenues for further research in section 7. Extensions to situations in which the SDF combines both traded and non-traded factors, or a gross return is added to the data at hand, are relegated to appendix A, while appendix B contains the proofs of our main results. We also include a supplemental appendix that discusses a model with an orthogonal factor, describes the Monte Carlo design, and contains a brief description of multifactor models and CU-GMM, together with some additional results.

2 Theoretical background

2.1 The SDF approach

Let $\mathbf{r}$ be an $n \times 1$ vector of excess returns, whose means we assume are not all equal to zero. Standard arguments such as lack of arbitrage opportunities or the first order conditions

of a representative investor imply that

$$E\left(m\mathbf{r}\right)=\mathbf{0}$$

for some random variable m called SDF, which discounts uncertain payoffs in such a way that their expected discounted value equals their cost.

The standard approach in empirical finance is to model m as an affine transformation of some risk factors, even though this ignores that m must be positive with probability 1 to avoid arbitrage opportunities (see Hansen and Jagannathan (1991)). With a single risky factor f, we can express the pricing equation as

$$E\left[\left(a+bf\right)\mathbf{r}\right]=\mathbf{0} \tag{1}$$

for some real numbers (a,b), which we can refer to as the intercept and slope of the affine SDF $a+bf$. For each asset i, the corresponding equation

$$E\left[\left(a+bf\right)r_i\right]=0,\quad (i=1,\ldots,n)$$

defines a straight line in (a,b) space. If asset markets were completely segmented, in the sense that the same source of risk is priced differently for different assets (see e.g. Stulz (1995)), those straight lines would be asset specific, and the only solution to the homogenous system of equations (1) would be the trivial one $(a,b)=(0,0)$, as illustrated in Figure 1a.

(FIGURE 1)

On the other hand, if there is complete market integration, all those n lines will coincide, as in Figure 1b. In that case, though, we can at best identify a direction in (a,b) space, which leaves both the scale and sign of the SDF undetermined, unless we add an asset whose price is different from 0, as in appendix A. As forcefully argued by Hillier (1990) for single equation IV models, this suggests that we should concentrate our efforts in estimating the identified direction, which can be easily achieved by using the polar coordinates $a=\sin\psi$ and $b=\cos\psi$ for $\psi\in[-\pi/2,\pi/2)$. However, empirical researchers often prefer to estimate points rather than directions, and for that reason they typically focus on some asymmetric scale normalization, such as $(1,b/a)$, although $(a/b,1)$ would also work. Figure 2a illustrates how different normalizations pin down different points along the identified direction. As we shall see below, this seemingly innocuous choice may have important empirical consequences.

(FIGURE 2)

We can also express the pricing conditions (1) in terms of central moments. Specifically, we can add and subtract $b\mu$ from $a + bf$, define $c = a + b\mu$ as the expected value of the affine SDF and express the pricing conditions as

$$E\left\{\begin{array}{c}[c + b\,(f - \mu)]\,\mathbf{r}\\ f - \mu\end{array}\right\} = \mathbf{0}. \tag{2}$$

The unknown parameters become (c, b, μ) instead of (a, b), but we have added an extra moment to estimate μ.[1] We refer to these two variants as the uncentred and centred SDF versions since they rely on either $E\,(\mathbf{r}f)$ or $Cov\,(\mathbf{r}, f)$ in explaining the cross-section of risk premia.[2]

Not surprisingly, under the null hypothesis of financial market integration we can only identify a direction in (c, b) space from (2), which again suggests that we should estimate the identified direction in terms of the polar coordinates $c = \sin\upsilon$ and $b = \cos\upsilon$ for $\upsilon \in [-\pi/2, \pi/2)$. Nevertheless, empirical work usually focuses on $(1, b/c)$. Alternatively, one could use $(c/b, 1)$, in which case the moment conditions would be linear in c/b and μ. Figure 2b shows how different normalizations pin down different points along the identified direction. Once more, we will discuss the empirical implications of these normalizations below.

2.2 The regression approach

Instead of explaining the cross-section of risk premia in terms of $Cov\,(\mathbf{r}, f)$, as in (2), we could equivalently use the vector $\boldsymbol{\beta} = Cov\,(\mathbf{r}, f)\,/V\,(f)$, which contains the slopes in the least squares projection of $\mathbf{r}$ onto the linear span of 1 and f. If $\boldsymbol{\phi} = E\,(\mathbf{r}) - \boldsymbol{\beta}E(f)$ denotes the corresponding vector of intercepts, the asset pricing restrictions (2) impose the parametric constraints

$$cE\,(\mathbf{r}) + bCov\,(\mathbf{r}, f) = c\boldsymbol{\phi} + d\boldsymbol{\beta} = \mathbf{0},$$

where $d = E\{[c + b(f - \mu)]f\}$ can be interpreted as the shadow cost of f. Hence, $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$ must belong to the same one-dimensional linear subspace. If we denote a basis for this subspace by the $n \times 1$ vector $\boldsymbol{\varphi}$, then we can impose the asset pricing constraint as $\boldsymbol{\phi} = -d\boldsymbol{\varphi}$ and $\boldsymbol{\beta} = c\boldsymbol{\varphi}$, so that the normal equations become

$$E\left[\begin{array}{c}\mathbf{r} + d\boldsymbol{\varphi} - c\boldsymbol{\varphi}f\\ (\mathbf{r} + d\boldsymbol{\varphi} - c\boldsymbol{\varphi}f)f\end{array}\right] = \mathbf{0}, \tag{3}$$

[1] Alternatively, we could work with covariances by centring $\mathbf{r}$ instead of f, which would require the addition of n moment conditions that define $E(\mathbf{r})$. We focus on (2) because it is more popular in empirical work as it involves fewer parameters for $n > 1$.

[2] Kan and Zhou (1999), Cochrane (2001b) and Jagannathan and Wang (2002) only study the first variant, but the second one is also widely used in the literature (see e.g. Parker and Julliard (2005) or Yogo (2006)).

where $(c, d, \boldsymbol{\varphi})$ are the new parameters to estimate.[3] As in the previous section, we can only identify a direction in (c, d) space. Once again, the usual asymmetric normalization in empirical work sets $(1, d/c)$, but we could also set $(c/d, 1)$ or indeed estimate the identified direction in terms of the polar coordinates $c = \sin\vartheta$ and $d = \cos\vartheta$ for $\vartheta \in [-\pi/2, \pi/2)$.

Alternatively, we could start from the uncentred variant of the SDF approach in (1), which explains the cross-section of risk premia in terms of $E(f\mathbf{r})$, and re-write the financial market integration restrictions using the vector $\boldsymbol{\eta} = E(f\mathbf{r})/E(f^2)$, which defines the regression slopes of the least squares projection of $\mathbf{r}$ onto the linear span of f only. Specifically, if $\boldsymbol{\rho} = E(\mathbf{r}) - \boldsymbol{\eta}\mu$ denotes the mean of the uncentred projection errors, the asset pricing restrictions (1) impose the parametric constraint

$$aE(\mathbf{r}) + bE(\mathbf{r}f) = a\boldsymbol{\rho} + d\boldsymbol{\eta} = \mathbf{0},$$

since $d = E[(a + bf) f]$. Hence, $\boldsymbol{\rho}$ and $\boldsymbol{\eta}$ must also belong to the same one dimensional subspace. If we denote a basis for this subspace by the $n \times 1$ vector $\boldsymbol{\varrho}$, then we can impose this constraint as $\boldsymbol{\rho} = -d\boldsymbol{\varrho}$ and $\boldsymbol{\eta} = a\boldsymbol{\varrho}$, so that the appropriate moment conditions would be

$$E\begin{bmatrix} \mathbf{r} + \boldsymbol{\varrho}d - a\boldsymbol{\varrho}f \\ (\mathbf{r} - a\boldsymbol{\varrho}f) f \end{bmatrix} = \mathbf{0}, \tag{4}$$

with $(a, d, \boldsymbol{\varrho})$ as the parameters to estimate. Once again, we can only identify a direction in (a, d) space, and the obvious asymmetric normalization would be $(1, d/a)$.

Given that (3) relies on covariances and (4) on second moments, we refer to these moment conditions as the centred and uncentred versions of the regression approach, respectively. However, since we are not aware of any empirical study based on (4), we shall not consider these moments conditions henceforth.

3 Traded factors

3.1 Moment conditions and parameters

Let us assume that the pricing factor f is itself the excess return on another asset, such as the market portfolio in the CAPM.[4] As forcefully argued by Shanken (1992), Farnsworth et al. (2002) and Lewellen, Nagel and Shanken (2010) among others, the pricing model applies to f too, which means that

$$E[(a + bf) f] = 0. \tag{5}$$

[3] An alternative, equivalent version of the second group of moment conditions in (3) would be $E[(r + d\boldsymbol{\varphi} - c\boldsymbol{\varphi})(f - \mu)] = 0$, which would require the addition of the moment condition $E(f - \mu) = 0$ to define μ. Theoretical and Monte Carlo results for these alternative moments are available on request.

[4] It is important to mention that our assets could include managed portfolios. Similary, the factor could also be a scaled version of a primitive excess return to accommodate conditioning information; see the discussion in chapter 8 of Cochrane (2001a).

The uncentred SDF approach relies on the $n+1$ moment conditions (1) and (5) once we choose a normalization for (a, b). As we mentioned before, the normalization could be asymmetric or symmetric. The latter relies on the directional coordinate ψ, while the former is typically implemented by factoring a out of the pricing conditions, leaving $\delta = -b/a$ as the only unknown parameter. Given moment condition (5), we will have that

$$\delta = -\cot\psi = \frac{\mu}{\gamma}, \tag{6}$$

where γ is the second moment of f, which allows us to interpret δ as a "price of risk" for the factor.

Similarly, the centred SDF approach works with the $n+2$ moment conditions (2) and

$$E\{[c + b\,(f-\mu)]f\} = 0. \tag{7}$$

Again, the normalization could be asymmetric or symmetric. The latter will make use of the polar coordinate υ, while the former is typically implemented by factoring c out of the pricing conditions, leaving $\tau = -b/c$ and μ as the only unknown parameters. Either way, we can use moment condition (7) to show that:

$$\tau = -\cot\upsilon = \frac{\mu}{\sigma^2}, \tag{8}$$

where $\sigma^2 = \gamma - \mu^2$ denotes the variance of f, which means that τ also has a "price of risk" interpretation.

When the risk factor coincides with the excess returns on a traded asset, its shadow cost d must coincide with its actual cost, which is 0. If we impose this constraint in the moment conditions (3), then the centred regression approach reduces to the $2n$ overidentified moment conditions

$$E\left[\begin{array}{c} \mathbf{r} - \boldsymbol{\beta} f \\ (\mathbf{r} - \boldsymbol{\beta} f)\, f \end{array}\right] = \mathbf{0}, \tag{9}$$

where the n unknown parameters are the elements of $\boldsymbol{\beta}$ because the regression intercepts must be 0 (see MacKinlay and Richardson (1991)).[5] As a result, the slope coefficients coincide with both $Cov(\mathbf{r},f)/V(f)$ and $E(\mathbf{r}f)/E(f^2)$ when (1) and (5) hold, so that the uncentred and centred variants of the regression (or beta) approach are identical in this case. The regression method identifies μ with the expected excess return of a portfolio whose "beta" is equal to 1. Thus, this parameter represents a "factor risk premium" when f is traded. To estimate it, we can add $f - \mu$ to (9), as in (2), and simultaneously estimate $\boldsymbol{\beta}$ and μ.

[5] These moment conditions confirm the result in Chamberlain (1983b) that says that $a + bf$ will constitute an admissible SDF if and only if f lies on the mean-variance frontier generated by f and $\mathbf{r}$. Then, the well-known properties of mean-variance frontiers imply that the least squares projection of $\mathbf{r}$ onto the linear span generated by a constant and f should be proportional to f.

Under standard regularity conditions (more on this in section 4.4), all three overidentifying restrictions (J) tests will follow an asymptotic chi-square distribution with n degrees of freedom when the corresponding moments are correctly specified.

The overidentification tests are regularly complemented by three standard evaluation measures. Specifically, we can define Jensen's alphas as $E(\mathbf{r}) - \boldsymbol{\beta} E(f)$ for the regression method, as well as the "pricing errors" associated to the uncentred SDF representation, $E(\mathbf{r}) - E(\mathbf{r}f)\,\delta$, and the centred SDF representation, $E(\mathbf{r}) - E[\mathbf{r}(f-\mu)]\tau$. In population terms, these three pricing errors coincide. In particular, they should be simultaneously 0 under the null hypothesis.

3.2 Numerical equivalence results

As we mentioned in the introduction, Kan and Zhou (1999, 2002), Cochrane (2001b), Jagannathan and Wang (2002), Burnside (2012) and Kan and Robotti (2008) compare some of the aforementioned approaches when researchers rely on traditional, two-step or iterated GMM procedures. In contrast, we show that all the methods coincide if one uses instead single-step procedures such as CU-GMM, which we describe in appendix F. More formally:

Proposition 1 *If we apply single-step procedures to the uncentred SDF method based on the moment conditions (1) and (5), the centred SDF method based on the moment conditions (2) and (7), and the regression method based on the moment conditions (9), then for a common specification of the characteristics of the HAC weighting matrix the following numerical equivalences hold for any finite sample size:*
1) The overidentification restrictions (J) tests regardless of the normalization used.
2) The direct estimates of (a, b) from (1) and (5), their indirect estimates from (2) and (7) that exploit the relationship $c = a + b\mu$, and the indirect estimates from (9) extended to include (μ, γ) which exploit the relationship $a\mu + b\gamma = 0$ when we use symmetric normalizations or compatible asymmetric ones. Analogous results apply to (c, b) and μ.
3) The estimates of Jensen's alphas $E(\mathbf{r}) - \boldsymbol{\beta} E(f)$ obtained by replacing $E(\cdot)$ by an unrestricted sample average and the elements of $\boldsymbol{\beta}$ by their direct estimates obtained from the regression method, and the indirect estimates obtained from SDF methods with symmetric normalizations and compatible asymmetric ones extended to include $\boldsymbol{\beta}$. Analogous results apply to the alternative pricing errors of the uncentred and centred SDF representations.

Importantly, these numerical equivalence results do not depend in any sense on the number of assets or indeed the number of factors, and remain true regardless of the validity of the asset pricing restrictions. In order to provide some intuition, imagine that for estimation purposes we assumed that the joint distribution of $\mathbf{r}$ and f is *i.i.d.* multivariate normal. In that context, we could test the mean-variance efficiency of f by means of a likelihood ratio (LR) test. We could then factorize the joint log-likelihood function of $\mathbf{r}$ and f as the marginal log-likelihood of f, whose parameters μ and σ^2 would be unrestricted, and the conditional log-likelihood of $\mathbf{r}$ given f. As a result, the LR version of the original Gibbons, Ross and Shanken (1989) test would be numerically identical to the LR test in the joint system irrespective of the chosen parameterization. The CU-GMM overidentification test, which implicitly uses the Gaussian

scores as influence functions, inherits the invariance of the LR test. The advantage, though, is that we can make it robust to departures from normality, serial independence or conditional homoskedasticity.

From a formal point of view, the equivalence between the two SDF approaches is a direct consequence of the fact that single-step procedures are numerically invariant to normalization, while the additional, less immediate results relating the regression and SDF approaches in proposition 1 follow from the fact that those GMM procedures are also invariant to reparameterizations and parameter dependent linear transformations of the moment conditions (see again appendix F).[6]

3.3 Starting values and other implementation details

One drawback of CU-GMM and other GEL estimators is that they involve a non-linear optimization procedure even if the moment conditions are linear in parameters, which may result in multiple local minima. In this sense, the uncentred SDF method has a non-trivial computational advantage because it contains a single unknown parameter.[7] At the same time, one can also exploit the numerical equivalence of the different approaches covered in proposition 1 to check that a global minimum has been reached. Likewise, one could also exploit the numerical equivalence of the Euclidean empirical likelihood and CU-GMM estimators of the model parameters (see Antoine, Bonnal and Renault (2006)). A much weaker convergence test is the fact that the value of the criterion function at the CU-GMM estimators cannot be larger than at the iterated GMM estimators, which do not generally coincide (see Hansen, Heaton and Yaron (1996)).

In any case, it is convenient to have good initial parameter values. For that reason, we propose to use as starting value a computationally simple intuitive estimator that is always consistent, but which would become efficient for *i.i.d.* elliptical returns, a popular assumption in finance because it guarantees the compatibility of mean-variance preferences with expected utility maximization regardless of investors' preferences (see Chamberlain (1983a) and Owen and Rabinovitch (1983)):

Lemma 1 *If $(\mathbf{r}_t, f_t)$ is an i.i.d. elliptical random vector with bounded fourth moments and the null hypothesis of linear factor pricing holds, then the most efficient GMM estimator of $\delta = -b/a$ obtained from (1) and (5) will be given by*

$$\dot{\delta}_T = \frac{\sum_{t=1}^{T} f_t}{\sum_{t=1}^{T} f_t^2}. \tag{10}$$

[6] Empirical researchers sometimes report the cross-sectional (squared) correlation between the actual and model implied risk premia. Proposition 1 trivially implies that they would also obtain a single number for each of the three approaches if they used single-step GMM.

