FEDERAL RESERVE BANK of ATLANTA

CenFIS

Center for Financial Innovation and Stability

A Unifying Approach to the Empirical Evaluation of Asset Pricing Models

Francisco Peñaranda and Enrique Sentana

CenFIS Working Paper 10-03
December 2010

Working Paper Series

CenFIS
Center for Financial Innovation and Stability

A Unifying Approach to the Empirical Evaluation of Asset Pricing Models

Francisco Peñaranda and Enrique Sentana

CenFIS Working Paper 10-03
December 2010

Abstract: Two main approaches are commonly used to empirically evaluate linear factor pricing models: regression and stochastic discount factor (SDF) methods, with centered and uncentered versions of the latter. We show that unlike standard two-step or iterated gerneralized method of moments (GMM) procedures, single-step estimators such as continuously updated GMMs yield numerically identical values for prices of risk, pricing errors, Jensen's alphas, and overidentifying restrictions tests irrespective of the model validity. Therefore, there is arguably a single approach regardless of the factors being traded or the use of excess or gross returns. We illustrate our results with the currency returns constructed by Lustig and Verdelhan (2007).

JEL classification: G11, G12, C12, C13

Key words: CU-GMM, factor pricing models, forward premium puzzle, generalized empirical likelihood, stochastic discount factor

Working Paper Series

The authors thank Abhay Abhyankar, Manuel Arellano, Antonio Díez de los Ríos, Lars Hansen, Raymond Kan, Craig MacKinlay, Cesare Robotti, Rosa Rodríguez, Amir Yaron, participants at the XVIII Finance Forum (Elche) and the XXXV Simposio de la Asociación Española de Economía (Madrid), and audiences at the Atlanta Fed, Princeton University, Universitat Pompeu Fabra, and the Wharton School for helpful comments, suggestions, and discussions. Financial support from the Spanish Ministry of Science and Innovation through grants ECO 2008-03066 (Peñaranda) and ECO 2008-00280 (Sentana) is gratefully acknowledged. Peñaranda also acknowledges the financial support of the Barcelona GSE and the Government of Catalonia and the hospitality of the Center for Financial Innovation and Stability at the Atlanta Fed. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Please address questions regarding content to Francisco Peñaranda, Assistant Professor, Department of Economics and Business, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain, (+34) 93 542 2638, fax (+34) 93 542 1746, francisco.penaranda@upf.edu, or Enrique Sentana, CEMFI, Casado del Alisal 5, 28014 Madrid, Spain, +34 914290551, fax +34 914291056, sentana@cemfi.es.

CenFIS Working Papers from the Federal Reserve Bank of Atlanta are available online at frbatlanta.org/frba/cenfis/. Subscribe online to receive e-mail notifications about new papers.

1 Introduction

Asset pricing theories are concerned with determining the expected returns of assets whose payoffs are risky. Specifically, these models analyse 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 a linear function of 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. 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 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 and another one that does not. Although 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 these approaches are often implemented by means of 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 advanced textbook treatments).

Unfortunately, though, each approach typically yields different estimates of prices of risk, pricing errors and Jensen's alphas, and different values for the overidentifying restrictions test. This begs the question of which approach is best, and there is some controversy surrounding the answer. For example, Kan and Zhou (2000) advocated the use of the regression method over the uncentred SDF method because the former provides more reliable risk premia estimators and more powerful specification tests than the latter. However, Cochrane (2001b) and Jagannathan and Wang (2002) criticised their conclusions on the grounds that they did not consider the estimation of factor means and variances. In this respect, Jagannathan and Wang (2002) showed that if excess returns and factors are independently and identically distributed as a 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 (2007) 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. At the same time, Kan and Robotti (2008, footnote 3) effectively exploit the invariance to coefficient normalisations 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 as part of the objective function.

In this context, the main contribution of our paper is to show the more subtle result that the application to both the regression and SDF approaches of single-step GMM methods, including CU-GMM, gives rise to numerically identical estimates of prices of risk, pricing errors, Jensen's alphas and overidentifying restrictions tests irrespective of the validity of the asset pricing model.

Therefore, one could argue that there is effectively a single method to empirically evaluate asset-pricing models. Although the rationale for our results is the well-known functional invariance of maximum likelihood estimators, our results do not depend on any distributional assumption, the number of assets, the specific combination of traded and non-traded factors, or the sample size, and remain true regardless of whether or not the researcher works with excess returns or gross returns, or the asset pricing restrictions hold. To ease the 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. Similarly, our empirical application includes both single and multifactor models.

Importantly, 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. Although the choice in Hansen and Jagannathan (1997) is reasonable in SDF contexts for the purposes of comparing several misspecified models, it does not have a natural regression counterpart. For analogous reasons, we do not consider the popular two-pass regressions, which do not have a natural SDF counterpart. In any case, those generally

suboptimal GMM estimators fall outside the realm of single-step methods such as CU-GMM, 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 because they do not require to choose the number of iterations, and more importantly, they 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 generalised empirical likelihood estimators over two-step GMM by going beyond the usual first-order asymptotic equivalence results.

In fact, the recent papers by Julliard and Gosh (2008), Almeida and Garcia (2009) or Campbell, Gilgio and Polk (2010) attest the increasing popularity of these modern GMM variants. However, the CU-GMM estimator and other single-step methods 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 (2010) 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 with the currency portfolios constructed by Lustig and Verdelhan (2007). We consider three popular linear factor pricing models: the CAPM, as well as linearlised versions of the Consumption CAPM and the Epstein and Zin (1989) model. 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. A different issue, though, is the interpretation of the restrictions that are effectively tested. In this sense, our results confirm Burnside's (2007) 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 to excess returns will automatically price those returns with a SDF whose mean is 0. Obviously, such a SDF is not satisfactory from an economic point of view, but strictly speaking, the vector of risk premia and the covariances

between excess returns and factors will be proportional. 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 organised as follows. We study the case of traded factors from the theoretical and empirical perspectives in section 2, while in sections 3 and 4 we analyse non-traded factors, and a mixture of traded and non-traded factors, respectively. In section 5 we show that our results hold not only for excess returns but also for gross returns. Finally, we summarise our conclusions in section 6. A brief description of CU-GMM as an example of single-step methods, multifactor models, and formal proofs are gathered in appendices A, B and C, respectively. We also include a supplemental appendix with additional results.

2 Traded factors

Let $\mathbf{r}$ be an $n \times 1$ vector of excess returns. Standard arguments such as lack of arbitrage opportunities or the first order conditions of a representative investor imply that

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

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 factor f, we can express the pricing equation as

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

for some real numbers (a, b).

Although $\mathbf{r}$ only contains assets with 0 cost, which leaves the *scale* and *sign* of m undetermined, we would like our candidate SDF to price other assets with positive prices. Therefore, we require a *scale normalisation* to rule out the trivial solution $(a, b) = (0, 0)$ (see Cochrane (2001a, pp. 256-258)). For example, we could choose the popular asymmetric normalisations $a = 1$ or $E(m) = a + b\mu = 1$, where $\mu = E(f)$. Alternatively, we could choose the symmetric normalisation $a^2 + b^2 = 1$, together with a sign restriction on one of the nonzero coefficients. As we shall see below, this seemingly innocuous issue may have important empirical consequences.

In this section we assume that the pricing factor f is itself the excess return on another

asset, such as the market portfolio in the CAPM.[1] As forcefully argued by Farnsworth et al. (2002) and Lewellen, Nagel and Shanken (2009) among others, the pricing model applies to f too, which means that

$$E\left[\left(a+bf\right)f\right]=0. \tag{2}$$

Following Chamberlain (1983b) we also know 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 a constant and f should be proportional to f. As result, we can equivalently write the above pricing equations as

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

where the parameter vector $\boldsymbol{\beta}$ represents the slopes of the projection.

2.1 Existing approaches

Equations (1) and (2) are particularly amenable to GMM estimation once we choose a normalisation for (a,b). As we mentioned before, there are two widespread asymmetric choices in empirical finance: $a=1$ and $a+b\mu=1$, with the corresponding SDFs typically expressed as $m=1-f\lambda$ and $m=1-\left(f-\mu\right)\tau$, respectively. Kan and Zhou (2000), 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)). Burnside (2007) refers to the first approach as the a-normalisation and to the second one as the ξ-normalisation. We will refer to them instead as the uncentred and centred SDF parametrisations since they differ in their use of either $E\left(\mathbf{r}f\right)$ or $Cov\left(\mathbf{r},f\right)$ in explaining the cross-section of risk premia.

Specifically, the uncentred SDF test relies on the overidentified, $n+1$ linear moment conditions

$$E\left[\begin{array}{c}\mathbf{r}\left(1-f\lambda\right)\\ f\left(1-f\lambda\right)\end{array}\right]=\mathbf{0}, \tag{4}$$

where the only unknown parameter is λ. Given that the last moment condition implies that $E\left(f\right)=E\left(f^{2}\right)\lambda$, we will have that

$$\lambda=\frac{\mu}{\gamma}, \tag{5}$$

where γ is the second moment of the factor, which allows us to interpret λ as a "price of risk" for the factor. Under standard regularity conditions, the overidentifying restrictions (J) test

[1] 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).

will follow an asymptotic chi-square distribution with n degrees of freedom (χ^2_n) when (4) is correctly specified.