[7] This advantage becomes more relevant as the number of factors k increases because the centred SDF method requires the additional estimation of k factor means and the regression method the estimation of $n \times k$ factor loadings.

Intuitively, this means that in those circumstances (5), which is the moment involving f, exactly identifies the parameter δ, while (1), which are the moments corresponding to $\mathbf{r}$, provide the n overidentification restrictions to test. Although the elliptical family is rather broad (see Fang, Kotz and Ng (1990)), and includes the multivariate normal and Student t distribution as special cases, it is important to stress that $\dot{\delta}_T$ will remain consistent under linear factor pricing even if the assumptions of serial independence and ellipticity are not totally realistic in practice.[8]

A rather different justification for (10) is that it coincides with the GMM estimator of δ that we would obtain from (1) and (5) if we used as weighting matrix the second moment of the vector of excess returns $\mathbf{x} = (f, \mathbf{r}')'$. Specifically, (10) minimizes the sample counterpart to the Hansen and Jagannathan (1997) distance

$$E\left[(1-\delta f)\,\mathbf{x}\right]' \left[E\left(\mathbf{x}\mathbf{x}'\right)\right]^{-1} E\left[(1-\delta f)\,\mathbf{x}\right]$$

irrespective of the distribution of returns and the validity of the asset pricing model.

Hansen, Heaton and Yaron (1996) also indicate that CU-GMM occasionally generates extreme estimators that lead to large pricing errors with even larger variances. In those circumstances, we would suggest the imposition of good deal restrictions (see Cochrane and Saa-Requejo (2000)) to rule out implausible results.[9]

4 Non-traded factors

4.1 Moment conditions and parameters

Let us now consider situations in which f is either a scalar non-traded factor, such as the growth rate of per capita consumption, or the empirical researcher ignores that it is traded. The main difference with the analysis in section 3 is that the pricing equations (5) and (7) are no longer imposed, so that the SDF is defined by (1) or (2) only. Similarly, the regression approach relies on (3) or (4) without the additional parametric constraint $d = 0$ implied by a traded factor. Obviously, the resulting reduction in the number of moment conditions or constraints yields a reduction in the degree of overidentification, which becomes $n - 1$.

[8] We can also prove that we obtain an estimator of δ that is asymptotically equivalent to (10) if we follow Spanos (1991) in assuming that the so-called *Haavelmo* distribution, which is the joint distribution of the $T(n+1)$ observed random vector $(\mathbf{r}_1, f_1, \ldots, \mathbf{r}_t, f_t, \ldots, \mathbf{r}_T, f_T)$, is an affine transformation of a scale mixture of normals, and therefore elliptical. Intuitively, the reason is that a single sample realization of such a Haavelmo distribution is indistinguishible from a realization of size T of an *i.i.d.* multivariate normal distribution for $(\mathbf{r}_t, f_t)$.

[9] Specifically, given that we know from Hansen and Jagannathan (1991) that

$$S^2 \cdot E^2(m)/V(m) = R^2,$$

where S is the maximum attainable Sharpe ratio of any portfolio of the assets under consideration, and R^2 is the coefficient of determination in the (theoretical) regression of f on a constant and the tradeable assets, one could estimate the linear factor pricing model subject to implicit restrictions that guarantee that the values of S or the coefficient of variation of m computed under the null should remain within some loose but empirically plausible bounds. In the case of traded factors both these bounds should coincide because $R^2 = 1$.

Nevertheless, we can still provide a "price of risk" interpretation to some parameters, but this time in terms of factor mimicking portfolios. In particular, (6) is replaced by

$$\delta = -\cot\psi = \frac{E(r^{+})}{E(r^{+2})}, \tag{11}$$

where

$$r^{+} = E(f\mathbf{r}')E^{-1}(\mathbf{r}\mathbf{r}')\mathbf{r} \tag{12}$$

is the uncentred least squares projection of f on $\mathbf{r}$. Similarly, (8) becomes

$$\tau = -\cot\upsilon = \frac{E(r^{++})}{V(r^{++})}, \tag{13}$$

where

$$r^{++} = Cov(f,\mathbf{r}')V^{-1}(\mathbf{r})\mathbf{r}$$

is the centred least squares projection of f on $\mathbf{r}$.

In turn, given that the standard implementation of the centred regression uses the asymmetric normalization $(1, d/c)$ in the $2n$ overidentified moment restrictions (3), and estimates the $n+1$ parameters $\varkappa = -d/c$ and $\boldsymbol{\beta} = \boldsymbol{\varphi}c$ (see Campbell, Lo and MacKinlay (1996, chap. 5)), we can interpret $\lambda = \varkappa + \mu$ as the "factor risk premium": the expected excess return of a portfolio whose "beta" is equal to 1.[10]

Finally, the expressions for the centred and uncentred SDF pricing errors at the end of section 3 continue to be valid, while Jensen's alphas are now defined as $E(\mathbf{r}) - \boldsymbol{\beta}\lambda$.

4.2 Numerical equivalence results

As in the case of traded factors, we can show that all the approaches discussed in the previous subsection coincide if one uses single-step methods. More formally

Proposition 2 *If we apply single-step procedures to the uncentred SDF method based on the moment conditions (1), the centred SDF method based on the moment conditions (2), and the centred regression method based on the moment conditions (3), then for a common specification of the characteristics of the HAC weighting matrix the following numerical equivalences hold for any finite sample size:*
1) The overidentification restrictions (J) tests regardless of the normalization used.
2) The direct estimates of (a, b) from (1), their indirect estimates from (2) that exploit the relationship $c = a + b\mu$, and the indirect estimates from (3) extended to include (μ, γ) that exploit the relationships $c = a + b\mu$ and $d = a\mu + b\gamma$ when we use symmetric normalizations or compatible asymmetric ones. Analogous results apply to (c, b) and (c, d).
3) The estimates of Jensen's alphas $E(\mathbf{r}) - \boldsymbol{\beta}\lambda$ obtained by replacing $E(\cdot)$ by an unrestricted sample average and the elements of $\boldsymbol{\beta}\lambda$ by their direct estimates obtained from the regression method, and the indirect estimates obtained from SDF methods with symmetric normalizations and compatible asymmetric ones extended to include $\boldsymbol{\beta}$, γ and μ. Analogous results apply to the alternative pricing errors of the uncentred and centred SDF representations.

[10] Jagannathan and Wang (2002) use $\lambda - \mu$ instead of $\varkappa$, and add the influence functions $f - \mu$ and $(f-\mu)^2 - \sigma^2$ to estimate μ and σ^2 too. The addition of these moments is irrelevant for the estimation of $\varkappa$ and the J test because they exactly identify μ and σ^2 (see e.g. pp. 196–197 in Arellano (2003) for a proof of the irrelevance of unrestricted moments).

Once again, we can gain some intuition by assuming that the joint distribution of $\mathbf{r}$ and f is *i.i.d.* multivariate normal. In that context, we could test the validity of the model by means of a LR test that compares the restricted and unrestricted criterion functions, as in Gibbons (1982). We could then factorize the joint log-likelihood function of $\mathbf{r}$ and f as the marginal log-likelihood of f, whose parameters μ and σ^2 would be unrestricted, and the conditional log-likelihood of $\mathbf{r}$ given f, which would have an affine mean and a constant variance. As a result, the LR version of the linear factor pricing test would be numerically identical to the LR test in the joint system irrespective of the chosen parameterization. The CU-GMM overidentification test, which implicitly uses the Gaussian scores as influence functions, inherits the invariance of the LR test. The advantage, though, is that we can make it robust to departures from normality, serial independence or conditional homoskedasticity.[11] As we shall see in section 4.4, though, we can encounter situations in which some of the popular asymmetric normalizations are incompatible the estimates obtained with the symmetric ones.

It is important to distinguish proposition 2 from the results in Jagannathan and Wang (2002) and Kan and Zhou (2002). These authors showed that the centred regression and uncentred SDF approaches lead to asymptotically equivalent inferences under the null and compatible sequences of local alternatives in single factor models. In contrast, proposition 2 shows that in fact both SDF approaches and the regression method yield numerically identical conclusions if we work with single-step GMM procedures. Since our equivalence result is numerical, it holds regardless of the validity of the pricing model and irrespective of n or the number of factors.[12]

4.3 Starting values and other implementation details

The numerical equivalence of the different approaches gives once more a non-trivial computational advantage to the uncentred SDF method, which only contains a single unknown parameter. At the same time, one can also exploit the fact that the approaches discussed in proposition 2 coincide to check that a global minimum has been obtained.

[11] Kan and Robotti (2008) also show that CU-GMM versions of the SDF approach are numerically invariant to affine transformations of the factors with known coefficients, which is not necessarily true of two-step or iterated GMM methods. Not surprisingly, it is easy to adapt the proof of Proposition 2 to show that the regression approach is also numerically invariant to such transformations.

[12] We could also consider a nonlinear SDF such as $m = f^{\zeta}$, with ζ unknown, so that the moments would become

$$E(\mathbf{r} f^{\zeta}) = \mathbf{0}.$$

In this context, we can easily show that a single-step overidentifying restrictions test would be numerically equivalent to the one obtained from the "regression"-based moment conditions

$$E\left[\begin{array}{c} (\mathbf{r} - \boldsymbol{\beta}_m(f^{\zeta} - \gamma_m/\mu_m)) \\ (\mathbf{r} - \boldsymbol{\beta}_m(f^{\zeta} - \gamma_m/\mu_m)))f^{\zeta} \\ f^{\zeta} - \mu_m \\ f^{2\zeta} - \gamma_m \end{array}\right] = \mathbf{0},$$

whose unkown parameters are $(\zeta, \boldsymbol{\beta}_m, \mu_m, \gamma_m)$.

Still, it is convenient to have good initial values. For that reason, we propose a computationally simple intuitive estimator that is always consistent, but which would become efficient when the returns and factors are *i.i.d.* elliptical, which nests the multivariate normal assumption in Jagannathan and Wang (2002):

Lemma 2 *If* $(\mathbf{r}_t, f_t)$ *is an i.i.d. elliptical random vector with bounded fourth moments such that* $E(\mathbf{r}_t f_t) \neq \mathbf{0}$ *and the null hypothesis of linear factor pricing holds, then the most efficient GMM estimator of* $\delta = -b/a$ *obtained from (1) will be given by*

$$\mathring{\delta}_T = \frac{\sum_{t=1}^{T} r_t^+}{\sum_{t=1}^{T} r_t^{+2}} \tag{14}$$

where r_t^+ *is the uncentred factor mimicking portfolio defined in (12), whose sample counterpart would be*

$$\tilde{r}_t^+ = \left(\sum_{s=1}^{T} f_s \mathbf{r}_s'\right)\left(\sum_{s=1}^{T} \mathbf{r}_s \mathbf{r}_s'\right)^{-1} \mathbf{r}_t.$$

Once again, it is important to stress that the feasible version of (14) will remain consistent under linear factor pricing even if the assumptions of serial independence and a multivariate elliptical distribution are not totally realistic in practice.

Importantly, (14) also coincides with the GMM estimator of δ that we would obtain from (1) if we used as weighting matrix the second moment of the excess returns in $\mathbf{r}$. In particular, the feasible version of $\mathring{\delta}_T$ minimizes the sample counterpart to the Hansen and Jagannathan (1997) distance

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

irrespective of the distribution of returns and the validity of the asset pricing model.

4.4 Problematic cases and tests to detect them

As we saw in section 2, the existence of a unique (up to scale) affine SDF $a + bf$ that correctly prices the vector of excess returns at hand means that the $n \times 2$ matrix with columns $E(\mathbf{r})$ and $E(\mathbf{r}f)$ has rank 1. Such a condition is related to the uncentred SDF approach. We also saw in the same section that we can transfer this rank 1 condition to a matrix constructed with $E(\mathbf{r})$ and $Cov(\mathbf{r}, f)$, which is related to the centred SDF approach, another matrix built from $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$ in the case of the centred regression, or indeed a matrix that concatenates $\boldsymbol{\rho}$ and $\boldsymbol{\eta}$ in an uncentred regression.

From an econometric perspective, those rank 1 matrices are important because their elements determine the expected Jacobian of the moment conditions with respect to the parameters. As is well known, one of the regularity conditions for standard GMM asymptotics is that the relevant Jacobian matrix must have full column rank in the population (see Hansen (1982)).

When the pricing factor is traded, we should add to these matrices a row whose second element is always different from 0. This additional row ensures that all the Jacobians have full rank when risk premia are not all simultaneously zero (see lemma G1 in appendix G).

When the pricing factor is non-traded, or treated as if it were so, all the symmetrically normalized moment conditions also have a full column rank Jacobian as long as risk premia are not zero (see lemma G2 in appendix G). As a result, if the additional GMM regularity conditions are satisfied, the unique single step overidentification test associated to all of them will be asymptotically distributed as χ^2_{n-1} under the null.[13] Moreover, the multistep overidentification tests will also share this asymptotic distribution.

In contrast, there are some special cases in which the population Jacobians of some of the asymmetrically normalized moment conditions do not have full rank.[14] Next, we study in detail the case of an uncorrelated factor, which is the most relevant one in empirical work.

4.4.1 An uncorrelated factor

As we show in lemma G3 in appendix G, when $Cov(\mathbf{r}, f) = \mathbf{0}$ but $E(\mathbf{r}) \neq \mathbf{0}$ the uncentred SDF moment conditions (1) asymmetrically normalized through the parameter δ will have a full rank Jacobian, with the "true value" being $\delta = 1/E(f)$ (see also section 5.1 of Burnside (2012), who uses the term "A-Normalization"). The centred SDF moment conditions (2) normalized with $(c/b, 1)$ and indeed the centred regression moment conditions (3) asymmetrically normalized with $(c/d, 1)$ are also well-behaved.

In contrast, (3) asymmetrically normalized in terms of $\varkappa$ will be set to 0 with $\boldsymbol{\beta} \to \mathbf{0}$ and $\varkappa\boldsymbol{\beta} \to E(\mathbf{r})$, but the expected Jacobian of these moment conditions will be increasingly singular along that path. Similarly, the moment conditions (2) asymmetrically normalized through the parameter τ, will be satisfied as $\mu \to E(f)$ and $\tau[\mu - E(f)] \to 1$ (see also appendix C in Burnside (2012), who talks about the "M-Normalization"), but again the expected Jacobian of these moment conditions will become increasingly singular. In those circumstances, the multistep J tests that use those problematic asymmetric normalizations will have a non-standard distribution under the null, which will lead to substantial size distortions in large samples if we rely on the χ^2_{n-1} critical values (see Dovonon and Renault (2013) for a thorough discussion of the properties

[13] This common asymptotic distribution would be shared with the Likelihood Ratio test of the asset pricing restrictions under the assumption that the distribution of $\mathbf{r}$ given f is jointly normal with an affine mean and a constant covariance matrix, which would also be invariant to reparameterization.

[14] In models defined by linear in parameters moment conditions, rank failure of the Jacobian is tantamount to underidentification. However, as forcefully argued by Sargan (1983), there are non-linear models in which the rank condition fails at the true values but not in their neighborhood, and yet the parameters are locally identified. In that case, we say that they are *first-order* underidentifiable. Similarly, if the expected value of the Jacobian of the Jacobian is also of reduced rank, then the parameters are said to be *second-order* underidentifiable, and so on. Obviously, if all the higher order Jacobians share a rank failure, the parameters will be locally underidentified (see also Arellano, Hansen and Sentana (2012)). In our case, the moment conditions are at most quadratic in the parameters, so second-order underidentifiability would be equivalent to local underidentifiability.

of the J test in an example of a quadratic in parameters model with rank failure of the Jacobian).

Intuitively, the reason for the differential behavior of the asymmetric normalizations $(1, b/a)$ and $(1, b/c)$ is the following. As illustrated in Figure 2a, the values of a and b are determined by the intersection between the straight lines (1) and $(1, b/a)$, which remains well defined even if the risk factor is uncorrelated with the vector of excess returns. In contrast, as $Cov(\mathbf{r}, f) \to \mathbf{0}$ the lines $(1, b/c)$ and the pricing condition in (2) cross at an increasingly higher value of b, and eventually become parallel (see Figures 2b and 3a). For analogous reasons, one cannot find any finite value of $\varkappa = -d/c$ that will satisfy (3) when $\boldsymbol{\beta} = c\boldsymbol{\varphi} \to \mathbf{0}$.

(FIGURE 3)

From an economic point of view, a risk factor for which $Cov(\mathbf{r}, f) = \mathbf{0}$ is not very attractive. The unattractiveness of f is confirmed by the fact that the centred mimicking portfolio r^{++} will be 0. In fact, it is easy to construct examples in which the true underlying SDF that prices all primitive assets in the economy is affine in another genuine risk factor, g say, and yet any SDF proportional to $1 - f/E(f)$ will be compatible with (1) for the vector of asset returns at hand if we choose f such that it is uncorrelated with $\mathbf{r}$ (see Burnside (2011)). Given that the J tests of the asset pricing conditions that do not impose the problematic asymmetric normalization $(1, b/c)$ will fail to reject their null, we propose a simple test to detect this special case.