In contrast, the centred SDF test relies on the overidentified, $n+2$ non-linear in parameters moment conditions

$$E\left[\begin{array}{c}\mathbf{r}\left(1-\left(f-\mu\right)\tau\right)\\ f\left(1-\left(f-\mu\right)\tau\right)\\ f-\mu\end{array}\right]=\mathbf{0}, \tag{6}$$

where the two unknown parameters are (μ,τ), with the additional moment condition allowing for the estimation of μ. Once again, we can use the middle moment condition to show that:

$$\tau=\frac{\mu}{\sigma^2}, \tag{7}$$

where $\sigma^2=\gamma-\mu^2$ denotes the variance of f, which means that τ also has a "price of risk" interpretation. Not surprisingly, the corresponding J test also converges in distribution to a χ^2_n under correct specification.

The regression (or beta) representation of the pricing model is also amenable to GMM estimation. In particular, we can follow MacKinlay and Richardson (1991) in regarding (3) as $2n$ overidentified, linear moment conditions, where the n unknown parameters are the slope coefficients $\boldsymbol{\beta}$, which under the null coincide with both $E(\mathbf{r}f)/E(f^2)$ and $Cov(\mathbf{r},f)/V(f)$. Therefore, the J test will be asymptotically distributed as χ^2_n under the null.

The regression method identifies μ with the expected excess return of a portfolio whose "beta" is equal to 1. Hence, this parameter may also be interpreted as an alternative "price of risk". To estimate it, we can add $f-\mu$ to (3) as in (6), and simultaneously obtain $\boldsymbol{\beta}$ and μ.

The overidentification tests are regularly complemented by three standard evaluation measures, which correspond to the value of the different moment conditions when the linear factor pricing model is incorrect. In this way, we define Jensen's alphas:

$$\boldsymbol{\alpha}=E\left(\mathbf{r}\right)-\boldsymbol{\beta}E\left(f\right) \tag{8}$$

for the regression method, the pricing errors obtained from the uncentred SDF representation:

$$\boldsymbol{\pi}=E\left(\mathbf{r}\right)-E\left(\mathbf{r}f\right)\lambda,$$

or from the centred SDF representation:

$$\boldsymbol{\psi}=E\left(\mathbf{r}\right)-E\left(\mathbf{r}\left(f-\mu\right)\right)\tau.$$

Under the null hypothesis these three measures should be simultaneously 0, but otherwise their values will be different.

2.2 Numerical equivalence results

As we mentioned in the introduction, Kan and Zhou (2000, 2002), Cochrane (2001b), Jagannathan and Wang (2002), Burnside (2007) and Kan and Robotti (2008) compare some of the aforementioned approaches when researchers rely on traditional, two-step or iterated GMM procedures. In contrast, our main result is that all the methods coincide if one uses instead single-step procedures such as CU-GMM. More formally:

Proposition 1 *If we apply single-step procedures to the uncentred SDF method based on the moment conditions (4), the centred SDF method based on the moment conditions (6), and the 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:*
1) The three overidentification restrictions (J) tests.
2) The direct estimate of the "price of risk" λ from (4), the indirect estimate

$$\lambda = \frac{\sigma^2}{\gamma}\tau,$$

from (6) extended to include γ, and the indirect estimate

$$\lambda = \frac{\mu}{\gamma}$$

from (3) extended to include (μ, γ). Analogous results apply to τ and μ.
3) The estimates of Jensen's alphas in (8) obtained by replacing $E(\cdot)$ by an unrestricted sample average and $\boldsymbol{\beta}$ by their direct estimates obtained from the regression method, or the indirect estimates obtained from SDF methods extended to include $\boldsymbol{\beta}$. Analogous results apply to the alternative pricing errors $\boldsymbol{\pi}$ and $\boldsymbol{\psi}$.

Importantly, these numerical equivalence results do not depend in any sense on the number of assets, the sample size, 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 factorise 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 parametrisation. 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.

Alternatively, we could understand the CU-GMM procedure as being based on the original moment conditions (1) and (2), which are valid for any normalisation of the SDF scale of the

form $a^2 + b^2 \neq 0$. In this light, the equivalence between the two SDF approaches is a direct consequence of the fact that single-step procedures are numerical invariant to normalisation, 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 parameter dependent linear transformations of the moments and reparametrisations (see appendix A).

One drawback of CU-GMM and other GEL procedures, though, is that they induce a non-linearity in the GMM objective function of the uncentred SDF and regression approaches, which may induce multiple local minima. In this sense, the uncentred SDF method has a non-trivial computational advantage because it only contains a single unknown parameter.[2] At the same time, one can also exploit the numerical equivalence of the three approaches to check that a global minimum has been reached. A much weaker convergence test is given by the fact that the value of the criterion function at the CU-GMM estimators cannot be larger than at the iterated GMM estimators (see Hansen, Heaton and Yaron (1996)).

In any case, it is convenient to have consistent initial parameter values. For that reason, we propose 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 maximisation regardless of investors' preferences (see Chamberlain (1983a) and Owen and Rabinovitch (1983)). In particular, if we derive the optimal moment condition for the uncentred SDF model (see Hansen (1982)), then we can immediately show that:

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 estimator of λ obtained from (4) will be given by*

$$\dot{\lambda}_T = \frac{\sum_{t=1}^{T} f_t}{\sum_{t=1}^{T} f_t^2}. \tag{9}$$

Intuitively, this means that under those circumstances (2), which is the moment involving f, exactly identifies the parameter λ, 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 e.g. Fang, Kotz and Ng (1990)), and includes the multivariate normal and Student t distribution as special cases, it is important to stress that $\dot{\lambda}_T$ will remain consistent under linear factor pricing even if the assumptions of serial independence and ellipticity are not totally realistic in practice.[3]

[2] 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, while the regression method the estimation of $n \times k$ factor loadings.

[3] We can also prove that we obtain an estimator of λ that is asymptotically equivalent to (9) 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,

A rather different justification for the estimator (9) is that it coincides with the GMM estimator of λ that we would obtain from (4) if we used as weighting matrix the second moment of the vector of excess returns $(\mathbf{r}, f)$, as recommended by Hansen and Jagannathan (1997).

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.[4]

2.3 Empirical application

Over the last thirty years most 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 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. Then they 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.

Table 1 reports some simple descriptive statistics for those portfolios. Interestingly, the broadly monotonic relationship between the level of interest rates differentials and risk premia provides informal evidence on the failure of uncovered interest rate parity.

---

and therefore elliptical. Intuitively, the reason is that a single sample realisation of such a Haavelmo distribution is indistinguishible from a realisation of size T of an *i.i.d.* multivariate normal distribution for $(\mathbf{r}_t, \mathbf{f}_t)$.

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

$$\overline{sr}^2 \cdot E^2(m)/V(m) = R^2,$$

where $\overline{sr}$ 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 $\overline{sr}$ 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 bounds should coincide because $R^2 = 1$.

(TABLE 1)

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 (MK), which we also identify with the CRSP value-weighted excess return. Table 2 contains the results of applying the different inference procedures previously discussed to this model. 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), except for the first-step estimators, for which we use the identity matrix as initial weighting matrix.[5]

The first thing to note is that the value of CU-GMM J statistic (=21.375, 0.6% p-value) is the same across the three methods, as stated in point 1 of Proposition 1. In contrast, there are marked numerical differences between the three two-step versions of the overidentification restrictions test. In particular, the centred SDF yields a much higher J statistic. These numerical differences are substantially reduced but not eliminated if we use those two-step GMM estimates to compute a new weighting matrix, which we then use to obtain three-step parameter estimates, and so on and so forth.[6] On the other hand, while the two-step, iterated and CU-GMM estimates of λ and μ are fairly similar, the CU-GMM estimate of τ is higher than its two-step counterpart, although the t-ratio is lower. In any case, all 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, yields a p-value of 0.3%.

(TABLE 2)

The J tests reported in Table 2 can also be interpreted as distance metric tests of the null hypothesis of zero pricing errors in the eight currency returns only. The rationale is as follows. A distance metric test, which is the GMM analogue to a likelihood ratio statistic, is given by the difference between the criterion function under the null and the alternative. But if we saturate (1) by adding n pricing errors, then the joint system of moment conditions becomes exactly

[5] We have also considered another two-step GMM procedure that as first-step estimator uses (9), whose numerical value is 0.0186, but the results are qualitatively similar.

[6] As Hansen, Heaton and Yaron (1996) show, though, such iterated GMM estimators do not generally coincide with the CU ones.

identified, which in turn implies that the optimal criterion function under the alternative will be zero.

Similarly, we can also consider the distance metric 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 2. Under the alternative, though, we need to conduct a new estimation. Specifically, if we saturate the moment condition (2) 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. Such a model delivers a borderline significant t ratio for λ, and a CU-GMM J test of 7.601, whose p-value is 0.42. As a result, the CAPM restrictions are not rejected when we do not force this model to price the market. In contrast, the distance metric test of zero pricing error for the traded factor, which is equal to the difference between this J statistic and the one reported in Table 2, is 14.314, 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 Farnsworth et al. (2002) and Lewellen, Nagel and Shanken (2009).

Importantly, these distance metric 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 (2010), and which would be particularly relevant in the elliptical case in view of Lemma 1.

Table E1 in the supplemental appendix also confirms the numerical equality of the CU-GMM estimators of prices of risk (λ, τ and μ) and pricing errors ($\boldsymbol{\alpha}$, $\boldsymbol{\pi}$ and $\boldsymbol{\psi}$) regardless of the approach used to estimate them, as expected from 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. Finally, Figure 1, which plots the CU-GMM criterion as a function of λ, confirms that we have obtained a global minimum.