It is easy to see that $Cov(\mathbf{r}, f) = \mathbf{0}$ is equivalent to all valid SDFs affine in f having a 0 mean. Therefore, we can re-estimate the different moment conditions with this additional restriction imposed, and compute a distance metric (DM) test, which is the GMM analogue to a LR statistic, as the difference between the criterion function under the null and the alternative.

In the case of the uncentred SDF moment conditions (1), the restriction can be imposed by adding the moment condition

$$E\left(a + bf\right) = 0 \tag{15}$$

expressed in such a way that it is compatible with the chosen asymmetric or symmetric normalization. Intuitively, this additional condition defines the expected value of the SDF, which we then set to 0 under the null. Consequently, the DM test will follow an asymptotic χ^2_1 distribution under the null of $Cov(\mathbf{r}, f) = \mathbf{0}$.[15]

[15] It is also straightforward to derive analogous distance metric tests associated to the moment conditions (2) and (3). However, since their single-step versions are numerically identical, we shall not discuss them any further.

4.4.2 Underidentification

Unfortunately, an intrinsic problem of any asymmetric normalizations is that there is always a configuration of the population first and second moments of $\mathbf{r}$ and f which is incompatible with it. For example, $E(\mathbf{r}f) = 0$ will be problematic for the normalization $(1, b/a)$ as illustrated in Figure 3b and described in detail in appendix C.[16] From an econometric point of view, though, the truly problematic case arises when $E(\mathbf{r}f) = \mathbf{0}$ and $E(\mathbf{r}) = \mathbf{0}$, which in turn implies that $Cov(\mathbf{r}, f) = \mathbf{0}$. In this situation, the asset pricing conditions (1) trivially hold, but the uncentred SDF parameters a and b are underidentified even after normalization, which renders standard GMM inferences invalid. Obviously, the same problem applies to all the other moment conditions.

Following Arellano, Hansen and Sentana (2012), this problematic case can be detected with the J test of the augmented set of $2n$ moment conditions

$$E\begin{pmatrix}\mathbf{r}\\ f\mathbf{r}\end{pmatrix} = \mathbf{0},$$

which involve no parameters (see Manresa, Peñaranda and Sentana (2014) for further details).[17]

5 Empirical application

Over the last thirty years many empirical studies have rejected the hypothesis of uncovered interest parity, which in its basic form implies that the expected return to speculation in the forward foreign exchange market conditioned on available information should be zero. Specifically, many of those studies find support for the so-called the "forward premium puzzle", which implies that, contrary to the theory, high domestic interest rates relative to those in the foreign country predict a future appreciation of the home currency. In fact, the so-called "carry trade", which involves borrowing low-interest-rate currencies and investing in high-interest-rate ones, constitutes a very popular currency speculation strategy developed by financial market practitioners to exploit this "anomaly" (see Burnside et al. (2006)).

One of the most popular explanations among economists is that such a seemingly anomalous pattern might reflect a reward to the exposure of foreign currency positions to certain systematic

[16] Similarly, if we work with the centred regression moment conditions (3) asymmetrically normalized in terms of $(c/d, 1)$ and the least squares projection of $\mathbf{r}$ onto (the span of) 1 and f is proportional to f, so that the nontraded factor effectively behaves as if it were traded, then this normalization will not be well-behaved (see again lemma G3). Likewise, the asymmetric normalization $(a/b, 1)$ applied to (1) will run into difficulties when $E(\mathbf{r}) = \mathbf{0}$ but $E(\mathbf{r}f) = Cov(\mathbf{r}, f) \neq \mathbf{0}$. Intuitively, the reason is that admissible SDFs must be constant when risk neutrality effectively holds in the data at hand.

[17] See also Kan and Zhang (1999), Burnside (2012) and appendix A for the implications that other types of identification failures have for GMM procedures.

risk factors. To study this possibility, Lustig and Verdelhan (2007) constructed eight portfolios of currencies sorted at the end of the previous year by their nominal interest rate differential to the US dollar, creating in this way annual excess returns (in real terms) on foreign T-Bill investments for a US investor over the period 1953-2002. Interestingly, the broadly monotonic relationship between the level of interest rates differentials and risk premia for those portfolios captured in Figure 1 of their paper provides informal evidence on the failure of uncovered interest rate parity.

Lusting and Verdelhan (2007) used two-pass regressions to test if some popular empirical asset pricing models that rely on certain domestic US risk factors were able to explain the cross-section of risk premia. In what follows, we use their data to estimate the parameters and assess the asset pricing restrictions of the different sets of moments conditions described in previous sections by means of two-step, iterated and CU-GMM.[18] In all cases, we estimate the asymptotic covariance matrix of the relevant influence functions by means of its sample counterpart, as in Hansen, Heaton and Yaron (1996). As for the first-step estimators, we use the identity matrix as initial weighting matrix given the prevalence of this practice in empirical work. Finally, we implicitly choose the leverage of the carry trades whose payoffs are the excess returns by systematically expressing all returns and factors as pure numbers. This scaling does not affect CU or iterated GMM, but it affects some of the two-step GMM procedures.[19]

5.1 Traded factor

Given that for pedagogical reasons we have only considered a single traded factor in our theoretical analysis, we focus on the CAPM. Following Lustig and Verdelhan (2007), we take the pricing factor to be the US market portfolio, which we also identify with the CRSP value-weighted excess return. Table 1 contains the results of applying the different inference procedures previously discussed to this model. Importantly, Figure G1a in appendix G, which plots the CU-GMM criterion as a function of δ, confirms that we have obtained a global minimum.

The first thing to note is that the value of the CU-GMM overidentification restriction statistic is the same across five different variants covered by proposition 1. In contrast, there are marked numerical differences between the corresponding two-step versions of the J test. In particular, an asymmetrically normalized version of the centred SDF approach yields a substantially higher value, while the two symmetric SDFs and the regression variants have p-values above 50%. These numerical differences are reduced but not eliminated as we update the weighting matrix.

[18] We have also considered other single step procedures such as empirical likelihood and exponentially-tilted methods, but since they yield J tests, parameter estimates and standard errors similar to their CU/Euclidean empirical likelihood counterparts, we do not report them in the interest of space.

[19] In contrast, the scale of the data does not affect those two-step GMM procedures that use (10) or (14) as first-step estimators instead of relying on the identity matrix.

In particular, iterated GMM applied to symmetric centred SDF gives a test statistic similar to CU, while its asymmetric version is still much higher.

(TABLE 1)

Table G1 in appendix G also confirms the numerical equality of the CU-GMM estimators of prices of risk (δ, τ and μ) and pricing errors regardless of the approach used to estimate them, as stated in points 2 and 3 of proposition 1. In contrast, two-step and iterated GMM yield different results, which explains the three different columns required for each of them.[20] In addition, the magnitudes of the two-step, iterated and CU-GMM estimates of δ and μ are broadly the same, while the CU-GMM estimate of τ is noticeably higher than its multistep counterparts.

In any case, most tests reject the null hypothesis of linear factor pricing. Interestingly, these rejections do not seem to be due to poor finite sample properties of the J statistics in this context since the F version of the Gibbons, Ross and Shanken (1989) regression test, which remains asymptotically valid in the case of conditional homoskedasticity, also yields a p-value of 0.3%.

The J tests reported in Table 1 can also be interpreted as DM tests of the null hypothesis of zero pricing errors in the eight currency returns only. The rationale is as follows. If we saturate (1) by adding n pricing errors, then the joint system of moment conditions becomes exactly identified, which in turn implies that the optimal criterion function under the alternative will be zero.

We can also consider the DM test of the null hypothesis of zero pricing error for the traded factor. Once again, the criterion function under the null takes the value reported in Table 1. Under the alternative, though, we need to conduct a new estimation. Specifically, if we saturate the moment condition (5) corresponding to the traded factor by adding a single pricing error, then the exact identifiability of this modified moment condition means that the joint system of moment conditions effectively becomes equivalent to another system that relies on (1) only. Treating the excess return on the US stock market as a nontraded factor delivers a CU-GMM J test of 6.87 (p-value 0.44). Hence, the CAPM restrictions are not rejected when we do not force this model to price the market, although the estimated τ is negative. In contrast, the DM test of zero pricing error for the traded factor, which is equal to the difference between this J

[20] The implied estimate of μ from the uncentred SDF approach also differs between two-step and iterated GMM (0.139 vs. 0.150), which are in turn different from the sample mean of f. The reason is that GMM equates to zero the average of the sample analogue of the orthogonalized influence function $(f-\mu) - E\left[(f-\mu)\, m\mathbf{r}\right]\left[E\left(m^2\mathbf{r}\mathbf{r}'\right)\right]^{-1}(m\mathbf{r})$, (assuming *i.i.d.* observations) where $m = 1-\delta f$, rather than the average of $f-\mu$. This residual depends on the estimate of δ, which differs between two-step and iterated GMM (4.455 vs. 4.534).

statistic and the one reported in Table 1, is 12.09, with a tiny p-value. Therefore, the failure of the CAPM to price the US stock market portfolio provides the clearest source of model rejection, thereby confirming the relevance of the recommendation in Shanken (1992), Farnsworth et al. (2002) and Lewellen, Nagel and Shanken (2010).

Importantly, these DM tests avoid the problems that result from the degenerate nature of the joint asymptotic distribution of the pricing error estimates recently highlighted by Gospodinov, Kan and Robotti (2012). This would be particularly relevant in the elliptical case because the moment condition (5) coincides with the optimal one in view of lemma 1.

5.2 Non-traded factor

Let us now explore a linearized version of the CCAPM, which defines the US per capita consumption growth of nondurables as the only pricing factor. Table 2 displays the results from the application of the different inference procedures previously discussed for the purposes of testing this model. Once again, Figure G1b in appendix G, which plots the CU-GMM criterion as a function of δ, confirms that we have obtained a global minimum.

In this case, the common CU-GMM J test (5.66, p-value 58%) does not reject the null hypothesis implicit in (1), (2) or (3), which is in agreement with the empirical results in Lustig and Verdelhan (2007). This conclusion is confirmed by a p-value of 83.9% for the test of the same null hypothesis computed from the regression using the expressions in Beatty, LaFrance and Yang (2005). Their F-type test is asymptotically valid in the case of conditional homoskedasticity, and may lead to more reliable inferences in finite samples.

In contrast, there are important numerical differences between the standard two-step GMM implementation of the five approaches, which lead to diverging conclusions at conventional significance levels. Specifically, while the asymmetric centred SDF approach rejects the null hypothesis, its symmetric version does not, with p-values of almost zero and 47%, respectively. These numerical differences are attenuated when we use iterated GMM procedures, but the contradicting conclusions remain.

(TABLE 2)

In contrast, when we look at the uncentred SDF (both symmetric and asymmetric variants) and regression approaches, the multistep GMM procedures yield results closer to CU-GMM. In particular, the two-step and iterated versions of the J test of the centred regression are closer to its uncentred SDF counterpart than to the centred SDF one. The reason is that in (3) we do not need to rescale the influence functions when we switch from the asymmetric normalization

$(1, d/c)$ to $(c/d, 1)$. Therefore, both normalizations are numerically equivalent not only with CU-GMM but also with two-step and iterated GMM. In contrast, in the centred SDF moments (2) we rescale the influence functions as we switch from the asymmetric normalization $(1, b/c)$ to $(c/b, 1)$.

Table G2 in appendix G also confirms the numerical equality of the CU-GMM estimators of prices of risk (δ, τ and λ) and pricing errors regardless of the approach used to estimate them, as expected from points 2 and 3 of proposition 2. In contrast, two-step and iterated GMM yield different results. In this case, all the estimates of δ and λ are fairly close, but the CU-GMM estimate of τ is much higher than its multistep counterparts. However, the directional estimates based on υ in the symmetric variant of the centred SDF approach behave very similarly across the different GMM implementations. Therefore, we can conclude that a very important driver of the differences between test statistics and parameter estimates is the normalization chosen, possibly even more than the use of centred or uncentred moments, or indeed the use of CU or iterated GMM.

The discrepancies that we observe suggest that we may have encountered one of the problematic situations described in section 4.4. The hypothesis of zero risk premia is clearly rejected with a J statistic of 39.97, whose p-value is effectively 0. Therefore, there are statistically significant risk premia in search of pricing factors to explain them. Similarly, the hypothesis of underidentification in section 4.4.2 is also rejected with a statistic of 53.04 and a negligible p-value, which confirms that the parameters appearing in (1), (2) and (3) are point identified after normalization.

Nevertheless, there is little evidence against the hypothesis of a zero mean SDF. Specifically, the DM test introduced in section 4.4.1 yields 2.73 and a p-value of almost 10%. The relevance of this p-value is reinforced by the findings of a Monte Carlo experiment reported in the next section, which suggest that this test tends to overreject.

It is worth noting that CU-GMM proves once again useful in unifying the empirical results in this context because the joint overidentification test of (1) and (15), which trivially coincides with the sum of the DM test of a SDF with zero mean and the J test of the CCAPM pricing restrictions, is numerically equivalent to a test of the null that all the betas are 0, whose p-value is 36%. For analogous reasons, we obtain the same J test whether we regress $\mathbf{r}$ on f or f on $\mathbf{r}$. This lack of correlation does not seem to be due to excessive reliance on asymptotic distributions, because it is corroborated by a p-value of 81.7% for the F test of the second univariate regression, which like the corresponding LR test, is also invariant to exchanging regressand and regressors. As explained by Savin (1983) using results from Scheffé (1953), the joint test of an uncorrelated

factor is effectively testing that any portfolio formed from the eight currency portfolios has zero correlation with US consumption growth (see also Gibbons, Ross and Shanken (1989) for a closely related argument). Obviously, if we computed t-tests between every conceivable portfolio and consumption growth, a non-negligible fraction of them will be statistically significant, so the usual trade off between power and size applies (see Lustig and Verdelhan (2011) and Burnside (2011) for further discussion of this point). In any case, the number of portfolios must be strictly larger than the number of pricing factors for (1) to have testable implications.

In summary, the fact that we cannot reject the asset pricing restrictions implicit in (1), (2) or (3) must be interpreted with some care. In this sense, the CCAPM results are very similar to the ones described at the end of the previous subsection when we treated the market portfolio as non-traded. This is not very surprising given that the correlations between the eight currency portfolios and the excess returns on the US market portfolio and consumption growth are of similar order.

6 Monte Carlo

In this section we report the results of some simulation experiments based on a linear factor pricing model with a nontraded factor. In this way we assess the reliability of the empirical evidence on the CCAPM we have obtained in section 5.2. Given that the number of mean, variance and correlation parameters for eight arbitrage portfolios and a risk factor is rather large, we have simplified the data generating process (DGP) as much as possible, so that in the end we only had to select a handful of parameters with simple interpretation; see appendix D for details.

We consider two different sample sizes: $T = 50$ and $T = 500$ and three designs (plus a fourth one in appendix C). In the first two, there is a valid SDF affine in the candidate risk factor, which gives rise to a 0 Hansen-Jagannathan distance, while in the third one, a second risk factor would be needed. In the interest of space, we only report results for the combination of normalizations, moments and initial conditions that we have analyzed in the empirical application. In view of the discussion of Table 2 in section 5, in the case of the multistep regression estimators we systematically computed the two asymmetric normalizations $(1, d/c)$ and $(c/d, 1)$ mentioned in section 2, and kept the results that provided the lower J statistic. We did so because the regression criterion function very often fails to converge in the neighborhood of $\boldsymbol{\beta} = \mathbf{0}$ (or $\boldsymbol{\phi} = \mathbf{0}$) even when the population values of those parameters are far away.

Although we are particularly interested in the finite sample rejection rates of the different versions of the overidentification test of the asset pricing restrictions and DM tests of the prob-

lematic cases, we also look at the distribution of the estimators of the different prices of risk. To do so, we have created "bicorne plots", which combine a kernel density estimate on top of a box plot. We use vertical lines to 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.

6.1 Baseline design

We set the mean of the risk factor to 1 in order to distinguish between centred and uncentred second moments in our experiment. We also set its standard deviation to 1 without loss of generality. Finally, we set the maximum Sharpe ratio achievable with excess returns to 0.5 and choose the R^2 of the regression of the factor on the excess returns to be 0.1. As in Burnside's (2012) related simulation exercise, all the underlying random variables are independent and identically distributed over time as multivariate Gaussian vectors.

We report the rejection rates of the different overidentification tests that rely on the critical values of a chi-square with 7 degrees of freedom in Tables 3 ($T = 50$) and G3 ($T = 500$). Given that the performance of two-step and iterated GMM is broadly similar, we will focus most of our comments on their differences with CU.