3 Non-traded factors

Let us now consider a situation in which f is a scalar non-traded factor, such as the growth rate of per capita consumption. The main difference with the analysis in section 2 is that the factor may not satisfy the pricing equation (2), so that the SDF is simply defined by (1). As in the case of a traded factor, we can equivalently write this pricing condition as a restriction on the least squares projection of $\mathbf{r}$ onto a constant and f. Specifically, if we define

$$\boldsymbol{\phi} = E(\mathbf{r}) - \boldsymbol{\beta}\mu$$

and

$$\boldsymbol{\beta} = Cov\left(\mathbf{r}, f\right) / \sigma^2 \tag{10}$$

as the vectors of intercepts and slopes in that projection, respectively, then (1) is equivalent to

$$E\left[\left(a + bf\right)\left(\boldsymbol{\phi} + \boldsymbol{\beta} f\right)\right] = c\boldsymbol{\phi} + d\boldsymbol{\beta} = \mathbf{0}, \tag{11}$$

where c and d are two scalars not simultaneously equal to 0, so that the projection and the SDF should be orthogonal. Intuitively, (1) implies that we can find a non-trivial linear combination of $E(\mathbf{r})$ and $E(\mathbf{r}f)$ (or $Cov(\mathbf{r}, f)$) that is zero, which in turn implies that we can find a non-trivial linear combination of $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$ that is zero too.[7]

In practice, we can easily impose the implicit constraint (11) by writing the moment conditions that define the projection as

$$E\left[\begin{array}{c} \left(\mathbf{r} - \boldsymbol{\varphi}\left(d - cf\right)\right) \\ \left(\mathbf{r} - \boldsymbol{\varphi}\left(d - cf\right)\right) f \end{array}\right] = \left(\begin{array}{c} \mathbf{0} \\ \mathbf{0} \end{array}\right), \tag{12}$$

where $\boldsymbol{\varphi}$ is some $n \times 1$ vector such that $\boldsymbol{\phi} = d\boldsymbol{\varphi}$ and $\boldsymbol{\beta} = -c\boldsymbol{\varphi}$. This moment condition closely resembles (3), except for the fact that when f is traded, the additional condition (2) implies that $\boldsymbol{\phi} = \mathbf{0}$ and $d = 0$. More generally, we need to solve the scale indeterminacy of c and d by choosing either the popular asymmetric normalisation $c = -1$, so that $\boldsymbol{\beta} = \boldsymbol{\varphi}$, or the symmetric normalisation $c^2 + d^2 = 1$. As we shall see below, this seemingly innocuous issue may once again have important empirical consequences.

3.1 Existing approaches

As we mentioned in the case of traded factors, some normalisation is required to identify (a, b) from (1). Specifically, the uncentred SDF approach implicitly sets $a = 1$, and relies on the n overidentified, linear moment conditions

$$E\left[\begin{array}{c} \mathbf{r}\left(1 - f\lambda\right) \end{array}\right] = \mathbf{0}, \tag{13}$$

where the only parameter to estimate is λ. This parameter still has a "price of risk" interpretation, as in (5), but this time in terms of factor mimicking portfolios. In particular

$$\lambda = \frac{E(r^+)}{E(r^{+2})}, \tag{14}$$

where

$$r^+ = E(f\mathbf{r}')E^{-1}(\mathbf{r}\mathbf{r}')\mathbf{r} \tag{15}$$

[7] Note that $c = a + b\mu$ is the mean of the SDF, while $d = a\mu + b\gamma$ would be the price of f if it was traded.

is the uncentred least squares projection of f on $\mathbf{r}$. As expected, the asymptotic distribution of the J test will be χ^2_{n-1} under the null. For that reason, in what follows we assume that the number of assets exceeds the number of factors to ensure that the linear factor pricing model imposes testable restrictions on asset returns.

In contrast, the centred SDF test implicitly sets $a + b\mu = 1$, and relies on the overidentified, $n + 1$ non-linear in parameters moment conditions

$$E\left[\begin{array}{c} \mathbf{r}\left(1 - (f-\mu)\tau\right) \\ f - \mu \end{array}\right] = \mathbf{0}, \tag{16}$$

where the two parameters to estimate are (μ, τ). Once again, τ has a "price of risk" interpretation as in (7), but in terms of factor mimicking portfolios. Specifically

$$\tau = \frac{E(r^{++})}{V(r^{++})}, \tag{17}$$

where

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

is the centred least squares projection of f on $\mathbf{r}$. The link to the uncentred "price of risk" parameter is simply

$$\tau = \frac{\lambda}{1 - \lambda\mu},$$

which effectively divides λ by the SDF mean since the centred SDF approach normalises that mean to 1. As before, the J test will asymptotically converge to a χ^2_{n-1} when the asset pricing model is correct.

On the other hand, if we normalise $c = -1$ then we can write (12) as the following $2n$ overidentified, non-linear in parameters moment conditions:

$$E\left[\begin{array}{c} (\mathbf{r} - \boldsymbol{\beta}(\varkappa + f)) \\ (\mathbf{r} - \boldsymbol{\beta}(\varkappa + f))f \end{array}\right] = \mathbf{0}, \tag{18}$$

where the $n + 1$ parameters to estimate are $(\varkappa, \boldsymbol{\beta})$, with $\varkappa = \delta - \mu$, so that we can interpret δ as the expected excess return of a portfolio whose "beta" is equal to 1.[8] Given that the risk premium of r^{++} is $\tau V(r^{++})$ and its beta is $V(r^{++})/\sigma^2$, there is again a simple connection to the "price or risk" of the centred SDF. Specifically,

$$\delta = \sigma^2\tau, \tag{19}$$

[8] Jagannathan and Wang (2002) use $\delta - \mu$ 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. Consequently, we will ignore them for the time being to simplify the exposition, although we will use them in our proofs to link the regression and SDF approaches.

which simply re-scales τ by the factor variance. This is the usual regression (or beta) test of the pricing model, which implicitly exploits the restrictions on the regression intercepts (see Campbell, Lo and MacKinlay (1996, chap. 5)).

Finally, the centred and uncentred pricing errors $\boldsymbol{\psi}$ and $\boldsymbol{\pi}$ are defined as in section 2, while Jensen's alpha is now defined as

$$\boldsymbol{\alpha} = E(\mathbf{r}) - \boldsymbol{\beta}\delta. \tag{20}$$

Unfortunately, the existing approaches may run into difficulties in those cases in which their implicit normalisations are invalid. For instance, if $E(\mathbf{r}f) = \mathbf{0}$ so that the true SDF would be proportional to f and $r^{+} = 0$, then the normalisation of (1) with $a = 1$ in the uncentred SDF approach (13) is not well defined in population terms. Similarly, if $Cov(\mathbf{r},f) = \mathbf{0}$ so that the true SDF would be proportional to $(f - \mu)$ and $r^{++} = 0$, then neither the normalisation of (1) with $a + b\mu = 1$ in the centred SDF approach (16), nor the normalisation of (12) with $c = -1$ in the centred regression approach (18) are properly defined. In contrast, the symmetric normalisations $a^2 + b^2 = 1$ and $c^2 + d^2 = 1$ continue to be well defined in those circumstances. We shall return to this issue in the empirical application.

3.2 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 such as CU-GMM. More formally[9]

Proposition 2 *If we apply single-step procedures to the uncentred SDF method based on the moment conditions (13), the centred SDF method based on the moment conditions (16), and the regression method based on the moment conditions (18), then for a common specification of the characteristics of the HAC weighting matrix the following numerical equivalences hold:*
1) The three overidentification restrictions (J) tests.
2) The direct estimate of the "price of risk" λ from (13), the indirect estimate

$$\lambda = \frac{\tau}{1 + \tau\mu}$$

from (16), and the indirect estimate of

$$\lambda = \frac{\delta}{\sigma^2 + \delta\mu}$$

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

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

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

$$E\begin{bmatrix} (\mathbf{r} - \boldsymbol{\beta}_m(f^{\theta} - \gamma_m/\mu_m)) \\ (\mathbf{r} - \boldsymbol{\beta}_m(f^{\theta} - \gamma_m/\mu_m)))f^{\theta} \\ f^{\theta} - \mu_m \\ f^{2\theta} - \gamma_m \end{bmatrix} = \mathbf{0},$$

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

from (18) parametrised in terms of δ and extended to include γ, provided that $1+\tau\mu = \left(\sigma^2+\delta\mu\right)/\sigma^2 \neq 0$. Analogous results apply to τ and δ.
3) The estimates of Jensen's alphas in (20) obtained by replacing $E(\cdot)$ by an unrestricted sample average and $\boldsymbol{\beta}$ and δ by their direct estimates obtained from the regression method or the indirect estimates obtained from SDF methods extended to include $\boldsymbol{\beta}$, γ and μ. Analogous results apply to the pricing errors $\boldsymbol{\pi}$ and $\boldsymbol{\psi}$.

It is important to distinguish this proposition 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 one-step GMM procedures such as CU-GMM. Since our equivalence result is numerical, it holds regardless of the validity of the pricing model, and does not depend on n, T or the number of factors.[10]

The numerical equivalence of the three approaches gives once more a non-trivial computational advantage to the uncentred SDF method, which only contains the single unknown parameter λ.[11] At the same time, one can also exploit the fact that the three approaches coincide to check that a global minimum has been obtained.

Still, it is convenient to have consistent 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 joint Gaussian assumption in Jagannathan and Wang (2002). In particular, if we derive the optimal moment condition for the uncentred SDF model in this context, then we can immediately show that:

Lemma 2 *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 estimator of λ obtained from (13) will be given by*

$$\mathring{\lambda}_T = \frac{\sum_{t=1}^T r_t^+}{\sum_{t=1}^T r_t^{+2}} \tag{21}$$

where r_t^+ is the uncentred factor mimicking portfolio defined in (15), whose sample counterpart would be

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

[10] 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.