(TABLE 3)

The most striking feature of those tables is the high rejection rates of the multistep J tests of the centred SDF moment conditions (2) asymmetrically normalized in terms of τ. These substantial overrejections are surprising since in this design the population Jacobians have full rank by construction. As expected, the size distortions are mitigated when $T = 500$, but the differences with the other tests still stand out. The Monte Carlo results in Burnside (2012) indicate a lower degree of over-rejection for the same moment conditions, which is probably due to the use of a sequential GMM procedure that fixes the factor mean to its sample counterpart. His implementation is widely used in the literature because of its linearity in τ when combined with multiple step GMM (see e.g. section 13.2 in Cochrane (2001a)), although Parker and Julliard (2005) and Yogo (2006) use optimal GMM in this context.

In contrast, the behavior of the multistep implementations of the J test of the centred SDF moment conditions (2) with a symmetric normalization is similar to the uncentred SDF and regression tests.

Tables 3 and G3 also report DM tests of the null hypothesis of an uncorrelated factor that we derived in section 4.4.1. As expected, we find high rejection rates, especially for $T = 500$.

As for the parameter estimators, the bicorne plots for the prices of risk in Figures 4 indicate that the three GMM estimators of υ and λ are rather similar for $T = 50$. In contrast, the CU estimates of τ are more disperse than their multistep counterparts, which on the other hand show substantial biases.

(FIGURE 4)

When the sample size increases to $T = 500$, CU and the other GMM implementations behave very similarly except for τ (see Figure G2).

6.2 Uncorrelated factor

In this case, we reduce the R^2 of the regression of the pricing factor on the excess returns all the way to 0, but leave the other DGP characteristics unchanged.

Tables 4 and G4 report the rejection rates for this design. Once again the most striking feature is the high rejection rates of the multistep J tests of the centred SDF moment conditions (2) asymmetrically normalized in terms of τ. Unlike what happens in the baseline design, though, those rejection rates do not converge to the nominal values for $T = 500$, which is not surprising given the failure of the GMM regularity conditions discussed in section 4.4.1 (see also Burnside (2012) for related evidence). In contrast, CU tends to underreject slightly for $T = 50$ but the distortion disappears with $T = 500$. As for the other J tests, they usually have rejection rates higher than size, especially the asymmetric uncentred SDF version.

(TABLE 4)

Table 4 also reports the DM test of the null hypothesis of an uncorrelated factor, which is true in this design. We find that the rejection rates are too high in the case of the zero SDF mean null when $T = 50$, but they converge to the nominal size for $T = 500$ in Table G4. We leave for further research the use of bootstrap methods to improve the finite sample properties of the DM tests.

The bicorne plots for the prices of risk shown in Figures 5 and G3 clearly indicate that the biggest difference across the GMM implementations corresponds to τ. In this sense, the sampling distribution of the CU estimator seems to reflect much better the lack of a finite true parameter value. In contrast, both two-step and iterated GMM may give the misleading impression that there is a finite true value when $T = 50$, and they still generate a bimodal bicorne plot with a

substantially lower dispersion when the sample size increases to $T = 500$ (see Hillier (1990) for related evidence in the case of single equation IV). In addition, all the estimators of λ show clear bimodality, which again reflects that this parameter does not have a finite true value either.

(FIGURE 5)

On the other hand, the three GMM estimators of υ behave reasonably well. Regarding ψ and δ, the CU estimators are more disperse, but once again they avoid the biases that plague the multistep estimators.

6.3 A missing risk factor

So far we have seen that GMM asymptotic theory provides a reliable guide for the CU version of the J test when the moment conditions hold, and the same applies to the CU parameter estimator when there exists a finite true value. In contrast, standard asymptotics seems to offer a poor guide to the finite sample rejection rates of the tests that rely on two step and iterated GMM applied to asymmetric normalizations, even in non-problematic cases. In addition, the sampling distributions of the multistep parameter estimators fail to properly reflect the inexistence of a finite parameter value in problematic cases, unlike what happens with single step estimators.

But it is also of interest to analyze the behavior of the different testing procedures when in effect the true SDF that prices all primitive assets in the economy depends on a second factor that the econometrician does not consider. To capture this situation, we simply change the baseline design by setting the Hansen-Jagannathan distance to 0.2.

Table 5 reports the rejection rates of the versions of the J tests that we have considered all along in this third design. Given the size distortions documented for the baseline case, it is not surprising that the CU test has lower rejection rates than the multistep tests, with the asymmetric centred SDF versions standing out again. However, the rejection rates become very similar once we adjust them for their nominal sizes under the null.

(TABLE 5)

Although those size-adjusted rates suggest low power, this is mostly due to the rather small value of the Hansen-Jagannathan distance we have chosen and the small sample size. For the same Hansen-Jagannathan distance, the rejection rates become very high when $T = 500$ (see Table G5). Moreover, the raw rejection rates of the different tests are similar for $T = 500$, which reflects the smaller size distortions in large samples.

7 Conclusions

There are two main approaches to evaluate linear factor pricing models in empirical finance. The oldest method relies on regressions of excess returns on factors, while the other more recent method relies instead on the SDF representation of the model. In turn, there are two variants of each approach, one that uses centred moments and another one which does not. In addition, an empirical researcher has to choose a specific normalization, and she can also transform her moment conditions to improve their interpretation or eliminate some exactly identified parameters. Given that such an unexpectedly large number of different procedures may lead to different empirical conclusions, it is perhaps not surprising that there has been some controversy about which approach is most adequate.

In this context, our paper shows that if we use single step methods such as CU-GMM instead of standard two-step or iterated GMM procedures, then all these procedures provide the same estimates of prices of risk, overidentifying restrictions tests, and pricing errors irrespective of the validity of the model, and regardless of the number asset payoffs and the sample size. In this way, we eliminate the possibility that different researches report potentially contradictory results with the same data set.

Our numerical equivalence results hold for any combination of traded and non-traded factors. We also show that if one uses single step methods, the empirical evidence in favour or against a particular valuation model is not affected by the addition of an asset with non-zero cost for the purposes of pinning down the scale of the SDF. Thus, we would argue that in effect there is a single optimal GMM procedure to empirically evaluate asset-pricing models.

For the benefit of practitioners, we also develop simple, intuitive consistent parameter estimators that can be used to obtain good initial conditions for single step methods, and which will be efficient for elliptically distributed returns and factors. Interestingly, these consistent estimators also coincide with the GMM estimators recommended by Hansen and Jagannathan (1997), which use as weighting matrix the second moment of returns.

Importantly, we propose several distance metric tests that empirical researchers should systematically report in addition to the J test to detect those situations in which the moment conditions are compatible with SDFs that are unattractive from an economic point of view. In particular, we propose tests of the null hypotheses that the mean of the SDF is 0, which corresponds to a risk factor uncorrelated with the vector of excess returns, and the intercept of the SDF is 0, which arises with orthogonal factors.

We illustrate our results with the currency portfolios constructed by Lustig and Verdelhan (2007). We consider some popular linear factor pricing models: the CAPM and linearized

versions of the Consumption CAPM, including the Epstein and Zin (1989) model in appendix A. Our findings clearly point out that the conflict among criteria for testing asset pricing models that we have previously mentioned is not only a theoretical possibility, but a hard reality. Nevertheless, such a conflict disappears when one uses single step methods.

A different issue, though, is the interpretation of the restrictions that are effectively tested. In this sense, our results confirm Burnside's (2011) suggestion that the discrepancies between traditional estimators are due to the fact that the US domestic risk factors seem poorly correlated with currency returns. In this regard, we find that if we force the CAPM to price the market portfolio, then we reject the asset pricing restrictions.

Nevertheless, the numerical coincidence of the different procedures does not necessarily imply that single step inferences are more reliable than their multistep counterparts. For that reason, we also conduct a detailed simulation experiment which shows that GMM asymptotic theory provides a reliable guide for the CU version of the J test when the moment conditions hold, and the same applies to the CU parameter estimator when there exists a finite true value. In fact, the same is true of all GMM implementations based on symmetric normalizations. In contrast, standard asymptotics seem to offer a poor guide to the finite sample rejection rates of those tests that rely on two-step and iterated GMM applied to asymmetric normalizations, even in non-problematic cases. In addition, the sampling distributions of the multistep parameter estimators fail to properly reflect the inexistence of a finite parameter value in problematic cases, unlike what happens with single step estimators.

From the econometric point of view, it would be useful to study in more detail possible ways of detecting the identification failures in asset pricing models with multiple factors discussed by Kan and Zhang (1999) and many others. In a follow up project (Manresa, Peñaranda and Sentana (2014)), we are currently exploring the application to linear factor pricing models of the underidentification tests recently proposed by Arellano, Hansen and Sentana (2012).

From the empirical point of view, an alternative application of our numerical equivalence results would be the performance evaluation of mutual and hedge funds. This literature can also be divided between papers that rely on regression methods, such as Kosowski et al. (2006), and papers that rely on SDF methods, such as Dahlquist and Soderlind (1999) and Farnsworth et al. (2002).

Undoubtedly, both these topics constitute interesting avenues for further research.

References

Almeida, C. and R. Garcia, "Assessing misspecified asset pricing models with empirical likelihood estimators", *Journal of Econometrics* 170 (2012), 519–537.

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.

Arellano, M., *Panel Data Econometrics* (Oxford: Oxford University Press, 2003).

Arellano, M., L.P. Hansen and E. Sentana, "Underidentification?", *Journal of Econometrics* 170 (2012), 256–280.

Bansal, R., D. Kiku and A. Yaron, "Risks for the long run: Estimation with time aggregation", mimeo, College of Business at Illinois (2012).

Beatty, T.K., J.T. LaFrance and M. Yang, "A simple Lagrange multiplier F-test for multivariate regression models", University of California, Berkeley, Department of Agricultural and Resource Economics CUDARE Working Paper 996 (2005).

Breeden, D.T., "An intertemporal asset pricing model with stochastic consumption and investment opportunities", *Journal of Financial Economics* 7 (1979), 265–296.

Burnside, C., M. Eichenbaum, I. Kleshchelski and S. Rebelo, "The returns to currency speculation", NBER Working Paper 12489 (2006).

Burnside, C., "The cross-section of foreign currency risk premia and consumption growth risk: Comment", *American Economic Review* 101 (2011), 3456–3476.

Burnside, C., "Identification and inference in linear stochastic discount factor models", mimeo, Duke University (2012).

Campbell, J.Y., S. Gilgio, and C. Polk, "Hard times", forthcoming in the *Review of Asset Pricing Studies* (2012).

Campbell, J.Y., A.W. Lo and A.C. MacKinlay, *The Econometrics of Financial Markets* (Princeton: Princeton University Press, 1997).

Chamberlain, G., "A characterization of the distributions that imply mean-variance utility functions", *Journal of Economic Theory* 29 (1983a), 185–201.

Chamberlain, G., "Funds, factors, and diversification in arbitrage pricing models", *Econometrica* 51 (1983b), 1305–1323.

Cochrane, J.H., *Asset Pricing*, (Princeton: Princeton University Press, 2001a).

Cochrane, J.H., "A rehabilitation of stochastic discount factor methodology", mimeo, GSB University of Chicago (2001b).

Cochrane, J. H., and J. Saa-Requejo, "Beyond arbitrage: good-deal asset price bounds in incomplete markets", *Journal of Political Economy* 108 (2000), 79–119.

Dahlquist, M. and P. Söderlind, "Evaluating portfolio performance with stochastic discount factors", *Journal of Business* 72 (1999), 347–83.

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

Dovonon, P. and Renault, E., "Testing for common GARCH factors", *Econometrica* 81 (2013), 2561–2586.

Epstein, L.G. and S.E. Zin, "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework", *Journal of Political Economy* 57 (1989), 937–969.

Fama, E.F. and K.R. French, "Common risk factors in the returns on stock and bonds", *Journal of Financial Economics* 33 (1993), 3–56.

Fang, K.-T., S. Kotz and K.-W. Ng, *Symmetric Multivariate and Related Distributions*, (London: Chapman and Hall, 1990).

Farnsworth, H., W. Ferson, D. Jackson, and S. Todd, "Performance evaluation with stochastic discount factors", *Journal of Business* 75 (2002), 473–503.

Gibbons, M., "Multivariate tests of financial models: a new approach", *Journal of Financial Economics* 10 (1982), 3–27.

Gibbons, M.R., S.A. Ross and J. Shanken, "A test of the efficiency of a given portfolio", *Econometrica* 57 (1989), 1121–1152.

Gospodinov, N., R. Kan and C. Robotti, "Further results on the limiting distribution of GMM sample moment conditions", *Journal of Business and Economic Statistics* 30 (2012), 494–504.

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, "Implications of security market data for models of dynamic economies", *Journal of Political Economy* 99 (1991), 225–262.

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

Hillier, G.H., "On the normalization of structural equations: properties of direct estimators", *Econometrica* 58 (1990), 1181–1194.

Hodrick, R. J. and Zhang, X., "Evaluating the specification errors of asset pricing models", *Journal of Financial Economics* 62 (2001), 327–376.

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

Jagannathan, R. and Z. Wang, "Empirical evaluation of asset-pricing models: a comparison of the SDF and beta methods", *Journal of Finance* 57 (2002), 2337–2367.

Julliard, C. and A. Ghosh, "Can rare events explain the equity premium puzzle?", *Review of Financial Studies* 25 (2012), 3037–3076.

Kan, R. and C. Robotti, "Specification tests of asset pricing models using excess returns", *Journal of Empirical Finance* 15 (2008), 816–38.

Kan, R. and C. Zhang, "GMM tests of stochastic discount factor models with useless factors", *Journal of Financial Economics* 54 (1999), 103–127.

Kan, R. and G. Zhou, "A critique of the stochastic discount factor methodology", *Journal of Finance* 54 (1999), 1221–1248.

Kan, R. and G. Zhou, "Empirical asset pricing: the beta method versus the stochastic discount factor method", mimeo, University of Toronto (2002).

Kosowski, R., A. Timmermann , R. Wermers, and H. White, "Can mutual fund stars really pick stocks? New evidence from a bootstrap analysis", *Journal of Finance* 61 (2006), 2551–2595.

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

Lintner, J., "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets", *Review of Economics and Statistics* 47 (1965), 13–37.

Lucas, R.E., "Asset prices in an exchange economy", *Econometrica* 46 (1978), 1429–1446.

Lustig, H. and A. Verdelhan, "The cross-section of foreign currency risk premia and consumption growth risk", *American Economic Review* 97 (2007), 89–17.

Lustig, H. and A. Verdelhan, "The cross-section of foreign currency risk premia and consumption growth risk: Reply", *American Economic Review* 101 (2011), 3477–3500.

MacKinlay, A.C. and M.P. Richardson, "Using generalized method of moments to test mean-variance efficiency", *Journal of Finance* 46 (1991), 511–527.

Magnus, J.R. and H. Neudecker, *Matrix Differential Calculus with Applications to Econometrics* (West Sussex: Wiley, 1988).

Manresa, E., Peñaranda, F. and Sentana, E., "Empirical evaluation of underidentified asset pricing models", work in progress, CEMFI (2014).

Mossin, J., "Equilibrium in a capital asset market", *Econometrica* 34 (1966), 768–783.

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.

Owen, J. and R. Rabinovitch, "On the class of elliptical distributions and their applications to the theory of portfolio choice", *Journal of Finance* 58 (1983), 745–752.

Parker, J.A. and C. Julliard, "Consumption risk and the cross-section of expected returns", *Journal of Political Economy* 113 (2005), 185–222.

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.

Rubinstein, M., "The valuation of uncertain income streams and the pricing of options", *Bell Journal of Economics* 7 (1976), 407–425.

Ross, S.A., "The arbitrage theory of capital asset pricing", *Journal of Economic Theory* 13 (1976), 341–360.

Sargan, J.D., "Identification and lack of identification" *Econometrica* 51 (1983), 1605–1634.

Savin, N.E., "Multiple hypothesis testing", in Z. Griliches and M.D. Intrilligator, eds., *Handbook of Econometrics* vol. II, 827–879 (Amsterdam: North Holland, 1984).

Scheffe, H., "A method of judging all contrasts in the analysis of variance", *Biometrika* 40 (1953), 87–104.

Shanken, J., "On the estimation of beta pricing models", *Review of Financial Studies* 5 (1992), 1–33.

Sharpe, W.F., "Capital asset prices: a theory of market equilibrium under conditions of risk", *Journal of Finance* 19 (1964), 425–442.

Spanos, A., "A parametric model to dynamic heteroskedasticity: the Student's t and related models", mimeo, Virginia Polytechnic Institute and State University (1991).

Stulz, R., "International asset pricing: An integrative survey," in R. Jarrow, M. Maximovich and W. Ziemba, eds. *Handbook of Modern Finance* (Amsterdam: North Holland, 1995).

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

Appendices

A Extensions

Mixed factors

Theoretical discussion

Let us consider a model with two pricing factors in which f_1 is traded, such as the market portfolio, and f_2 is nontraded, such as the growth rate of per capita consumption. An important example would be the linearized CCAPM with Epstein and Zin (1989) preferences. To avoid trivial situations, we assume that the linear span of $(1, f_1, f_2)$ is of dimension 3, which is the multifactor counterpart to the assumption $V(f) > 0$ in single factor models.