[11] 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, while the regression method the estimation of $n \times k$ factor loadings if we use the parametrisation in terms of $\varkappa$, and even more elements if we use the parametrisation in terms of $\boldsymbol{\delta}$.

Once again, it is important to stress that the feasible version of $\hat{\lambda}_T$ 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.

Similarly, (21) also coincides with the GMM estimator of λ that we would obtain from (13) if we used as weighting matrix the second moment of the excess returns in $\mathbf{r}$, as recommended by Hansen and Jagannathan (1997).

3.3 Empirical application

We look again at the eight currency portfolios in Lustig and Verdelhan (2007), but this time we focus on linear factor pricing models with non-traded factors. Given that for pedagogical reasons we have only considered single factor models in our theoretical analysis, we consider a linearised version of the CCAPM, which defines the US per capita consumption growth of nondurables as the only pricing factor.

Table 3 displays the results from the application of the different inference procedures previously discussed to this data set for the purposes of testing the CCAPM. The computation of the weighting matrix in two-step and CU-GMM is the same as we explained for the case of traded factors.[12] As stated in point 1 of Proposition 2, the value of CU-GMM J statistic (=5.663, 58% p-value) is the same across the three methods. Therefore, in this case this J test does not reject the null hypothesis implicit in (1) and (12), which is in agreement with the empirical results in Lustig and Verdelhan (2007). This result 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 again numerical differences between the standard two-step GMM implementation of the three existing approaches. Unlike what happened in the case of traded factors, though, in this case the numerical differences lead to different conclusions at conventional significance levels. Specifically, while the centred SDF approach rejects the null hypothesis, the uncentred SDF and the regression approach do not. These numerical differences are attenuated when we use iterated GMM procedures, but the conflicting conclusions remain. On the other hand, while the two-step, iterated and CU-GMM estimates of λ and δ are fairly close, the CU-GMM estimate of τ is higher than its two-step and iterated counterparts, although the t-ratio is much lower.

(TABLE 3)

[12] We have also considered another two-step GMM procedure that as first-step estimator uses (21), whose numerical value is 0.492, but the results are qualitatively similar.

Table E2 in the supplemental appendix also confirms the numerical equality of the CU-GMM estimators of prices of risk (λ, τ and μ) and pricing errors (α, π and ψ) 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, which explains the three different columns required for each of them. Finally, Figure 2, which plots the CU-GMM criterion as a function of λ, confirms that we have obtained a global minimum.

Importantly, Burnside (2007) argues that the usual two-step implementation of the uncentred SDF approach has no power against potentially misspecified SDFs when the population covariance of the pricing factors with the excess returns on the assets is 0. Similarly, one could easily modify his arguments to say that the usual two-step implementation of the centred SDF approach would have no power if the cross moment between pricing factors and excess returns were 0, and the same would apply to the centred regression approach.

Given the numerical equivalence of the single-step implementation of the three approaches in Proposition 2, in our view the focus should not be on the statistical properties of the different estimators and tests, but rather, on the interpretation of the restrictions that are effectively tested in those two special cases.

For the sake of clarity, let us study these issues with a single factor. Specifically, when $Cov(\mathbf{r}, f) = \mathbf{0}$ but $E(\mathbf{r}) \neq \mathbf{0}$, (12) will be satisfied with $c = 0$ and $\boldsymbol{\varphi} d = E(\mathbf{r})$, while the moment conditions (1) will be satisfied by any affine transformation of f such that $a + b\mu = 0$. Therefore, the value of λ given by (14) will trivially satisfy (13), as Burnside (2007) shows. In contrast, one cannot find any finite value of τ that will satisfy (16) because the centred mimicking representing portfolio r^{++} will be 0. Likewise, one cannot find any finite value of $\varkappa$ that will satisfy (18) because $\boldsymbol{\beta} = \mathbf{0}$. Therefore, the lack of correlation between excess returns and consumption growth could explain the striking numerical differences between the empirical results obtained with the centred and uncentred SDF moment conditions using two-step and iterated GMM.

Similarly, when $E(\mathbf{r}f) = \mathbf{0}$ but $E(\mathbf{r}) \neq \mathbf{0}$, (12) will be satisfied with $d\mu - c\gamma = 0$ and $\boldsymbol{\varphi} d\left(\sigma^2/\gamma\right) = E(\mathbf{r})$, while the moment conditions (1) will be satisfied by any SDF which is exactly proportional to f (so that $a = 0$). Therefore, the value of τ given by (17) will trivially satisfy (16), and the same applies to (18) with (10) and $\varkappa = -\gamma/\mu$. In contrast, one cannot find any finite value of λ that will satisfy (13) because the uncentred mimicking portfolio r^{+} will be 0.

From an economic point of view those solutions are clearly unsatisfactory, but strictly speaking the corresponding SDF's correctly price the vector of excess returns at hand. In our view, the best way to solve these problems would be to add assets whose cost is not 0, which would

implicitly fix the scale of the SDF by fixing its mean; see e.g. Hodrick and Zhang (2001) and Farnsworth et al. (2002). For that reason, we devote the section 5 to this case. Unfortunately, most empirical studies, including Lustig and Verdelhan (2008) only include zero cost assets.

From an econometric point of view, though, the truly problematic case arises when (1) and (12) hold but $E(\mathbf{r}) = \mathbf{0}$, in which case both $Cov(\mathbf{r}, f)$ and $E(\mathbf{r}f)$ must be $\mathbf{0}$ too. In this situation, the SDF parameters a and b and the projection parameters c and d are underidentified even after normalisation, which renders standard GMM inferences invalid.[13]

To investigate whether these theoretical situations are empirically relevant, we perform single-step GMM overidentification tests of the following null hypotheses:
(a) the mean excess return is 0, i.e.

$$E(\mathbf{r}) = \mathbf{0},$$

(b) the cross moment between excess returns and factor is 0, i.e.

$$E(\mathbf{r}f) = \mathbf{0},$$

and (c) the covariance between excess returns and factor is 0, i.e.

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

where $\boldsymbol{\varphi}$ is a vector of parameters to estimate.

Hypothesis (a) is clearly rejected with a J statistic of 39.97, whose p-value is essentially 0. Therefore, there are statistically significant risk premia in search of pricing factors to explain them. The next step is to investigate if US consumption growth can play such a role.

Hypothesis (b) is also unambiguously rejected with a statistic of 34.85 and a p-value close to 0, but there is not much evidence against hypothesis (c) for non-durable consumption, with a test statistic of 8.39 and p-value of almost 40%. Thus, we cannot reject that this factor is jointly uncorrelated with the currency portfolios, which indicates that the seemingly positive evaluation of the consumption based asset pricing model in Table 3 must be interpreted with some care. In this sense, note that the joint test of (c) is effectively testing that any portfolio formed from the original eight currency portfolios in Lustig and Verdelhan (2008) is uncorrelated to US consumption growth. In this sense, it is worth noting that once again CU-GMM proves useful to unify the empirical results because the zero covariance test above is numerically equivalent to a test that all the betas are 0. 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

[13] See Kan and Zhang (1999) or Burnside (2007) for the implications that identification failures have for two-step GMM procedures.

on asymptotic distributions, because it is corroborated by a p-value of 81.7% for the F test of the latter regression, which like the corresponding LR test, is also invariant to exchanging regressand and regressors. 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 (2008) and Burnside (2009) 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.

Finally, the joint hypothesis (a)+(b), or equivalently (a)+(c), is also rejected with a statistic of 53.039 and a negligible p-value, which confirms that the parameters appearing in (1) and (12) are point identified in a single factor model. Interestingly, this test coincides with a simple version of the underidentification test of Arellano, Hansen and Sentana (2009) adapted to linear factor pricing models by Manresa (2008).

In summary, our empirical results with a non-traded factor indicate that although we cannot reject the overidentification restrictions implicit in (1) and (12), this may be due to the fact that US consumption growth is uncorrelated to the eight portfolios of currencies in Lustig and Verdelhan (2007). In this sense, the CCAPM results are very similar to the ones obtained with the CAPM when we treat the market portfolio as non-traded, as we mentioned at the end of section 2, which 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.

4 Mixed factors

Let us now consider a model with two pricing factors in which f_1 is a scalar traded factor, such as the market portfolio, and f_2 is another scalar non-traded factor, such as growth rate of per capita consumption. An important example of such a mixed linear factor pricing model will be the linearised CCAPM with Epstein and Zin (1991) preferences. Apart from the multifactor nature of this model, the main difference with the analysis in the previous sections is that while f_1 must satisfy the pricing equation (2), f_2 does not. As a result, if we define $\mathbf{f}' = (f_1, f_2)$, then the SDF will be implicitly defined by

$$E\left[\begin{array}{c}(a+\mathbf{b}'\mathbf{f})\,\mathbf{r}\\(a+\mathbf{b}'\mathbf{f})\,f_1\end{array}\right]=\mathbf{0}, \tag{22}$$

subject to some scale normalisation of the vector $(a, \mathbf{b})$.

In this context, the uncentred test relies on the overidentified, $n+1$ linear moment conditions

$$E\left[\begin{array}{c}\mathbf{r}\left(1-\mathbf{f}'\boldsymbol{\lambda}\right)\\ f_1\left(1-\mathbf{f}'\boldsymbol{\lambda}\right)\end{array}\right]=\mathbf{0},$$

where the two unknown parameters are contained in $\boldsymbol{\lambda}$. In contrast, the centred SDF test relies on the $n+3$ overidentified, nonlinear in parameters moment conditions

$$E\left[\begin{array}{c}\mathbf{r}\left(1-\left(\mathbf{f}-\boldsymbol{\mu}\right)'\boldsymbol{\tau}\right)\\ f_1\left(1-\left(\mathbf{f}-\boldsymbol{\mu}\right)'\boldsymbol{\tau}\right)\\ \mathbf{f}-\boldsymbol{\mu}\end{array}\right]=\mathbf{0},$$

whose four unknown parameters are in $\boldsymbol{\mu}$ and $\boldsymbol{\tau}$. Finally, the regression version can be written in terms of the following $3n$ overidentified, nonlinear in parameters moment conditions:

$$E\left[\begin{array}{c}\mathbf{r}-\boldsymbol{\beta}_1 f_1-\boldsymbol{\beta}_2\left(\varkappa_2+f_2\right)\\ \left(\mathbf{r}-\boldsymbol{\beta}_1 f_1-\boldsymbol{\beta}_2\left(\varkappa_2+f_2\right)\right)f_1\\ \left(\mathbf{r}-\boldsymbol{\beta}_1 f_1-\boldsymbol{\beta}_2\left(\varkappa_2+f_2\right)\right)f_2\end{array}\right]=\mathbf{0},$$

where the $2n+1$ parameters to estimate are $(\varkappa_2,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)$, with $(\boldsymbol{\beta}_1|\boldsymbol{\beta}_2)=cov(\mathbf{r},\mathbf{f})V^{-1}(\mathbf{f})$ and $\varkappa_2=\delta_2-\mu_2$, so that we can interpret δ_2 as the expected excess return of a portfolio whose "betas" on f_1 and f_2 are equal to 0 and 1, respectively.

Not surprisingly, it is straightforward to combine the proofs of Propositions 1 and 2 to show that all these three approaches numerically coincide if one uses single-step methods such as CU-GMM.

Table 4 contains the results of estimating a linearised version of the CCAPM with Epstein and Zin (1991) preferences with the dataset that we have been considering. 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 previous sections.

(TABLE 4)

As can be seen from Table 4, the common CU-GMM J statistic is 4.785, with a p-value of 68.6%, so like the CCAPM considered in the previous section, the combined model is not rejected even though we attempt to price the market portfolio as in the CAPM. In addition, the estimates of $\boldsymbol{\lambda}$ show that consumption growth and not the market portfolio is the driving force behind this model. In addition, Figure 3, which plots the CU-GMM criterion as a function of the vector $\boldsymbol{\lambda}$, confirms that we have obtained a global minimum.

Once again, the two-step, iterated and CU GMM implementations of the uncentred SDF approach provide similar results, with slightly higher differences for the regression method. But the results of the two-step implementation of the centred SDF approach clearly diverge from the CU-GMM results, with negative estimates of the prices of risk and a very high J statistic. Moreover, in this case the iterated GMM fails to converge, cycling over three different solutions, and for that reason we do not report them.

Once again, the cause of the wedge between the results obtained with the centred and uncentred SDF moment conditions using two-step GMM seems to be the lack of correlation between consumption growth and asset returns. Specifically, if $Cov(\mathbf{r}, f_2) = \mathbf{0}$ and $Cov(f_1, f_2) = 0$ then (22) will be satisfied with $b_1 = 0$ and $a + b_2\mu_2 = 0$ even though $E(\mathbf{r})$ is not proportional to $Cov(r, f_1)$. This is another example of an economically unattractive SDF with $E(m) = 0$, which nevertheless satisfies the moment restrictions, as the CU GMM estimates confirm. In this context, the uncentred SDF equations will trivially hold, while the centred SDF equations cannot hold because they implicitly normalise with $E(m) = 1$. Not surprisingly, exactly the opposite situation would arise if $E(\mathbf{r}f_2) = \mathbf{0}$ and $E(f_1 f_2) = 0$.

As we mentioned in section 3, the most satisfactory solution to these normalisation-induced problems would be to add assets with non-zero cost, which are unfortunately not readily available in the Lustig and Verdelhan (2008) dataset. Still, it is worth extending our theoretical results to this case.

5 Extension to gross returns

Let us define $\mathbf{R}$ as an $n \times 1$ vector of gross returns, which are such that their cost is given by the $n \times 1$ vector of ones ℓ_n. We focus again on the case of a single factor f to simplify the exposition, and relegate the general multifactor case to the supplemental appendix.

In this context, the analogue to the SDF pricing equation (1) is

$$E\left((a + bf)\,\mathbf{R}\right) = \ell_n \tag{23}$$

for some real numbers (a, b). As we mentioned in the previous section, there is no longer any need to normalise (a, b), unlike in the case of excess returns.

Not surprisingly, we can equivalently write this pricing condition as a restriction on the least squares projection of $\mathbf{R}$ onto a constant and f. Specifically, if we define $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$ as the vectors of intercepts and slopes in that projection, respectively, then (23) is equivalent to

$$c\boldsymbol{\phi} + d\boldsymbol{\beta} = \ell_n, \tag{24}$$

where c is the SDF mean and d is the shadow price of the factor, or its actual price when is traded.

When the factor is a gross return, it will also satisfy

$$E\left(\left(a+bf\right)f\right)=1. \tag{25}$$

Therefore, the moments of the uncentred SDF method will be (23) and (25). Since the only parameters are (a,b), the J test will have $n-1$ degrees of freedom under correct specification.

We can also reparametrise the SDF in terms of c instead of a and rely on the moment conditions

$$E\left[\begin{array}{c}\mathbf{R}\left(c-\left(f-\mu\right)\tau\right)-\ell_n\\ f\left(c-\left(f-\mu\right)\tau\right)-1\\ f-\mu\end{array}\right]=\mathbf{0},$$

whose parameters are (c,τ,μ), which are analogous to (6).

Regarding the regression approach, the fact that the factor is traded implies that $d=1$, which simplifies the least squares constraint to

$$\boldsymbol{\phi}=\kappa\left(\ell_n-\boldsymbol{\beta}\right),\qquad \kappa=1/c,$$

when $c\neq 0$.[14] The parameter κ is usually referred to as the zero-beta return since it corresponds to the expectation of returns uncorrelated with f.

Therefore, the moments associated to the regression method will be

$$E\left[\begin{array}{c}\left(\left(\mathbf{R}-\kappa\ell_n\right)-\boldsymbol{\beta}\left(f-\kappa\right)\right)\\ \left(\left(\mathbf{R}-\kappa\ell_n\right)-\boldsymbol{\beta}\left(f-\kappa\right)\right)f\end{array}\right]=\mathbf{0},$$

with parameters $(\boldsymbol{\beta},\kappa)$.

We show in the supplemental appendix that there is a direct counterpart to Proposition 1 for gross returns too, so that all the approaches are numerically equivalent when implemented by single-step GMM methods.

Let us now consider the case of a non-traded factor, in which case the moments of the SDF method are simply (23) with parameters (a,b). Consequently, the J test has $n-2$ degrees of freedom under correct specification.

[14] The case $c=0$ is such that f is equal to the minimum variance portfolio, whose return R^{**} satisfies

$$Cov\left(\mathbf{R},R^{**}\right)=Var\left(R^{**}\right)\ell_n.$$

It is well known that this asset is the only element of the mean-variance frontier that cannot be used as a benchmark in beta pricing.

Once again, we can also reparametrise the SDF in terms of c instead of a and rely on the moment conditions

$$E\left[\begin{array}{c}\mathbf{R}\left(c-\left(f-\mu\right)\tau\right)-\ell_n\\ f-\mu\end{array}\right]=\mathbf{0}$$

with parameters (c,τ,μ), which are analogous to (16).

On the other hand, when $c\neq 0$ we can re-express the least squares constraint as

$$\boldsymbol{\phi}=\kappa\ell_n+\boldsymbol{\beta}\varkappa,\qquad \kappa=1/c,\qquad \varkappa=-d/c.$$

Therefore, the moments of the regression method are

$$E\left[\begin{array}{c}\left(\left(\mathbf{R}-\kappa\ell_n\right)-\boldsymbol{\beta}\left(f+\varkappa\right)\right)\\ \left(\left(\mathbf{R}-\kappa\ell_n\right)-\boldsymbol{\beta}\left(f+\varkappa\right)\right)f\end{array}\right]=\mathbf{0},$$

with parameters $(\boldsymbol{\beta},\kappa,\varkappa)$.

We show in the supplemental appendix that there is a direct counterpart to Proposition 2 for gross returns too, so that all the approaches are numerically equivalent when implemented by single-step GMM methods. More generally, it is straightforward to show that analogous results hold for the mixed case discussed in section 4.

Finally, it is worth explaining the advantages of working with gross returns in the situation discussed in section 3.3 in which the non-traded factor is uncorrelated with excess returns. For simplicity, let us assume that the reference asset used in computing $\mathbf{r}$ is a portfolio whose gross return has zero variance, so that $Cov(\mathbf{R},f)=\mathbf{0}$ if and only if $Cov(\mathbf{r},f)=\mathbf{0}$. As we mentioned in that section, the moment condition (1) will hold with $a=-b\mu$ regardless of $E\left(\mathbf{r}\right)$, which corresponds to a valid but economically unsatisfactory SDF. In contrast, (23) cannot be satisfied unless $E\left(\mathbf{R}\right)$ is proportional to ℓ_n. If this "risk neutral" condition does not hold, then the J test based on $\mathbf{R}$ will reject with probability 1 in large samples, while the rejection probability of the J test based on $\mathbf{r}$ will coincide with its size.