Aside from dealing with several factors, the main difference with the analysis in the previous sections is that while f_1 must satisfy the pricing equation (5), f_2 does not. As a result, the uncentred SDF approach will be defined by the $n+1$ moment restrictions:

$$E\left[\begin{array}{c}(a+b_1f_1+b_2f_2)\,\mathbf{r}\\(a+b_1f_1+b_2f_2)\,f_1\end{array}\right]=\mathbf{0}. \tag{A1}$$

These conditions must be supplemented by some scaling of the vector (a, b_1, b_2), such as a symmetric normalization in terms of spherical coordinates or an asymmetric one like $(1, b_1/a, b_2/a)$, both of which leave two free parameters to estimate.

The centred SDF approach relies on the $n+3$ moment conditions:

$$E\left\{\begin{array}{c}[c+b_1(f_1-\mu_1)+b_2(f_2-\mu_2)]\,\mathbf{r}\\[c+b_1(f_1-\mu_1)+b_2(f_2-\mu_2)]\,f_1\\f_1-\mu_1\\f_2-\mu_2\end{array}\right\}=\mathbf{0}, \tag{A2}$$

which again requires a normalization, leaving two free parameters to estimate in addition to the two factor means.

In turn, the centred regression approach can be written in terms of the following $3n$ moment conditions:

$$E\left[\begin{array}{c}\mathbf{r}-\boldsymbol{\beta}_1f_1+d\boldsymbol{\varphi}-c\boldsymbol{\varphi}f_2\\(\mathbf{r}-\boldsymbol{\beta}_1f_1+d\boldsymbol{\varphi}-c\boldsymbol{\varphi}f_2)f_1\\(\mathbf{r}-\boldsymbol{\beta}_1f_1+d\boldsymbol{\varphi}-c\boldsymbol{\varphi}f_2)f_2\end{array}\right]=\mathbf{0}, \tag{A3}$$

where there are $2n+1$ free parameters to estimate after normalizing (c, d).

We can extend the proofs of Propositions 1 and 2 to show that all these three approaches numerically coincide if one uses single-step methods. Moreover, lemma G4 in appendix G shows that all the symmetrically normalized variants of the moment conditions in this section also

have a full column rank Jacobian when risk premia are not all zero and the traded factor alone cannot explain them. As a result, if the other GMM regularity conditions are satisfied, both the unique single step overidentification test and the corresponding multistep overidentification tests will be asymptotically distributed as χ^2_{n-1} under the null.

In contrast, there are some special cases analogous to the ones discussed in section 4.4.1 and appendix C in which the Jacobians of some of the asymmetrically normalized moment conditions do not have full rank (see lemma G5 in appendix G). For example, the counterpart to an uncorrelated factor would arise when the residual of projecting f_2 on a constant and f_1 is not correlated with $\mathbf{r}$. In that case, all valid SDFs affine in (f_1, f_2) would also have a 0 mean. Therefore, if we were using the uncentred SDF moment conditions (A1), we could detect this problematic situation by means of a DM test of the additional restriction

$$E\left(a + b_1 f_1 + b_2 f_2\right) = 0$$

expressed in such a way that it is compatible with the asymmetric or symmetric normalization used.

Finally, we may also find underidentified situations analogous to the one described in section 4.4.2 (see Manresa, Peñaranda and Sentana (2014) for further details).

Empirical application

Table A1 contains the results of estimating a linearized version of the CCAPM with Epstein and Zin (1989) preferences with the same dataset as in section 5. This amounts to identifying f_1 with the US market portfolio and f_2 with US per capita consumption growth. Therefore, this model nests both the CAPM and the CCAPM studied in section 5.

(TABLE A1)

Figure G1c in appendix G, which plots the CU-GMM criterion as a function of the vector of uncentred risk prices $\boldsymbol{\delta}$, confirms that we have obtained a global minimum. Given that the common CU-GMM J statistic is 4.93 with a p-value of 66.8%, this mixed factor model is not rejected even though we attempt to price the market portfolio as in the CAPM. In addition, the t-ratios of $\boldsymbol{\delta}$ show that consumption growth rather than the market portfolio seems to be the driving force behind risk premia.

The two-step, iterated and CU-GMM implementations of the (symmetric and asymmetric) uncentred SDF approach provide similar results, with slightly higher differences for the asymmetric centred regression method. Like in the CCAPM tests reported in section 5.2, though, the

results of the two-step implementation of the asymmetric centred SDF approach clearly diverge from both the CU-GMM results and its two-step symmetric normalization, with much lower estimates of the prices of risk and a very large J statistic. Moreover, iterated GMM fails to converge, cycling over four different solutions, which we do not report for the sake of brevity.

The wedge between the results obtained with the asymmetric centred SDF and the other implementations may be caused by the absence of correlation between excess returns on the currency portfolios and the residuals from regressing US consumption growth on a constant and the excess returns on the US market (cf. section 5). If we apply the relevant DM test, we obtain a statistic of 3.07, with a p-value of 8%. Therefore, the seemingly positive evaluation of the consumption based asset pricing model in Table A1 must be interpreted with some care once again.

Adding a gross return

Theoretical discussion

Many empirical studies only include assets with zero cost. As we saw in section 2, this implies that the SDF is only identified up to scale and sign changes. Let us now see what happens if we add an asset whose cost is not 0, as in Hodrick and Zhang (2001), Farnsworth et al. (2002), and section 7.1 of Burnside (2012).

Let us assume that our data are given by the same $n \times 1$ vector of excess returns $\mathbf{r}$, together with an additional gross return R. We focus again on the case of a single factor f to simplify the exposition. In this context, the relevant moment conditions are (1) plus the pricing of the gross return

$$E[(a + bf)\, R - 1] = 0. \tag{A4}$$

Equation (A4) defines another straight line in (a, b) space, whose intersection with the line defined by the moment conditions (1) will uniquely identify a particular point (see Figure A1a). Therefore, it is no longer necessary to rely on an arbitrary normalization to identify the SDF.

(FIGURE A1)

Similarly, if the researcher prefers the centred SDF version of the moment conditions (2), then she can use the additional moment

$$E\{[c + b\,(f - \mu)]R - 1\} = 0. \tag{A5}$$

As Figure A1b shows, this generally leads to point identification without any need for arbitrary normalizations.

In terms of centred regressions, the moment conditions will be (3), plus the two moments that define the projection of the gross return R on a constant and the risk factor:

$$E\left[\begin{array}{c} R-\phi_R-\beta_R f \\ (R-\phi_R-\beta_R f)\, f \end{array}\right]=\mathbf{0}, \tag{A6}$$

where (ϕ_R, β_R) are the unknown intercept and slope of the projection. Once again, we do not need a normalization because (c, d) must satisfy the pricing constraint

$$1 = c\phi_R + d\beta_R.$$

Empirical researchers, though, typically divide this pricing constraint by c, so that it can be expressed as

$$\phi_R = \varkappa\beta_R + \varkappa_R,$$

where $\varkappa_R = 1/c$ can be interpreted as the "zero-beta" return: the expected return of a unit cost asset whose beta with respect to the factor is 0.

If the factor itself is the excess return on a traded asset, then we should add (5) or (7) to the set of moments as we did in section 3. Similarly, we should impose $d = 0$ in the regression approach as we did in (9).

Note that in the case of the uncentred and centred SDF approaches, there is a new moment and a new parameter to estimate, while in the case of the centred regression there are two new moments and two new parameters. Hence, the J test will have the same number of degrees of freedom as when we use excess returns only. In fact, it turns out that the empirical results with and without the gross return coincide if one uses single-step methods:

Proposition 3 *When we add a gross return R to the vector of n excess returns $\mathbf{r}$, then single-step GMM methods yield the same J test regardless of the normalization for a common specification of the characteristics of the HAC weighting matrix, and irrespective of the pricing factor being traded or not. Analogous results apply to the estimates of a, b, c and d or μ for compatible normalizations.*

Therefore, single step inferences in favour or against a model are not affected by the addition of a gross return, unlike what can happen with multistep methods. We can also show that when we add a gross return there are direct counterparts to proposition 1 and 2. As a result, single step methods yield again numerical equivalent results for SDF and regression approaches, and their uncentred and centred variants.

If the usual GMM regularity conditions are satisfied (see lemma G6 in appendix G), both the unique single step overidentification test and the corresponding multistep overidentification tests will be asymptotically distributed as a χ^2 with the same number of degrees of freedom under the null irrespective of whether we include R.

In this sense, it is worth mentioning that the addition of a gross return generally solves the inference problems associated to a nontraded factor that is either uncorrelated or orthogonal to $\mathbf{r}$. In particular, given that we no longer need to normalize the parameters c and b after the addition of R, the Jacobian of the moment conditions (2) and (A5) with respect to (c, b, μ) will have full column rank when $Cov(\mathbf{r}, f) = \mathbf{0}$ unless we also have $Cov(R, f) = 0$. Similar comments apply to the uncentred SDF conditions when $E(\mathbf{r}f) = \mathbf{0}$ provided that $E(Rf) \neq 0$.

Still, there are situations other than $Cov[f, (\mathbf{r}', R)] = \mathbf{0}$ and $E[f(\mathbf{r}', R)] = \mathbf{0}$ in which the Jacobians of the moment conditions do not have full rank (see lemma G7 in appendix G for further details). Intuitively, the problematic situations arise when the straight line corresponding to (A4) in Figure A1 is parallel to the line generated by (1). We could easily develop a DM test in the system (1) and (A4) to detect this situation.

Finally, the underidentified case that we study in section 4.4.2 remains problematic after the addition of the gross return. Although the addition of R provides an additional equation that the parameters must satisfy, this is not enough to pin down a unique point on the (a, b) space when the vector $\mathbf{r}$ does not provide any information to identify (a, b).

Empirical application

Given that an asset with nonzero cost is not readily available in the Lustig and Verdelhan (2007) dataset, we use the Treasury bill return from Kenneth French's web page deflated by the CPI from FRED as a measure of the cost of the short leg of the carry trades. This gross return has a mean of 1.0136, a standard deviation of 0.0208, and a correlation with US consumption growth of 25.7%, which is higher than the highest correlation of the eight currency portfolios with consumption (19.7%). A positive correlation between those two variables is to be expected through the usual incentives to save, even though this real gross return is not conditionally riskless.

Table A2 reports the empirical evaluation of the CCAPM with this additional payoff. Given proposition 3, checking the convergence of the CU-GMM criterion is straightforward.

(TABLE A2)

In particular, the CU J test is the same as in Table 2 and the parameter estimates are also concordant. For example, $-b/a$ and $-b/c$ in Table A2 coincide with δ and τ respectively in Table 2. Similarly, the DM test of the null hypothesis that the SDF has 0 mean provides exactly the same result whether or not we include a gross return.

However, the uncentred and centred SDF approaches, which become numerically equivalent to each other once we add a gross return, reject the model in their multistep GMM implementations. In particular, the p-value of the asymmetric uncentred SDF decreases from around 57% in Table 2 to less than 5% in Table A2. Therefore, a naive researcher looking at these results may conclude that the excess return data on their own did not have enough power to reject the model, but the addition of the gross return was decisive.

Finally, given that proposition 3 also applies to a traded factor model, we have repeated the same exercise replacing consumption growth with the excess returns on the US stock market portfolio. As expected, the CU J statistic is still the same, unlike what happens with multistep GMM.

B Proofs

All proofs consider the multifactor context of appendix E instead of the simplifying single factor set up in the main text. In addition, the proofs of proposition 1, 2 and 3 do not rely on any particular normalization since they are irrelevant for single-step methods, although some of the asymmetric normalizations might be ill-defined for some problematic configurations, as illustrated in section 4.4. Nevertheless, the overidentification tests are always well defined.

In what follows we represent a set of k factors by the vector $\mathbf{f}$, but maintain the assumption that the number of assets exceeds the number of factors ($n > k$). We also replace the vectors $\boldsymbol{\beta}$ and $\boldsymbol{\varphi}$ in section 2 by the $n \times k$ matrices

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

respectively.

Proposition 1:

In order to show that single-step methods yield numerical equivalent parameter estimators and tests, we prove that we can write $\mathbf{h}_R(\mathbf{r}, \mathbf{f}; \mathbf{B})$ in (E12) and $\mathbf{h}_U(\mathbf{r}, \mathbf{f}; a, \mathbf{b})$ in (E10) as parameter dependent linear transformation of each other after a suitable reparameterization and an augmentation with unrestricted moments, so that the associated criterion functions will be numerically identical for any compatible set of parameter values. Specifically, let us define an extended regression system that adds the estimation of $(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ to $\mathbf{h}_R(\mathbf{r}, \mathbf{f}; \mathbf{B})$:

$$\mathfrak{h}_R(\mathbf{r}, \mathbf{f}; \mathbf{B}, \boldsymbol{\mu}, vech(\boldsymbol{\Gamma})) = \begin{bmatrix} \mathbf{h}_R(\mathbf{r}, \mathbf{f}; \mathbf{B}) \\ \mathbf{f} - \boldsymbol{\mu} \\ vech\left(\mathbf{f}\mathbf{f}' - \boldsymbol{\Gamma}\right) \end{bmatrix} = \begin{bmatrix} \mathbf{r} - \mathbf{B}\mathbf{f} \\ vec\left(\left(\mathbf{r} - \mathbf{B}\mathbf{f}\right)\mathbf{f}'\right) \\ \mathbf{f} - \boldsymbol{\mu} \\ vech\left(\mathbf{f}\mathbf{f}' - \boldsymbol{\Gamma}\right) \end{bmatrix}.$$

Importantly, by adding the exactly identified parameters $(\boldsymbol{\mu}, vech(\boldsymbol{\Gamma}))$, $\mathfrak{h}_R(\mathbf{r}, \mathbf{f}; \mathbf{B}, \boldsymbol{\mu}, vech(\boldsymbol{\Gamma}))$ will be numerically equivalent to $\mathbf{h}_R(\mathbf{r}, \mathbf{f}; \mathbf{B})$ in terms of both the estimates of the original parameters $\mathbf{B}$ and the J test.

We are interested in parameter values that yield a nonsingular $\boldsymbol{\Gamma}$. As a result, the system of equations

$$a\boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{b} = \mathbf{0},$$

defines a unique value for $(a, \mathbf{b})$ with $a \neq 0$ for any admissible normalization.

Then we can carry out the following transformations of the system $\mathfrak{h}_R(\mathbf{r}, \mathbf{f}; \mathbf{B}, \boldsymbol{\mu}, vech(\boldsymbol{\Gamma}))$

$$\left[\begin{array}{cccc} a\mathbf{I}_n & \mathbf{b}' \otimes \mathbf{I}_n & a\mathbf{B} & (\mathbf{b}' \otimes \mathbf{B})\mathcal{D} \end{array}\right] \left[\begin{array}{c} \mathbf{r} - \mathbf{B}\mathbf{f} \\ vec\left((\mathbf{r} - \mathbf{B}\mathbf{f})\mathbf{f}'\right) \\ \mathbf{f} - \boldsymbol{\mu} \\ vech\left(\mathbf{f}\mathbf{f}' - \boldsymbol{\Gamma}\right) \end{array}\right] =$$

$$\left[\mathbf{r}\left(a + \mathbf{b}'\mathbf{f}\right)\right] - \mathbf{B}\left[a\boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{b}\right] = \mathbf{r}\left(a + \mathbf{b}'\mathbf{f}\right),$$

where $\mathcal{D}$ denotes the usual duplication matrix (see Magnus and Neudecker (1988)). Similarly

$$\left[\begin{array}{cccc} \mathbf{0} & \mathbf{0} & a\mathbf{I}_k & (\mathbf{b}' \otimes \mathbf{I}_k)\mathcal{D} \end{array}\right] \left[\begin{array}{c} \mathbf{r} - \mathbf{B}\mathbf{f} \\ vec\left((\mathbf{r} - \mathbf{B}\mathbf{f})\mathbf{f}'\right) \\ \mathbf{f} - \boldsymbol{\mu} \\ vech\left(\mathbf{f}\mathbf{f}' - \boldsymbol{\Gamma}\right) \end{array}\right] =$$

$$\left[\mathbf{f}\left(a + \mathbf{f}'\mathbf{b}\right)\right] - \left[a\boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{b}\right] = \mathbf{f}\left(a + \mathbf{f}'\mathbf{b}\right).$$

As we mentioned before, single-step methods are numerically invariant to normalization, bijective reparameterizations and parameter-dependent linear transformations of the moment conditions. Therefore, for a given choice of HAC weighting matrix, those methods render the extended regression system $\mathfrak{h}_R(\mathbf{r}, \mathbf{f}; \mathbf{B}, \boldsymbol{\mu}, vech(\boldsymbol{\Gamma}))$ and the system

$$\mathfrak{h}_S(\mathbf{r}, \mathbf{f}; a, \mathbf{b}, \mathbf{B}, vech(\boldsymbol{\Gamma})) = \left[\begin{array}{c} \mathbf{r}(a + \mathbf{f}'\mathbf{b}) \\ \mathbf{f}(a + \mathbf{f}'\mathbf{b}) \\ vec\left((\mathbf{r} - \mathbf{B}\mathbf{f})\mathbf{f}'\right) \\ vech\left(\mathbf{f}\mathbf{f}' - \boldsymbol{\Gamma}\right) \end{array}\right]$$

$$= \left(\begin{array}{cccc} a\mathbf{I}_n & \mathbf{b}' \otimes \mathbf{I}_n & a\mathbf{B} & (\mathbf{b}' \otimes \mathbf{B})\mathcal{D} \\ \mathbf{0} & \mathbf{0} & a\mathbf{I}_k & (\mathbf{b}' \otimes \mathbf{I}_k)\mathcal{D} \\ \mathbf{0} & \mathbf{I}_{nk} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{k(k+1)/2} \end{array}\right) \mathfrak{h}_R(\mathbf{r}, \mathbf{f}; \mathbf{B}, \boldsymbol{\mu}, vech(\boldsymbol{\Gamma}))$$

numerically equivalent. Note that the transformation is nonsingular because the parameter values satisfy $a \neq 0$. In particular, the estimates of $\mathbf{B}$ and $vech(\mathbf{\Gamma})$ are the same, the implied $\boldsymbol{\mu} = -\mathbf{\Gamma}\mathbf{b}/a$ is the same, and so is the J test.

Given the definition of $\mathbf{h}_U(\mathbf{r}, \mathbf{f}; a, \mathbf{b})$ in (E10), the last system can also be expressed as

$$\mathfrak{h}_U(\mathbf{r}, \mathbf{f}; a, \mathbf{b}, \mathbf{B}, vech(\mathbf{\Gamma})) = \begin{bmatrix} \mathbf{h}_U(\mathbf{r}, \mathbf{f}; a, \mathbf{b}) \\ vec((\mathbf{r} - \mathbf{B}\mathbf{f})\mathbf{f}') \\ vech(\mathbf{f}\mathbf{f}' - \mathbf{\Gamma}) \end{bmatrix},$$

where the influence functions added to $\mathbf{h}_U(\mathbf{r}, \mathbf{f}; a, \mathbf{b})$ are exactly identified for $(\mathbf{B}, vech(\mathbf{\Gamma}))$. Thus $\mathfrak{h}_U(\mathbf{r}, \mathbf{f}; a, \mathbf{b}, \mathbf{B}, vech(\mathbf{\Gamma}))$ is numerically equivalent to relying on the first block $\mathbf{h}_U(\mathbf{r}, \mathbf{f}; a, \mathbf{b})$ in terms of both the original parameters $(a, \mathbf{b})$ and the J test. Therefore, single-step methods render the systems $\mathbf{h}_R(\mathbf{r}, \mathbf{f}; \mathbf{B})$ and $\mathbf{h}_U(\mathbf{r}, \mathbf{f}; a, \mathbf{b})$ numerically equivalent.

Given the previous arguments, it is trivial to show the numerical equivalence between the systems $\mathbf{h}_U(\mathbf{r}, \mathbf{f}; a, \mathbf{b})$ in (E10) and $\mathbf{h}_C(\mathbf{r}, \mathbf{f}; c, \mathbf{b}, \boldsymbol{\mu})$ in (E11) because they are related by the reparameterization $c = a + \mathbf{b}'\boldsymbol{\mu}$ and the addition of the exactly identified influence functions $\mathbf{f} - \boldsymbol{\mu}$.

Finally, the numerical equivalence of the pricing errors follows trivially from that of the parameter estimators. □

Lemma 1:

We assume that the vector $\mathbf{x} = (\mathbf{f}', \mathbf{r}')'$ follows an elliptical distribution, and denote the corresponding coefficient of multivariate excess kurtosis as κ, which is equal to $\kappa = 2/(\nu - 4)$ in the case of Student t with ν degrees of freedom, and $\kappa = 0$ under normality (see Fang, Kotz and Ng (1990) and the references therein for further details).

Let us order the estimating functions in (1) and (5) for a multifactor model as

$$\mathbf{h}(\mathbf{x}; \boldsymbol{\delta}) = \begin{bmatrix} \mathbf{f}(1 - \mathbf{f}'\boldsymbol{\delta}) \\ \mathbf{r}(1 - \mathbf{f}'\boldsymbol{\delta}) \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1(\mathbf{f}; \boldsymbol{\delta}) \\ \mathbf{h}_2(\mathbf{r}, \mathbf{f}; \boldsymbol{\delta}) \end{bmatrix},$$

where we are using the asymmetric normalization $(1, b/a)$. Thus, we can define the relevant Jacobian as

$$\mathbf{D} = E\left[\frac{\partial \mathbf{h}(\mathbf{x}; \boldsymbol{\delta})}{\partial \boldsymbol{\delta}'}\right] = \begin{pmatrix} \mathbf{\Gamma} \\ E(\mathbf{r}\mathbf{f}') \end{pmatrix} = \begin{pmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \end{pmatrix}.$$

Similarly, we can decompose the relevant asymptotic covariance matrix as

$$\mathbf{S} = avar\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{h}(\mathbf{x}_t; \boldsymbol{\delta})\right] = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix}.$$

If we apply lemma D1 in Peñaranda and Sentana (2012), then we find

$$\mathbf{S}_{11} = \omega_1 \boldsymbol{\Gamma} + \omega_2 \boldsymbol{\mu}\boldsymbol{\mu}',$$

$$\omega_1 = (1-\mathcal{H})(1+\kappa\mathcal{H}), \qquad \omega_2 = -2(1-\mathcal{H})^2 + \left(3\mathcal{H}^2 - 5\mathcal{H} + 2\right)\kappa,$$

where $\mathcal{H} = E(\mathbf{y})' E^{-1}(\mathbf{y}\mathbf{y}') E(\mathbf{y})$, and

$$\mathbf{S}_{21} = \omega_1 E\left(\mathbf{r}\mathbf{f}'\right) + \omega_2 E(\mathbf{r})\boldsymbol{\mu}'.$$

Thus, we only need to check that condition (C1) in lemma C1 in Peñaranda and Sentana (2012) holds, which in our context becomes

$$\mathbf{D}_2\mathbf{D}_1^{-1}\mathbf{S}_{11} = \mathbf{S}_{21}.$$

This restriction will be satisfied as

$$\mathbf{D}_2\mathbf{D}_1^{-1}\mathbf{S}_{11} = E\left(\mathbf{r}\mathbf{f}'\right)\boldsymbol{\Gamma}^{-1}\left[\omega_1\boldsymbol{\Gamma} + \omega_2\boldsymbol{\mu}\boldsymbol{\mu}'\right] = \omega_1 E\left(\mathbf{r}\mathbf{f}'\right) + \omega_2 E\left(\mathbf{r}\mathbf{f}'\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}'$$

and

$$\omega_1 E\left(\mathbf{r}\mathbf{f}'\right) + \omega_2 E\left(\mathbf{r}\mathbf{f}'\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}' = \omega_1 E\left(\mathbf{r}\mathbf{f}'\right) + \omega_2 E(\mathbf{r})\boldsymbol{\mu}' = \mathbf{S}_{21}$$

because $E(\mathbf{r}) = E\left(\mathbf{r}\mathbf{f}'\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{\mu}$ under the null of a valid SDF. Therefore, the linear combinations of the moment conditions in $E[\mathbf{h}(\mathbf{x};\boldsymbol{\delta})] = \mathbf{0}$ that provide the most efficient estimators of δ will be given by

$$E(\mathbf{f}\mathbf{f}'\boldsymbol{\delta} - \mathbf{f}) = \mathbf{0}.$$

□

Proposition 2:

As in the proof of proposition 1, let us define an extended regression system that adds the estimation of $(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ to the influence functions $\mathbf{g}_R(\mathbf{r}, \mathbf{f}; \mathbf{P}, c, \mathbf{d})$ defined in (E15),

$$\mathfrak{g}_R(\mathbf{r}, \mathbf{f}; \mathbf{P}, c, \mathbf{d}, \boldsymbol{\mu}, vech(\boldsymbol{\Gamma})) = \begin{bmatrix} \mathbf{g}_R(\mathbf{r}, \mathbf{f}; \mathbf{P}, c, \mathbf{d}) \\ \mathbf{f} - \boldsymbol{\mu} \\ vech\left(\mathbf{f}\mathbf{f}' - \boldsymbol{\Gamma}\right) \end{bmatrix} = \begin{bmatrix} \mathbf{r} - \mathbf{P}(c\mathbf{f} - \mathbf{d}) \\ vec\left((\mathbf{r} - \mathbf{P}(c\mathbf{f} - \mathbf{d}))\mathbf{f}'\right) \\ \mathbf{f} - \boldsymbol{\mu} \\ vech\left(\mathbf{f}\mathbf{f}' - \boldsymbol{\Gamma}\right) \end{bmatrix}.$$

We are adding exactly identified parameters $(\boldsymbol{\mu}, vech(\boldsymbol{\Gamma}))$, so that $\mathfrak{g}_R(\mathbf{r}, \mathbf{f}; \mathbf{P}, c, \mathbf{d}, \boldsymbol{\mu}, vech(\boldsymbol{\Gamma}))$ is numerically equivalent to the influence functions in $\mathbf{g}_R(\mathbf{r}, \mathbf{f}; \mathbf{P}, c, \mathbf{d})$ in terms of both the original parameter estimates and the J test.

We are interested in parameter values with a nonsingular $\boldsymbol{\Sigma}$. Hence we can choose $(a, \mathbf{b})$ such that

$$\begin{pmatrix} a \\ \mathbf{b} \end{pmatrix} = \begin{pmatrix} 1 & \boldsymbol{\mu}' \\ \boldsymbol{\mu} & \boldsymbol{\Gamma} \end{pmatrix}^{-1} \begin{pmatrix} c \\ \mathbf{d} \end{pmatrix} = \begin{pmatrix} c\left(1+\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right) - \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{d} \\ \boldsymbol{\Sigma}^{-1}\left(\mathbf{d} - c\boldsymbol{\mu}\right) \end{pmatrix}.$$

Then we can compute the following $n \times 1$ transformation of $\mathfrak{g}_R\left(\mathbf{r}, \mathbf{f}; \mathbf{P}, c, \mathbf{d}, \boldsymbol{\mu}, vech\left(\boldsymbol{\Gamma}\right)\right)$:

$$\left[\begin{array}{cccc} a\mathbf{I}_n & \mathbf{b}' \otimes \mathbf{I}_n & \mathbf{P}\left(ac\mathbf{I}_k - \mathbf{d}\mathbf{b}'\right) & c\left(\mathbf{b}' \otimes \mathbf{P}\right)\mathcal{D} \end{array}\right] \left[\begin{array}{c} \mathbf{r} - \mathbf{P}\left(c\mathbf{f} - \mathbf{d}\right) \\ vec\left(\left(\mathbf{r} - \mathbf{P}\left(c\mathbf{f} - \mathbf{d}\right)\right)\mathbf{f}'\right) \\ \mathbf{f} - \boldsymbol{\mu} \\ vech\left(\mathbf{f}\mathbf{f}' - \boldsymbol{\Gamma}\right) \end{array}\right] =$$

$$\left[\mathbf{r}\left(a + \mathbf{f}'\mathbf{b}\right)\right] + \mathbf{P}\left[\mathbf{d}\left(a + \mathbf{b}'\boldsymbol{\mu}\right) - c\left(a\boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{b}\right)\right] = \mathbf{r}\left(a + \mathbf{f}'\mathbf{b}\right).$$

Accordingly, we can also reparameterize $(c, \mathbf{d})$ in terms of the other parameters in the second block of influence functions and then construct the system

$$\mathfrak{g}_U\left(\mathbf{r}, \mathbf{f}; a, \mathbf{b}, \mathbf{P}, \boldsymbol{\mu}, vech\left(\boldsymbol{\Gamma}\right)\right) = \left[\begin{array}{c} \mathbf{r}\left(a + \mathbf{f}'\mathbf{b}\right) \\ vec\left(\left(\mathbf{r} - \mathbf{P}\left(c\mathbf{f} - \mathbf{d}\right)\right)\left(\mathbf{f} + \mathbf{b}\right)'\right) \\ \mathbf{f} - \boldsymbol{\mu} \\ vech\left(\mathbf{f}\mathbf{f}' - \boldsymbol{\Gamma}\right) \end{array}\right]$$

$$= \begin{pmatrix} a\mathbf{I}_n & \mathbf{b}' \otimes \mathbf{I}_n & \mathbf{P}\left(ac\mathbf{I}_k - \mathbf{d}\mathbf{b}'\right) & c\left(\mathbf{b}' \otimes \mathbf{P}\right)\mathcal{D} \\ \mathbf{b} \otimes \mathbf{I}_n & \mathbf{I}_{nk} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{k(k+1)/2} \end{pmatrix} \mathfrak{g}_R\left(\mathbf{r}, \mathbf{f}; \mathbf{P}, c, \mathbf{d}, \boldsymbol{\mu}, vech\left(\boldsymbol{\Gamma}\right)\right),$$

where

$$\begin{pmatrix} c \\ \mathbf{d} \end{pmatrix} = \begin{pmatrix} 1 & \boldsymbol{\mu}' \\ \boldsymbol{\mu} & \boldsymbol{\Gamma} \end{pmatrix} \begin{pmatrix} a \\ \mathbf{b} \end{pmatrix}.$$

Importantly, given that we are ruling out the trivial solution $\mathbf{b} = \mathbf{0}$ and $a = 0$, the above transformation is always non-singular. Hence, single step methods applied to the influence functions $\mathfrak{g}_U\left(\mathbf{r}, \mathbf{f}; a, \mathbf{b}, \mathbf{P},\ \boldsymbol{\mu}, vech\left(\boldsymbol{\Gamma}\right)\right)$ will provide the same estimates and J test as applied to $\mathfrak{g}_R\left(\mathbf{r}, \mathbf{f}; \mathbf{P}, c, \mathbf{d}, \boldsymbol{\mu}, vech\left(\boldsymbol{\Gamma}\right)\right)$ for a specific choice of HAC estimator. As a result, the estimator of $(c, \mathbf{d})$ obtained from $\mathfrak{g}_R\left(\mathbf{r}, \mathbf{f}; \mathbf{P}, c, \mathbf{d}, \boldsymbol{\mu}, vech\left(\boldsymbol{\Gamma}\right)\right)$ and $(a, \mathbf{b})$ from $\mathfrak{g}_U\left(\mathbf{r}, \mathbf{f}; a, \mathbf{b}, \mathbf{P}, \boldsymbol{\mu}, vech\left(\boldsymbol{\Gamma}\right)\right)$ coincide with their implied counterparts in the other system.

This last system can be related to the influence function $\mathbf{g}_U\left(\mathbf{r}, \mathbf{f}; a, \mathbf{b}\right)$ defined in (E13), where the influence functions that are added are exactly identified for $\left(\mathbf{P}, \boldsymbol{\mu}, vech\left(\boldsymbol{\Gamma}\right)\right)$ given $(a, \mathbf{b})$. Thus $\mathfrak{g}_U\left(\mathbf{r}, \mathbf{f}; a, \mathbf{b}, \mathbf{P}, \boldsymbol{\mu}, vech\left(\boldsymbol{\Gamma}\right)\right)$ is numerically equivalent to relying on $\mathbf{r}\left(a + \mathbf{b}'\mathbf{f}\right)$ in terms of both

the estimates of the common parameters $(a, \mathbf{b})$ and the J test. Therefore, single-step methods render the systems $\mathbf{g}_R(\mathbf{r}, \mathbf{f}; \mathbf{P}, c, \mathbf{d})$ and $\mathbf{g}_U(\mathbf{r}, \mathbf{f}; a, \mathbf{b})$ numerically equivalent.

Given the previous arguments, it is trivial to show the numerical equivalence between the systems $\mathbf{g}_U(\mathbf{r}, \mathbf{f}; a, \mathbf{b})$ in (E13) and $\mathbf{g}_C(\mathbf{r}, \mathbf{f}; c, \mathbf{b}, \boldsymbol{\mu})$ in (E14) because they are related by the reparameterization $c = a + \mathbf{b}'\boldsymbol{\mu}$ and the addition of the exactly identified influence functions $\mathbf{f} - \boldsymbol{\mu}$.

Once again, the numerical equivalence of the pricing errors follows trivially from that of the parameter estimators. □

Lemma 2:

As we have already mentioned, 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. As a result, we need $E(\mathbf{r}\mathbf{f}')$ to have full column rank so that we can define $\boldsymbol{\delta} = -\mathbf{b}/a$ with $a \neq 0$.

We also need to extend the results in appendix D in Peñaranda and Sentana (2012) for elliptical distributions to the case of non-traded factors. The optimal moments are given by the linear combinations $\mathbf{D}'\mathbf{S}^{-1}\bar{\mathbf{h}}_T(\boldsymbol{\theta})$. The uncentred SDF method (1) with the asymmetric normalization $(1, b/a)$ has the following long-run variance under the null

$$avar\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\mathbf{r}\left(1-\mathbf{f}'\boldsymbol{\delta}\right)\right]\right] = \frac{(1+\kappa)H_1+1}{(H_2+1)^2}E(\mathbf{r}\mathbf{r}') - \frac{\kappa H_1 + 2(1-\kappa)}{(H_2+1)^2}E(\mathbf{r})E(\mathbf{r})',$$

where $H_1 = \boldsymbol{\lambda}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}$ and $H_2 = \boldsymbol{\lambda}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}$. This asymptotic variance represents a multifactor extension to elliptical distributions of the Gaussian single factor computations in Jagannathan and Wang (2002).