On the other hand, an unsatisfactory SDF that will satisfy (23) would be one in which the factor is correlated with returns but cannot discriminate across assets, so that $Cov(\mathbf{R},f)=k\ell_n$ for some scalar constant $k\neq 0$. Similar arguments apply to $E(\mathbf{R}f)$.

6 Conclusions

There are two main approaches in empirical finance to evaluate linear factor pricing models. The oldest method relies on regressions of excess returns on factors, while the other more recent methods rely instead on the SDF representation of the model. There are two variants of the SDF approach, one that subtracts the mean of the factors and another one which does not.

Given that these 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, pricing errors, and Jensen's alphas irrespective of the validity of the model, and regardless of n and T. In this way, we eliminate the possibility that different researches report potentially contradictory results with the same data set.

We prove our numerical equivalence results hold for any combination of traded and non-traded factors, and also for excess returns and gross returns. Thus, we would argue that in effect the regression and SDF approaches are different representations of a single method to empirically evaluate asset-pricing models. Nevertheless, the uncentred SDF method has a non-trivial computational advantage because it contains fewer unknown parameters. At the same time, one can also exploit the numerical equivalence of the three approaches to check that a global minimum has been reached.

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

We illustrate our results with the currency portfolios constructed by Lustig and Verdelhan (2007). We consider three popular linear factor pricing models: the CAPM, as well as linearised versions of the Consumption CAPM and the Epstein and Zin (1989) model. 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 (2007) findings that the discrepancies between traditional estimators are due to the fact that US consumption growth seems to be uncorrelated to currency returns. We also find that if we force the CAPM to price the market portfolio, then we reject the asset pricing restrictions.

Given that our results are not specific to CU-GMM, it might be worth comparing the properties in finite samples of different members of the generalised empirical likelihood family of estimators. In addition, it would be interesting to compare them in our empirical application.

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, these topics constitute interesting avenues for further research.

References

Almeida, C. and R. Garcia (2009): "Assessing misspecified asset pricing models with empirical likelihood estimators", mimeo, Fundaçäo Getulio Vargas.

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

Arellano, M., L.P. Hansen and E. Sentana (2009): "Underidentification?", CEMFI Working Paper 0905.

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

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

Burnside, C., M. Eichenbaum, I. Kleshchelski and S. Rebelo (2006): "The returns to currency speculation," NBER Working Paper 12,489.

Burnside, C. (2007): "Empirical asset pricing and statistical power in the presence of weak risk factors", NBER Working Paper 13,357.

Burnside, C. (2009): "The cross-section of foreign currency risk premia and consumption growth risk: A comment", mimeo.

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

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

Campbell, J.Y., S. Gilgio, and C. Polk (2010): "Hard times", mimeo, Harvard University.

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

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

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

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

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

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

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

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

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

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

Gospodinov, N., R. Kan and C. Robotti (2010): "Further results on the limiting distribution of GMM sample moment conditions", Atlanta Fed Working Paper 2010-11.

Hansen, L.P. (1982): "Large sample properties of generalized method of moments estimators", *Econometrica* 50, 1029-1054.

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

Hansen, L.P. and R. Jagannathan (1991): "Implications of security market data for models of dynamic economies", *Journal of Political Economy* 99, 225-262.

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

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

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

Julliard, C. and A. Ghosh (2008): "Can rare events explain the equity premium puzzle?", LSE FMG Discussion Paper 610.

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

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

Kan, R. and G. Zhou (2000): "A critique of the stochastic discount factor methodology", *Journal of Finance* 54, 1221-1248.

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

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

Lewellen J., S. Nagel and J. Shanken (2009): "A skeptical appraisal of asset-pricing tests, *Journal of Financial Economics*, forthcoming

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

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

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

Lustig, H. and A. Verdelhan (2008): "Note on the cross-section of foreign currency risk premia and consumption growth risk", NBER WP 13812.

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

Magnus, J.R. and H. Neudecker (1990): *Matrix Differential Calculus with Applications to Econometrics*, Wiley.

Manresa, E. (2008): "Application of rank tests to linear factor asset pricing models", unpublished Master's thesis, CEMFI.

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

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

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

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

Peñaranda, F. and E. Sentana (2010): "Spanning tests in portfolio and stochastic discount factor mean-variance frontiers: a unifying approach", mimeo, UPF.

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

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

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

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

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

Appendices

A Single-step methods: Continuously Updated GMM

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

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

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

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

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

where

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

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

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

defined as

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

where

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

Peñaranda and Sentana (2010) discuss how to express the CU-GMM criterion in terms of OLS output, which facilitates its optimisation. Although this estimator is often more difficult to compute than two-step and iterated estimators, particularly in linear models, an important advantage is that it is numerically invariant to normalisation, bijective reparametrisations and parameter-dependent linear transformations of the moment conditions, which will again prove useful in our context. In contrast, these properties do not necessarily hold for two-step or iterated GMM.

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

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

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

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

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

where

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

Finally, $T\cdot J_T(\hat{\boldsymbol{\theta}}_T)$ and $T\cdot\tilde{J}_T(\tilde{\boldsymbol{\theta}}_T)$ will be asymptotically distributed as the same chi-square with $\dim(\mathbf{h}) - \dim(\boldsymbol{\theta})$ degrees of freedom if $E[\mathbf{h}(\mathbf{x};\boldsymbol{\theta})] = \mathbf{0}$ holds, so that we can use those statistics to compute overidentifying restrictions (J) tests.

B Multifactor models

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

$$\begin{aligned} E(\mathbf{r})a + E(\mathbf{r}\mathbf{f}')\mathbf{b} &= E(\mathbf{r})(a + \mathbf{b}'\boldsymbol{\mu}) + Cov(\mathbf{r}, \mathbf{f})\mathbf{b} = \boldsymbol{\phi}(a + \mathbf{b}'\boldsymbol{\mu}) + \mathbf{B}\boldsymbol{\mu}(a + \mathbf{b}'\boldsymbol{\mu}) + \mathbf{B}\boldsymbol{\Sigma}\mathbf{b} \\ &= \boldsymbol{\phi}(a + \mathbf{b}'\boldsymbol{\mu}) + \mathbf{B}\left(a\boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{b}\right) = \boldsymbol{\phi}c + \mathbf{B}\mathbf{d} = \mathbf{0}, \end{aligned} \tag{B1}$$

where c is the mean of the SDF, $\mathbf{d}$ is the shadow price of $\mathbf{f}$ (or its actual price if it is a vector of traded payoffs) and

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

is the $n \times k$ matrix of regression slopes. We can interpret condition (B1) as reflecting the orthogonality between the SDF and the projection of $\mathbf{r}$ onto a constant and $\mathbf{f}$ because $\boldsymbol{\phi}c + \mathbf{B}\mathbf{d} = E\left[\left(a + \mathbf{b}'\mathbf{f}\right)\left(\boldsymbol{\phi} + \mathbf{B}\mathbf{f}\right)\right] = \mathbf{0}$.

Traded factors

If the factors are excess returns themselves then condition (B1) is equivalent to $\boldsymbol{\phi} = \mathbf{0}$ because $c = a + \mathbf{b}'\boldsymbol{\mu} \neq 0$ and the price of $\mathbf{f}$ must satisfy $\mathbf{d} = a\boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{b} = \mathbf{0}$. Therefore, we can evaluate the corresponding asset pricing model by means of the SDF influence functions

$$\mathbf{h}_S\left(\mathbf{r}, \mathbf{f}; a, \mathbf{b}\right) = \begin{bmatrix} \mathbf{r}\left(a + \mathbf{f}'\mathbf{b}\right) \\ \mathbf{f}\left(a + \mathbf{f}'\mathbf{b}\right) \end{bmatrix}, \tag{B2}$$

or the regression influence functions

$$\mathbf{h}_R\left(\mathbf{r}, \mathbf{f}; \mathbf{B}\right) = \begin{bmatrix} \mathbf{r} - \mathbf{B}\mathbf{f} \\ vec\left(\left(\mathbf{r} - \mathbf{B}\mathbf{f}\right)\mathbf{f}'\right) \end{bmatrix}. \tag{B3}$$

The SDF functions require some normalisation in their implementation. A symmetrically normalised version of the SDF approach would use the normalisation $a^2 + \mathbf{b}'\mathbf{b} = 1$ but asymmetric normalisations are more common in empirical work. The uncentred SDF method imposes $a = 1$ and relies on the influence functions

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

with parameters $\boldsymbol{\lambda}$, while the influence functions of the centred SDF method impose $a + \boldsymbol{\mu}'\mathbf{b} = 1$, and become

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

with parameters $(\boldsymbol{\mu}, \boldsymbol{\tau})$. The link between both sets of parameters is

$$\boldsymbol{\mu} = \boldsymbol{\Gamma}\boldsymbol{\lambda} = \boldsymbol{\Sigma}\boldsymbol{\tau}.$$

In all these methods the degrees of freedom of the corresponding J tests are n regardless of the number of factors k. The Jensen's alphas and pricing errors of excess returns $\mathbf{r}$ are defined by

$$\begin{aligned}
\boldsymbol{\alpha} &= E(\mathbf{r}) - \mathbf{B}E(\mathbf{f}), \\
\boldsymbol{\pi} &= E(\mathbf{r}) - E\left(\mathbf{r}\mathbf{f}'\right)\boldsymbol{\lambda}, \\
\boldsymbol{\psi} &= E(\mathbf{r}) - E\left(\mathbf{r}\left(\mathbf{f}-\boldsymbol{\mu}\right)'\right)\boldsymbol{\tau}.
\end{aligned}$$