Given that $\mathbf{D} = -E(\mathbf{r}\mathbf{f}')$ for the asymmetric uncentred SDF method, the optimal moments are then proportional to the linear transformation

$$E(\mathbf{f}\mathbf{r}')\left[E(\mathbf{r}\mathbf{r}') - \omega E(\mathbf{r})E(\mathbf{r})'\right]^{-1}, \qquad \omega = \frac{\kappa H_1 + 2(1-\kappa)}{(1+\kappa)H_1+1}.$$

Computing the inverse, we obtain

$$E(\mathbf{f}\mathbf{r}')\left[E^{-1}(\mathbf{r}\mathbf{r}') + \frac{\omega}{1-\omega E(\mathbf{r})'E^{-1}(\mathbf{r}\mathbf{r}')E(\mathbf{r})}E^{-1}(\mathbf{r}\mathbf{r}')E(\mathbf{r})E(\mathbf{r})'E^{-1}(\mathbf{r}\mathbf{r}')\right]$$

and imposing the null hypothesis $E(\mathbf{r}) = E(\mathbf{r}\mathbf{f}')\boldsymbol{\delta}$, we get

$$E(\mathbf{f}\mathbf{r}')\left[E^{-1}(\mathbf{r}\mathbf{r}') + \frac{\omega}{1-\omega\boldsymbol{\delta}'E(\mathbf{f}\mathbf{r}')E^{-1}(\mathbf{r}\mathbf{r}')E(\mathbf{r}\mathbf{f}')\boldsymbol{\delta}}E^{-1}(\mathbf{r}\mathbf{r}')E(\mathbf{r}\mathbf{f}')\boldsymbol{\delta}\boldsymbol{\delta}'E(\mathbf{f}\mathbf{r}')E^{-1}(\mathbf{r}\mathbf{r}')\right]$$
$$= \left[\mathbf{I}_k + \frac{\omega}{1-\omega\boldsymbol{\delta}'E(\mathbf{f}\mathbf{r}')E^{-1}(\mathbf{r}\mathbf{r}')E(\mathbf{r}\mathbf{f}')\boldsymbol{\delta}}E(\mathbf{f}\mathbf{r}')E^{-1}(\mathbf{r}\mathbf{r}')E(\mathbf{r}\mathbf{f}')\boldsymbol{\delta}\boldsymbol{\delta}'\right]E(\mathbf{f}\mathbf{r}')E^{-1}(\mathbf{r}\mathbf{r}').$$

Since the $k \times k$ matrix in brackets has full rank, we can conclude that the optimal estimator of $\boldsymbol{\delta}$ solves the sample moments

$$\frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{r}_t^{+}\left(1-\mathbf{f}_t'\boldsymbol{\delta}_T\right)\right]=\mathbf{0},$$

with

$$\mathbf{r}_t^{+}=E\left(\mathbf{f}\mathbf{r}'\right)E^{-1}\left(\mathbf{r}\mathbf{r}'\right)\mathbf{r}_t.$$

Finally, note that to implement this optimal estimator in practice, we need consistent estimators of $E\left(\mathbf{f}\mathbf{r}'\right)$ and $E\left(\mathbf{r}\mathbf{r}'\right)$, which we can easily obtain from their unrestricted sample counterparts. □

Proposition 3:

When we add a gross return to the uncentred SDF influence functions (E10) or (E13) but we retain the same normalization for a and $\mathbf{b}$, there is an additional influence function

$$\left(a+\mathbf{b}'\mathbf{f}\right)R-q \tag{B7}$$

and a new parameter q that captures the scale of the SDF $(a/q)+\left(\mathbf{b}'/q\right)\mathbf{f}$. Similarly, the centred SDF influence functions become (E11) or (E14) plus

$$[c+\mathbf{b}'\left(\mathbf{f}-\boldsymbol{\mu}\right)]R-q. \tag{B8}$$

In contrast, the centred regression method adds to (E12) or (E15) the $k+1$ new influence functions

$$\begin{bmatrix} R-\phi_R-\boldsymbol{\beta}_R'\mathbf{f} \\ vec\left(\left(R-\phi_R-\boldsymbol{\beta}_R'\mathbf{f}\right)\mathbf{f}'\right) \end{bmatrix},$$

which depend on the $k+1$ new parameters $(\phi_R,\boldsymbol{\beta}_R)$ because q is given by $c\phi_R+\mathbf{d}'\boldsymbol{\beta}_R$, so that it collapses to $q=c\phi_R$ when $\mathbf{f}$ represents excess returns on traded factors. Given that we are adding unrestricted moments in all three instances, the original parameter estimators will be unaffected, and the optimal criterion function will remain the same. But since single-step methods are invariant to reparameterizations and parameter dependent linear transformations of the influence functions, the same results hold even if we use the $q=1$ normalization in (A4), unlike what happens with multistep methods. □

Table 1: Empirical evaluation of the CAPM

	CU	Iterated	2S
Uncentred SDF - Symmetric normalization (ψ)			
Market	-0.204	-0.157	-0.168
	(0.022)	(0.023)	(0.034)
J test	18.959	21.809	6.536
	(0.015)	(0.005)	(0.587)
Centred SDF - Symmetric normalization (υ)			
Market	-0.059	-0.025	-0.048
	(0.020)	(0.015)	(0.040)
J test	18.959	22.037	6.463
	(0.015)	(0.005)	(0.595)
Uncentred SDF - Asymmetric normalization (δ)			
Market	4.826	4.534	4.455
	(0.542)	(0.541)	(0.518)
J test	18.959	19.299	26.844
	(0.015)	(0.013)	(0.000)
Centred SDF - Asymmetric normalization (τ)			
Market	16.945	3.290	2.724
	(17.340)	(1.228)	(1.072)
J test	18.959	46.880	48.957
	(0.015)	(0.000)	(0.000)
Regression (μ)			
Market	0.148	0.142	0.118
	(0.014)	(0.017)	(0.012)
J test	18.959	22.907	5.143
	(0.015)	(0.003)	(0.742)

Note: This table displays estimates of the spherical coordinates (ψ and υ) and the "prices of risk" (δ, τ or μ) with standard errors in parenthesis, as well as the J tests with p-values in parenthesis. We implement each method by continuously updated (CU), iterated and two-step (2S) GMM. The payoffs to price are the annual excess returns on the 8 Lustig-Verdelhan currency portfolios (1953-2002).

Table 2: Empirical evaluation of the (linearized) CCAPM

	CU	Iterated	2S
Uncentred SDF - Symmetric normalization (ψ)			
Nondurables	-0.020	-0.020	-0.020
	(0.002)	(0.002)	(0.002)
J test	5.663	5.784	6.588
	(0.580)	(0.565)	(0.473)
Centred SDF - Symmetric normalization (υ)			
Nondurables	-0.002	-0.002	-0.002
	(0.001)	(0.001)	(0.001)
J test	5.663	5.784	6.589
	(0.580)	(0.565)	(0.473)
Uncentred SDF - Asymmetric normalization (δ)			
Nondurables	49.507	48.835	48.85
	(4.166)	(4.332)	(4.322)
J test	5.663	5.691	5.711
	(0.580)	(0.576)	(0.574)
Centred SDF - Asymmetric normalization (τ)			
Nondurables	438.769	115.428	120.114
	(572.698)	(47.746)	(18.067)
J test	5.663	16.925	91.626
	(0.580)	(0.018)	(0.000)
Centred regression - Asymmetric normalization (λ)			
Nondurables	0.056	0.056	0.024
	(0.021)	(0.016)	(0.008)
J test	5.663	5.677	4.303
	(0.580)	(0.578)	(0.744)

CU tests of problematic cases

Uncorrelated f	Underidentification
2.726	53.039
(0.099)	(0.000)

Note: This table displays estimates of the spherical coordinates (ψ and υ) and the "prices of risk" (δ, τ or λ) with standard errors in parenthesis, as well as the J tests with p-values in parenthesis. We implement each method by continuously updated (CU), iterated and two-step (2S) GMM. The J tests are complemented with CU tests of some problematic cases. The payoffs to price are the annual excess returns on the 8 Lustig-Verdelhan currency portfolios (1953-2002).

Table 3: Rejection rates in the baseline design (T=50)

	Nominal size		
	10	5	1
J tests			
CU	11.67	5.14	0.64
Uncentred SDF - Symmetric normalization			
Iterated	15.63	8.74	2.09
2S	16.42	9.28	2.40
Centred SDF - Symmetric normalization			
Iterated	15.62	8.42	1.97
2S	22.51	14.05	4.68
Uncentred SDF - Asymmetric normalization			
Iterated	16.81	9.67	2.62
2S	17.69	10.28	2.90
Centred SDF - Asymmetric normalization			
Iterated	50.08	40.88	25.93
2S	57.07	47.09	28.73
Centred regression - Asymmetric normalization			
Iterated	13.29	6.70	1.44
2S	13.49	6.95	1.42
CU DM tests of problematic cases			
Uncorrelated f	67.02	57.86	37.55

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 4: Rejection rates in the uncorrelated factor design (T=50)

	Nominal size		
	10	5	1
J tests			
CU	8.8	3.33	0.40
Uncentred SDF - Symmetric normalization			
Iterated	18.48	10.68	2.99
2S	18.34	10.84	3.05
Centred SDF - Symmetric normalization			
Iterated	12.92	6.42	1.18
2S	16.70	9.62	2.42
Uncentred SDF - Asymmetric normalization			
Iterated	26.46	17.05	5.84
2S	25.00	15.83	5.32
Centred SDF - Asymmetric normalization			
Iterated	66.98	56.49	36.79
2S	64.78	52.74	31.83
Centred regression - Asymmetric normalization			
Iterated	10.45	4.52	0.80
2S	12.44	6.08	0.96
CU DM tests of problematic cases			
Uncorrelated f	31.92	22.42	8.56

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 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 5: Rejection rates in the missing factor design (T=50)

	Asymptotic critical values			Monte Carlo critical values		
	10	5	1	10	5	1
CU	17.78	8.39	1.1	15.51	8.08	1.81
Uncentred SDF - Symmetric normalization						
Iterated	24.17	14.21	3.99	16.08	8.52	2.09
2S	24.83	14.90	4.51	16.17	8.06	1.85
Centred SDF - Symmetric normalization						
Iterated	22.59	12.91	3.23	15.12	8.01	1.58
2S	30.65	20.60	7.46	15.39	8.07	1.77
Uncentred SDF - Asymmetric normalization						
Iterated	26.20	16.22	4.84	16.59	8.61	2.34
2S	27.14	16.95	5.13	16.79	8.72	1.94
Centred SDF - Asymmetric normalization						
Iterated	60.71	51.60	34.59	13.85	6.90	1.59
2S	66.74	56.81	37.00	14.12	7.42	1.55
Centred regression - Asymmetric normalization						
Iterated	20.12	11.02	2.38	15.95	8.14	1.84
2S	20.45	11.44	2.62	15.64	7.90	1.91

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 A1: Empirical evaluation of the (linearized) Epstein-Zin model

	CU	Iterated	2S
Uncentred SDF - Symmetric normalization (ψ_1,ψ_2)			
ψ_1	-0.021	-0.020	-0.020
	(0.002)	(0.002)	(0.002)
ψ_2	0.012	0.006	0.007
	(0.012)	(0.011)	(0.009)
J test	4.932	5.405	6.654
	(0.668)	(0.611)	(0.466)
Centred SDF - Symmetric normalization (υ_1,υ_2)			
υ_1	-0.002	-0.002	-0.002
	(0.001)	(0.001)	(0.001)
υ_2	0.012	0.006	0.007
	(0.012)	(0.011)	(0.009)
J test	4.932	5.406	6.922
	(0.668)	(0.611)	(0.437)
Uncentred SDF - Asymmetric normalization (δ_1,δ_2)			
Market	0.593	0.474	0.501
	(0.537)	(0.536)	(0.531)
Nondurables	47.883	48.082	47.441
	(5.343)	(5.543)	(5.128)
J test	4.932	4.988	5.455
	(0.668)	(0.661)	(0.605)
Centred SDF - Asymmetric normalization (τ_1,τ_2)			
Market	4.991		1.575
	(5.792)		(0.583)
Nondurables	402.811		114.235
	(397.268)		(18.215)
J test	4.932		74.335
	(0.668)		(0.000)
Centred regression - Asymmetric normalization (μ,λ)			
Market	0.079	0.072	0.070
	(0.022)	(0.021)	(0.022)
Nondurables	0.057	0.055	0.047
	(0.021)	(0.015)	(0.012)
J test	4.932	5.225	6.043
	(0.668)	(0.633)	(0.535)

Note: This table displays estimates of the spherical coordinates and the "prices of risk" with standard errors in parenthesis, as well as the J tests with p-values in parenthesis. We implement each method by continuously updated (CU), iterated and two-step (2S) GMM. Iterated GMM does not converge for the asymmetric normalization of the centred SDF . The payoffs to price are the annual excess returns on the 8 Lustig-Verdelhan currency portfolios (1953-2002).

Table A2: Empirical evaluation of the (linearized) CCAPM with a gross return

	CU	Iterated	2S
Uncentred SDF - Parameters (a,b)			
Constant	8.838	2.848	2.535
	(5.139)	(0.866)	(0.466)
Nondurables	-437.522	-117.692	-98.976
	(276.178)	(48.185)	(29.736)
J test	5.663	16.346	29.024
	(0.580)	(0.022)	(0.000)
Centred SDF - Parameters (c,b)			
Mean	0.997	0.995	0.995
	(0.325)	(0.010)	(0.202)
Nondurables	-437.522	-117.692	-98.976
	(276.178)	(48.185)	(29.739)
J test	5.663	16.346	29.024
	(0.580)	(0.022)	(0.000)
Centred regression - Parameters (x_0,λ)			
Zero-beta	1.003	1.003	1.006
	(0.013)	(0.011)	(0.012)
Nondurables	0.056	0.056	0.065
	(0.021)	(0.016)	(0.019)
J test	5.663	5.677	6.177
	(0.580)	(0.578)	(0.531)

Note: This table displays estimates of several methods with standard errors in parenthesis, as well as the J tests with p-values in parenthesis. We implement each method by continuously updated (CU), iterated and two-step (2S) GMM. The payoffs to price are the annual excess returns on the 8 Lustig-Verdelhan currency portfolios, and the real Treasury bill return (1953-2002).