Non-traded factors

If the factors are not traded payoffs then we need $n > k$ so that there are some overidentifying restrictions to test. As in the case of traded factors, the condition (B1) holds for a valid asset pricing model. Such a constraint is equivalent to both $\boldsymbol{\phi}$ and $\mathbf{B}$ belonging to the span of some $n \times k$ matrix that we can denote as

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

Assuming that $\mathbf{B}$ has full column rank, we can impose this implicit constraint on the intercepts and slopes of the regression of $\mathbf{r}$ on a constant and $\mathbf{f}$ as follows:

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

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

$$\mathbf{g}_S(\mathbf{r}, \mathbf{f}; a, \mathbf{b}) = \left[\mathbf{r}\left(a + \mathbf{f}'\mathbf{b}\right)\right], \tag{B4}$$

or the regression influence functions

$$\mathbf{g}_R(\mathbf{r}, \mathbf{f}; \mathbf{P}, c, \mathbf{d}) = \begin{bmatrix} \mathbf{r} - \mathbf{P}(\mathbf{d} - c\mathbf{f}) \\ vec\left(\left(\mathbf{r} - \mathbf{P}(\mathbf{d} - c\mathbf{f})\right)\mathbf{f}'\right) \end{bmatrix}. \tag{B5}$$

The SDF functions require some normalisation in their implementation. A symmetrically normalised version of the SDF approach would use the normalisation $a^2 + \mathbf{b}'\mathbf{b} = 1$, while the regression would rely on the normalisation $c^2 + \mathbf{d}'\mathbf{d} = 1$, but asymmetric normalisations are more common in empirical work. The uncentred SDF method imposes $a = 1$ and relies on the influence functions

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

with parameters $\boldsymbol{\lambda}$. The influence functions of the centred SDF method impose $a+\boldsymbol{\mu}'\mathbf{b}=1$ and they are

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

with parameters $(\boldsymbol{\mu},\boldsymbol{\tau})$. The usual regression imposes $c=-1$ and relies on the influence functions

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

with parameters $(\varkappa,\mathbf{B})$. Alternatively we can define the vector $\boldsymbol{\delta}$ of factor risk premia as $E\left(\mathbf{r}\right)=\mathbf{B}\boldsymbol{\delta}$, so that $\boldsymbol{\delta}=\varkappa+\boldsymbol{\mu}$, and add the estimation of $\boldsymbol{\mu}$. This yields

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

Assuming that $\mathbf{B}$ has full column rank, the link between these parameters is

$$\boldsymbol{\delta}=\left[\boldsymbol{\Sigma}+\boldsymbol{\delta}\boldsymbol{\mu}'\right]\boldsymbol{\lambda}=\boldsymbol{\Sigma}\boldsymbol{\tau}.$$

In any case, the degrees of freedom of the J test will be $n-k$. Regarding Jensen's alphas and pricing errors, they are defined by

$$\begin{aligned} \boldsymbol{\alpha} &= E\left(\mathbf{r}\right)-\mathbf{B}\boldsymbol{\delta}, \\ \boldsymbol{\pi} &= E\left(\mathbf{r}\right)-E\left(\mathbf{r}\mathbf{f}'\right)\boldsymbol{\lambda}, \\ \boldsymbol{\psi} &= E(\mathbf{r})-E\big(\mathbf{r}\left(\mathbf{f}-\boldsymbol{\mu}\right)'\big)\boldsymbol{\tau}. \end{aligned}$$

C Proofs

All proofs consider the multifactor context of appendix B instead of the simplifying single factor set up in the main text. In addition, the proofs of Proposition 1 and 2 do not rely on any particular normalisation since they are irrelevant for single-step methods, even though the proposition statements in the main text refer to the usual normalisations in empirical work.

Proposition 1:

Let us define an extended regression system that adds the estimation of $(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ to $\mathbf{h}_R(\mathbf{r}, \mathbf{f}; \mathbf{B})$ defined in (B3). Specifically

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

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.

If we choose $(a, \mathbf{b})$ such that

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

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

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

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

where the matrix $\mathcal{D}$ denotes the appropriate duplication matrix, that is, the matrix such that $vec(\cdot) = \mathcal{D}vech(\cdot)$ (see Magnus and Neudecker (1990)). Similarly

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

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

As we mentioned before, single-step methods are numerically invariant to normalisation, bijective reparametrisations and parameter-dependent linear transformations of the moment conditions. Therefore, for a given choice of HAC weighting matrix CU-GMM renders the extended

regression system $\mathfrak{h}_R(\mathbf{r},\mathbf{f};\mathbf{B},\boldsymbol{\mu},vech(\boldsymbol{\Gamma}))$ and the system

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

numerically equivalent. In particular, the estimates of $\mathbf{B}$ and $vech(\boldsymbol{\Gamma})$ are the same, the implied $\boldsymbol{\mu} = -\boldsymbol{\Gamma}\mathbf{b}/a$ is the same, and the J test is the same.

Given the definition of $\mathbf{h}_S(\mathbf{r},\mathbf{f};a,\mathbf{b})$ in (B2), the last system can also be expressed as

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

where the influence functions added to $\mathbf{h}_S(\mathbf{r},\mathbf{f};a,\mathbf{b})$ are exactly identified for $(\mathbf{B},vech(\boldsymbol{\Gamma}))$. Thus $\mathfrak{h}_S(\mathbf{r},\mathbf{f};a,\mathbf{b},\mathbf{B},vech(\boldsymbol{\Gamma}))$ is numerically equivalent to relying on the first block $\mathbf{h}_S(\mathbf{r},\mathbf{f};a,\mathbf{b})$ 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 $\mathbf{h}_S(\mathbf{r},\mathbf{f};a,\mathbf{b})$ numerically equivalent.

Finally, it is worth noting that in the case of the extended systems the numerical equivalence of the CU-GMM criterion function will hold not only at the optimum but also for any compatible set of parameter values. □

Lemma 1:

We assume that the vector $\mathbf{y} = (\mathbf{f}',\mathbf{r}')'$ follows an elliptical distribution, and denote the corresponding coefficient of multivariate excess kurtosis as κ, which is equal to $\kappa = 2/(\upsilon-4)$ in the case of Student t with υ 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 (4) for a multifactor model as

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

Thus, we can define the relevant Jacobian as

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

and similarly we can decompose the relevant asymptotic covariance matrix as

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

If we apply Lemma D1 in Peñaranda and Sentana (2010), 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 similarly

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

Thus, we only need to check that condition (C1) in Lemma C1 in Peñaranda and Sentana (2010) holds, which in our context becomes

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

This restriction will be satisfied because

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

and, since $E(\mathbf{r}) = E\left(\mathbf{r}\mathbf{f}'\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{\mu}$ under the null of tangency,

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

Therefore, the linear combinations of the moment conditions in $E(\mathbf{h}(\mathbf{y};\boldsymbol{\lambda})) = \mathbf{0}$ that provide the most efficient estimators of $\boldsymbol{\lambda}$ will be given by

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

□

Proposition 2:

Let us define a 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 (B5),

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

We are adding exactly identified parameters $(\boldsymbol{\mu}, vech(\boldsymbol{\Gamma}))$, and hence $\mathfrak{g}_R(\mathbf{r}, \mathbf{f}; \mathbf{P}, c, \mathbf{d}, \boldsymbol{\mu}, vech(\boldsymbol{\Gamma}))$ 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.

If we 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 $\mathfrak{g}_R\left(\mathbf{r}, \mathbf{f}; \mathbf{P}, c, \mathbf{d}, \boldsymbol{\mu}, vech\left(\boldsymbol{\Gamma}\right)\right)$:

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

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

where the matrix $\mathcal{D}$ denotes the corresponding duplication matrix.

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

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

where

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

Under CU-GMM, and a specific choice of HAC estimator, $\mathfrak{g}_S\left(\mathbf{r}, \mathbf{f}; a, \mathbf{b}, \mathbf{P}, \boldsymbol{\mu}, vech\left(\boldsymbol{\Gamma}\right)\right)$ provides the same estimates and J test as $\mathfrak{g}_R\left(\mathbf{r}, \mathbf{f}; \mathbf{P}, c, \mathbf{d}, \boldsymbol{\mu}, vech\left(\boldsymbol{\Gamma}\right)\right)$ because single-step procedures are numerically invariant to normalisation, bijective reparametrisations and parameter-dependent linear transformations of the moment conditions. 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 $\mathfrak{g}_S\left(\mathbf{r}, \mathbf{f}; a, \mathbf{b}, \mathbf{P}, \boldsymbol{\mu}, vech\left(\boldsymbol{\Gamma}\right)\right)$ coincide with their implied counterparts in the other system.

This last system can be related to the influence function $\mathbf{g}_S\left(\mathbf{r}, \mathbf{f}; a, \mathbf{b}\right)$ defined in (B4), where the influence functions that are added are exactly identified for $(\mathbf{P}, \boldsymbol{\mu}, vech\left(\boldsymbol{\Gamma}\right))$ given $(a, \mathbf{b})$. Thus $\mathfrak{g}_S\left(\mathbf{r}, \mathbf{f}; a, \mathbf{b}, \mathbf{P}, \boldsymbol{\mu}, vech\left(\boldsymbol{\Gamma}\right)\right)$ 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 $\mathbf{g}_R\left(\mathbf{r}, \mathbf{f}; \mathbf{P}, c, \mathbf{d}\right)$ and $\mathbf{g}_S\left(\mathbf{r}, \mathbf{f}; a, \mathbf{b}\right)$ numerically equivalent.