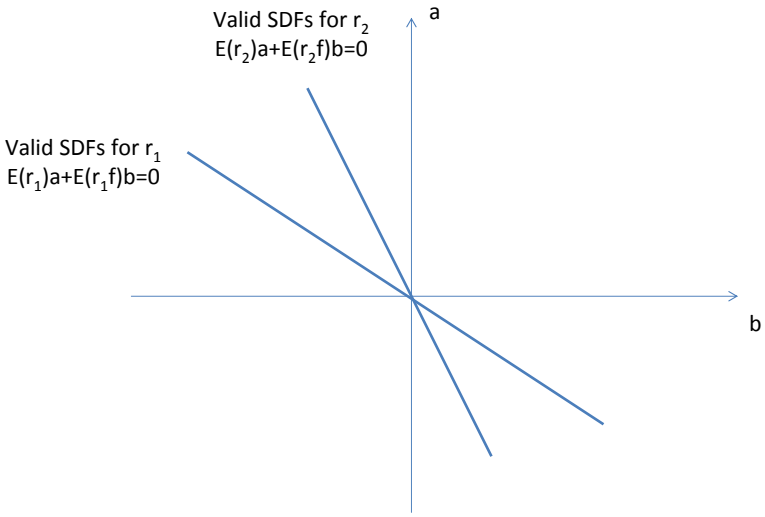


Figure 1a : Market segmentation

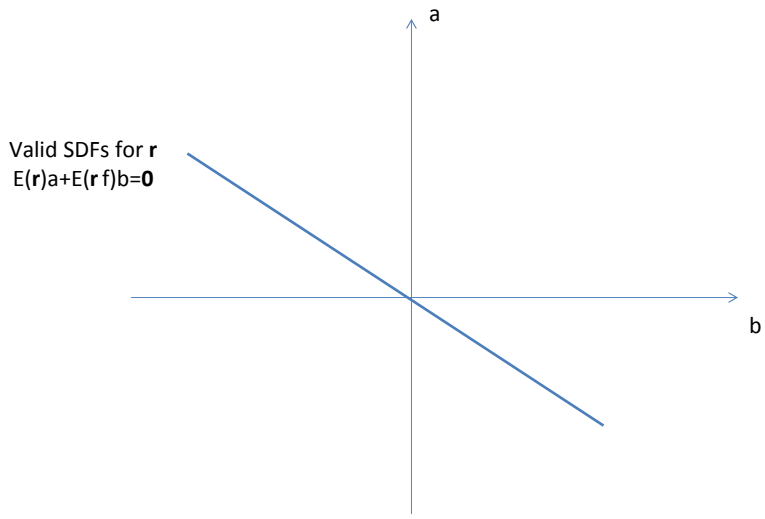


Figure 1b : Market integration

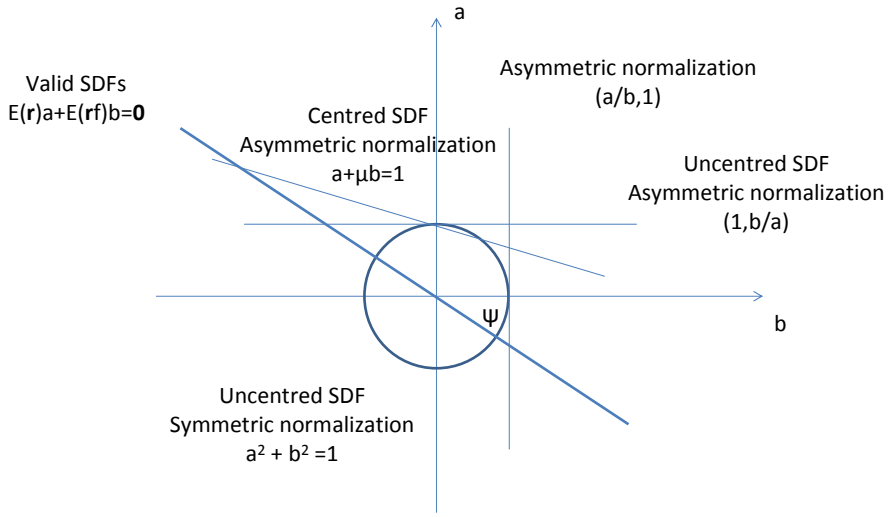


Figure 2a : Alternative normalizations in (a,b) space

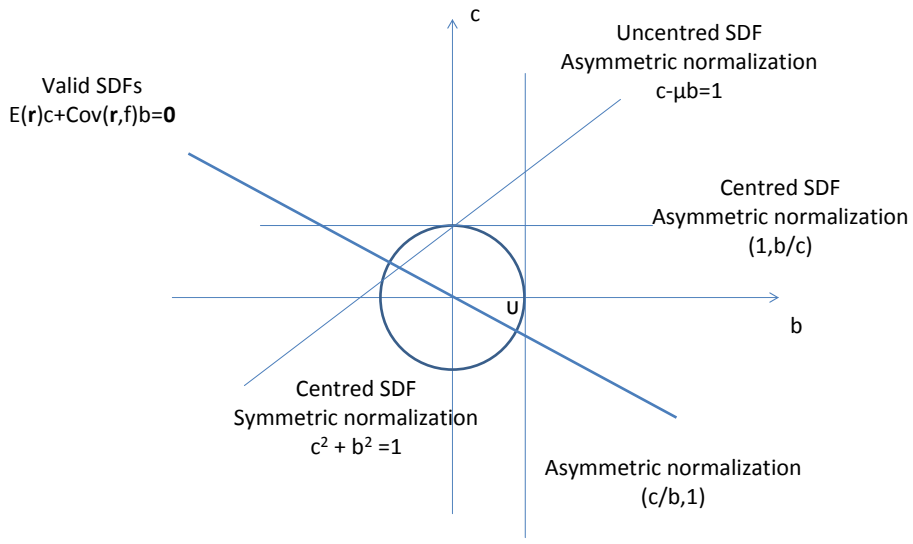


Figure 2b : Alternative normalizations in (c,b) space

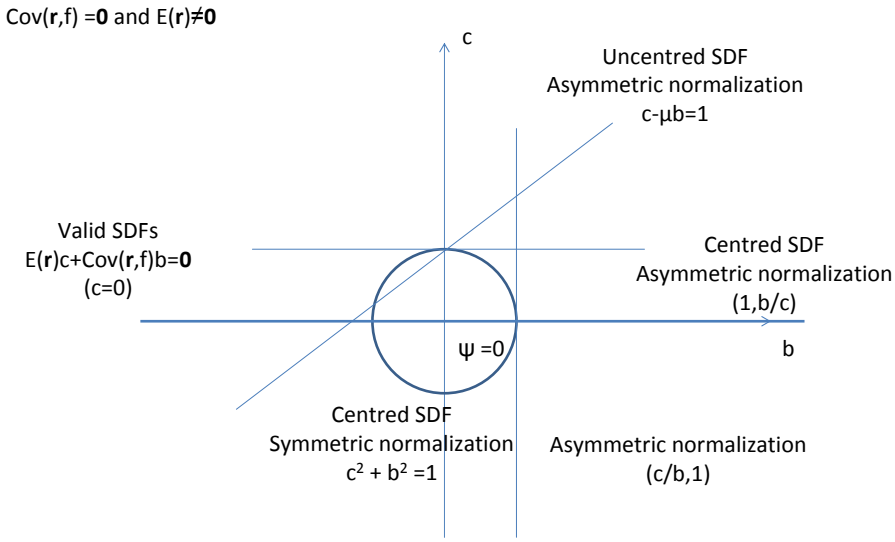


Figure 3a : Uncorrelated factor in (c,b) space

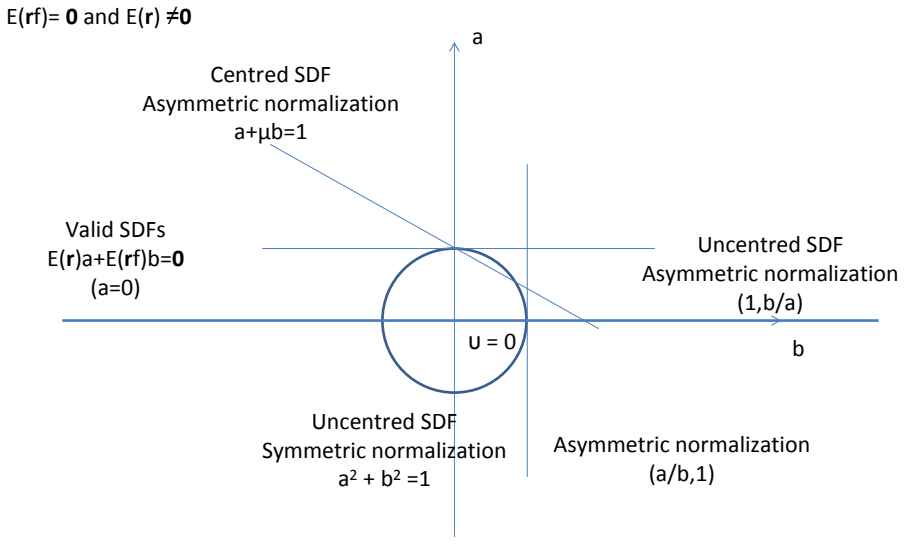
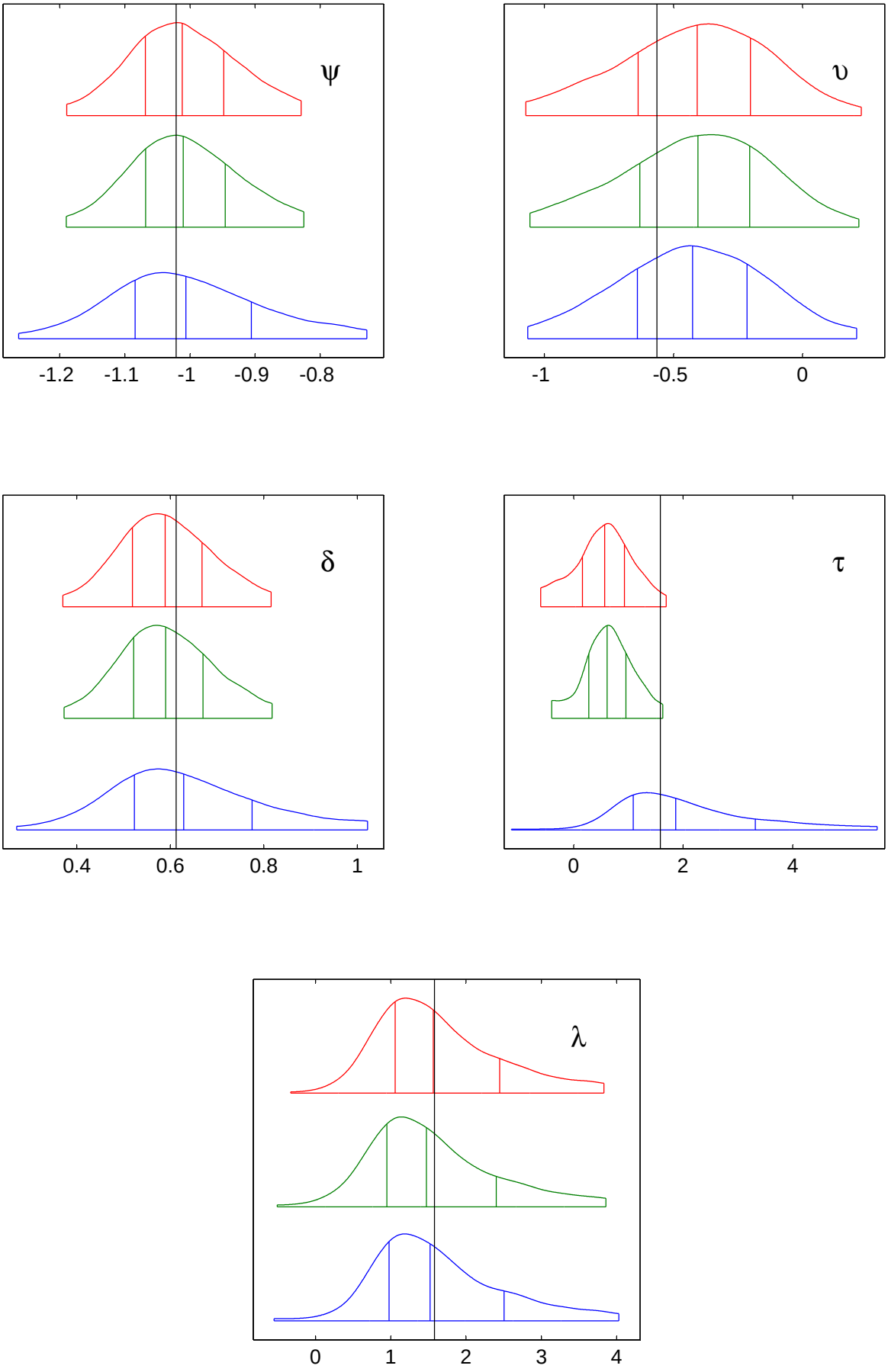


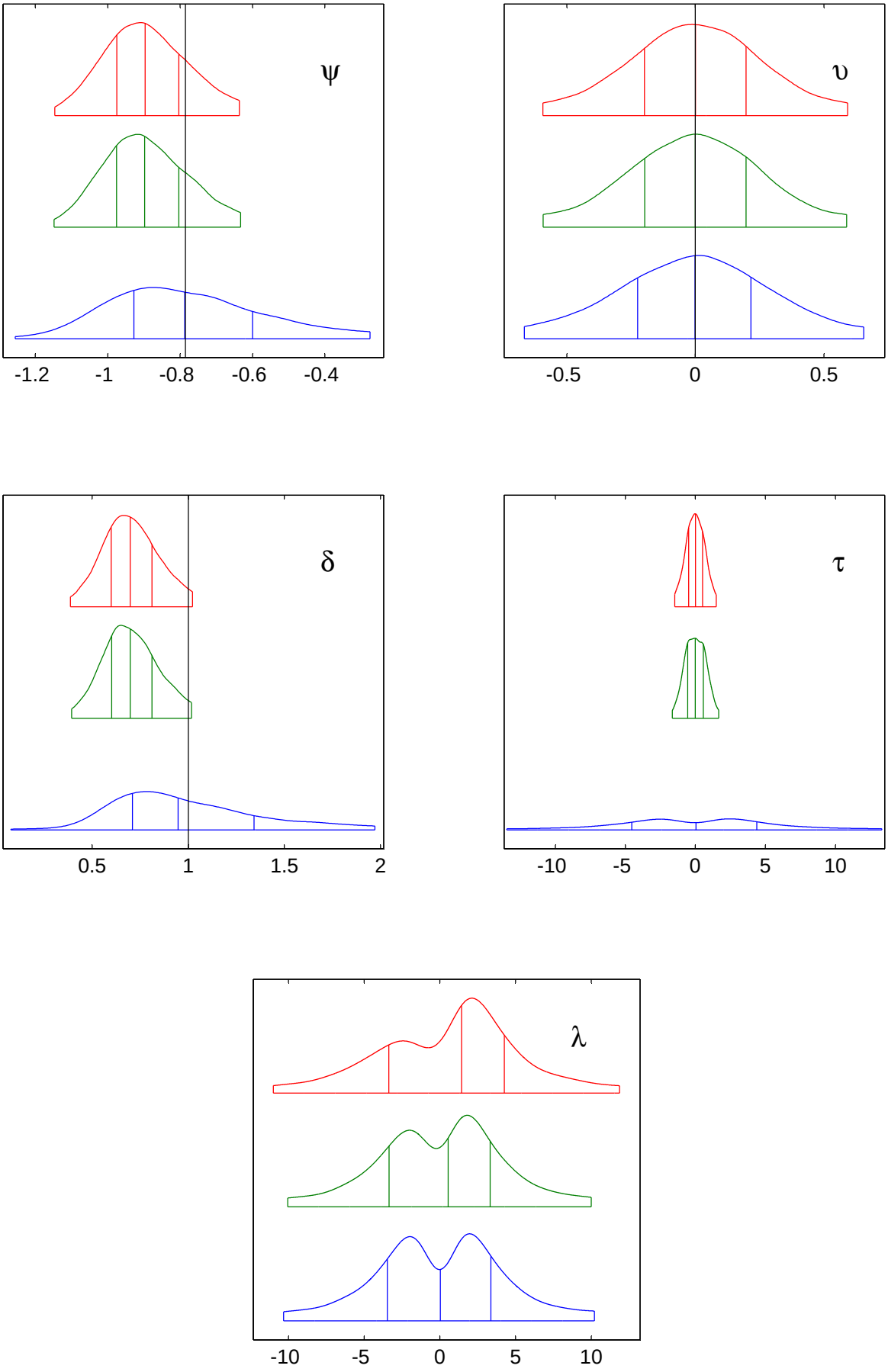
Figure 3b : Orthogonal factor in (a,b) space

Figure 4: Parameter estimates in baseline 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 5: Parameter estimates in uncorrelated 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.

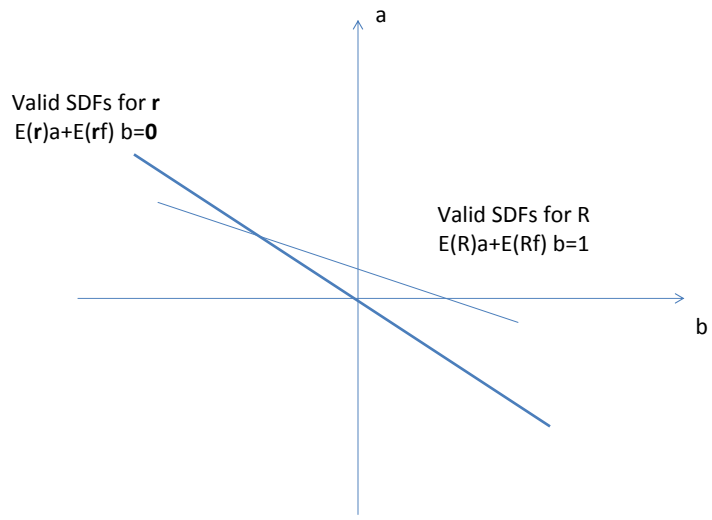


Figure A1a : Adding a gross return in (a,b) space

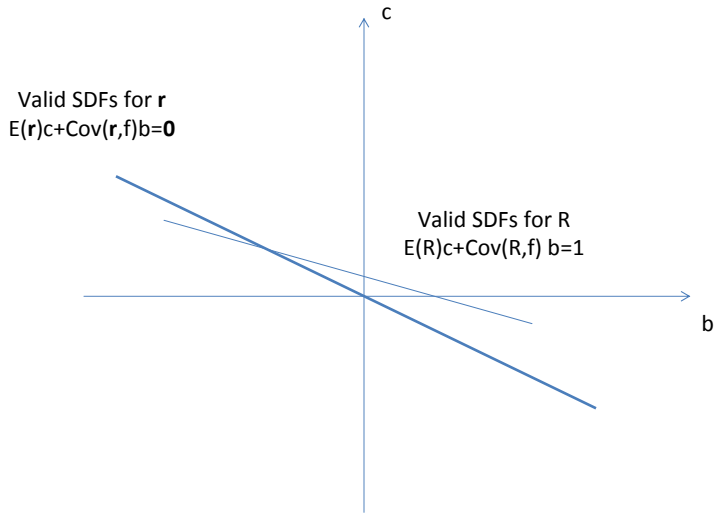


Figure A1b : Adding a gross return in (c,b) space