Once again, note that in the case of the extended systems the numerical equivalence of the CU-GMM criterion function will hold not only at the optimum but also for any compatible set of parameter values. □

Lemma 2:

We extend the results in appendix D in Peñaranda and Sentana (2010) 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 has the following long-run variance under the null

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

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

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

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

Computing the inverse, we obtain

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

and imposing the null hypothesis $E(\mathbf{r})=E\left(\mathbf{rf}'\right)\boldsymbol{\lambda}$, we get

$$E\left(\mathbf{fr}'\right)\left[E^{-1}\left(\mathbf{rr}'\right)+\frac{\omega}{1-\omega\boldsymbol{\lambda}'E\left(\mathbf{fr}'\right)E^{-1}\left(\mathbf{rr}'\right)E\left(\mathbf{rf}'\right)\boldsymbol{\lambda}}E^{-1}\left(\mathbf{rr}'\right)E\left(\mathbf{rf}'\right)\boldsymbol{\lambda}\boldsymbol{\lambda}'E\left(\mathbf{fr}'\right)E^{-1}\left(\mathbf{rr}'\right)\right]$$
$$=\left[\mathbf{I}_k+\frac{\omega}{1-\omega\boldsymbol{\lambda}'E\left(\mathbf{fr}'\right)E^{-1}\left(\mathbf{rr}'\right)E\left(\mathbf{rf}'\right)\boldsymbol{\lambda}}E\left(\mathbf{fr}'\right)E^{-1}\left(\mathbf{rr}'\right)E\left(\mathbf{rf}'\right)\boldsymbol{\lambda}\boldsymbol{\lambda}'\right]E\left(\mathbf{fr}'\right)E^{-1}\left(\mathbf{rr}'\right).$$

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

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

with

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

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

Table 1: Descriptive statistics in annual % terms, 1953-2002

	1	2	3	4	5	6	7	8
Mean	-2.336	-0.873	-0.747	0.329	-0.151	-0.213	2.988	2.031
SD	6.346	6.628	6.614	8.415	7.443	8.121	8.090	12.417
Sharpe ratio	-0.368	-0.132	-0.113	0.039	-0.020	-0.026	0.369	0.164

Real excess returns on the 8 currency portfolios in Lustig and Verdelhan (2007) for a US investor. Portfolio 1 contains currencies with the lowest interest rates, while portfolio 8 contains currencies with the highest interest rates.

Table 2: Empirical evaluation of the CAPM

	CU	Iterated	2S
	Uncentred SDF (λ)		
Market	0.043	0.039	0.038
	(0.005)	(0.005)	(0.004)
J test	21.375	22.003	28.068
	(0.006)	(0.005)	(0.000)
	Centred SDF (τ)		
Market	0.122	0.031	0.026
	(0.070)	(0.010)	(0.009)
J test	21.375	43.259	46.046
	(0.006)	(0.000)	(0.000)
	Regression (μ)		
Market	14.974	14.468	14.627
	(1.694)	(1.939)	(2.071)
J test	21.375	24.920	28.202
	(0.006)	(0.002)	(0.000)

This table displays estimates of "prices of risk" (λ, τ or μ) with standard errors in parenthesis, and the J test with p-values in parenthesis. We implement each method by CU, iterated and two-step GMM.

Table 3: Empirical evaluation of the (linearised) CCAPM

	CU	Iterated	2S
	Uncentred SDF (λ)		
Nondurables	0.495	0.488	0.488
	(0.042)	(0.043)	(0.043)
J test	5.663	5.691	5.711
	(0.580)	(0.576)	(0.574)
	Centred SDF (τ)		
Nondurables	4.388	1.154	1.201
	(5.727)	(0.477)	(0.181)
J test	5.663	16.925	91.624
	(0.580)	(0.018)	(0.000)
	Regression (δ)		
Nondurables	5.566	5.611	5.308
	(2.056)	(1.623)	(1.483)
J test	5.663	5.677	5.627
	(0.580)	(0.578)	(0.584)

This table displays estimates of "prices of risk" (λ, τ or δ) with standard errors in parenthesis, and the J test with p-values in parenthesis. We implement each method by CU, iterated and two-step GMM.

Table 4: Empirical evaluation of the (linearised) Epstein-Zin model

	CU	Iterated	2S
		Uncentred SDF (λ)	
Market	0.007	0.006	0.006
	(0.004)	(0.005)	(0.005)
Nondurables	0.476	0.479	0.474
	(0.052)	(0.054)	(0.051)
J test	4.785	4.827	5.171
	(0.686)	(0.681)	(0.639)
		Centred SDF (τ)	
Market	0.057		-0.002
	(0.053)		(0.006)
Nondurables	3.948		-0.706
	(3.659)		(0.220)
J test	4.785		69.285
	(0.686)		(0.000)
		Regression (μ, δ)	
Market	7.861	7.351	7.284
	(2.347)	(2.279)	(2.312)
Nondurables	5.746	5.582	4.748
	(2.130)	(1.638)	(1.271)
J test	4.785	4.964	5.283
	(0.686)	(0.664)	(0.626)

This table displays estimates of "prices of risk" (λ, τ or μ and δ) with standard errors in parenthesis, and the J test with p-values in parenthesis. We implement each method by CU, iterated and two-step GMM.

Supplemental Appendix

D Evaluation with gross returns

The SDF pricing equation is

$$E\left(\left(a+\mathbf{b}'\mathbf{f}\right)\mathbf{R}\right)=\ell_n,$$

where $\mathbf{R}$ is an $n\times 1$ vector of gross returns and ℓ_n is an $n\times 1$ vector of ones. Therefore, we find

$$E\left(\mathbf{R}\right)a+E\left(\mathbf{R}\mathbf{f}'\right)\mathbf{b}=E\left(\mathbf{R}\right)c+Cov\left(\mathbf{R},\mathbf{f}\right)\mathbf{b}=\ell_n,$$

where $c=a+\mathbf{b}'\boldsymbol{\mu}$ is the SDF mean. We can also relate the pricing of $\mathbf{R}$ to a constraint on the least squares intercepts $\boldsymbol{\phi}$ and slopes $\mathbf{B}$. Specifically,

$$E\left(\mathbf{R}\right)c+Cov\left(\mathbf{R},\mathbf{f}\right)\mathbf{b}=\boldsymbol{\phi}c+\mathbf{B}\mathbf{d}=\ell_n,$$

where $\mathbf{d}=\boldsymbol{\mu}c+\boldsymbol{\Sigma}\mathbf{b}$ is the shadow price of the factors, or their actual prices if they are traded.

When the factors are gross returns themselves, they also satisfy

$$E\left(\left(a+\mathbf{b}'\mathbf{f}\right)\mathbf{f}\right)=\ell_k, \tag{D6}$$

where ℓ_k is a $k\times 1$ vector of ones, which implies that

$$\mathbf{d}=\boldsymbol{\mu}a+\boldsymbol{\Gamma}\mathbf{b}=\boldsymbol{\mu}c+\boldsymbol{\Sigma}\mathbf{b}=\ell_k.$$

In that case, the least squares constraint simplifies to

$$\boldsymbol{\phi}c+\mathbf{B}\ell_k=\ell_n,$$

which can also be expressed as

$$\boldsymbol{\phi}=\kappa\left(\ell_n-\mathbf{B}\ell_k\right),\qquad \kappa=1/c,$$

when $c\neq 0$.

The influence functions of the SDF and regression methods with traded factors are

$$\mathbf{h}_S\left(\mathbf{R},\mathbf{f};a,\mathbf{b}\right)=\begin{bmatrix}\mathbf{R}\left(a+\mathbf{f}'\mathbf{b}\right)-\ell_n\\ \mathbf{f}\left(a+\mathbf{f}'\mathbf{b}\right)-\ell_k\end{bmatrix},$$

$$\mathbf{h}_R\left(\mathbf{R},\mathbf{f};\mathbf{B},\kappa\right)=\begin{bmatrix}\left(\mathbf{R}-\kappa\ell_n\right)-\mathbf{B}\left(\mathbf{f}-\kappa\ell_k\right)\\ vec\left(\left(\left(\mathbf{R}-\kappa\ell_n\right)-\mathbf{B}\left(\mathbf{f}-\kappa\ell_k\right)\right)\mathbf{f}'\right)\end{bmatrix},$$

and the J tests have $n-1$ degrees of freedom.

If the factors are not traded then we discard the moment condition (D6) that prices the factors. At the same time, we can re-express the least squares constraint as

$$\boldsymbol{\phi} = \kappa \ell_n + \mathbf{B}\varkappa, \qquad \kappa = 1/c, \qquad \varkappa = -\mathbf{d}/c.$$

when $c \neq 0$.

In this context, the influence functions of the SDF and regression approaches with non-traded factors are

$$\mathbf{g}_S(\mathbf{R}, \mathbf{f}; a, \mathbf{b}) = \left[\mathbf{R}\left(a + \mathbf{f}'\mathbf{b}\right) - \ell_n\right],$$

$$\mathbf{g}_R(\mathbf{R}, \mathbf{f}; \mathbf{B}, \kappa, \varkappa) = \begin{bmatrix} (\mathbf{R} - \kappa\ell_n) - \mathbf{B}(\mathbf{f} + \varkappa) \\ vec\left(((\mathbf{R} - \kappa\ell_n) - \mathbf{B}(\mathbf{f} + \varkappa))\,\mathbf{f}'\right) \end{bmatrix},$$

and the J tests will have $n - (k+1)$ degrees of freedom.

Counterpart to Proposition 1 with gross returns:

We can obtain the SDF influence functions as a transformation and reparametrisation of the regression influence functions augmented with exactly identified moment conditions. In particular, if we choose $(a, \mathbf{b})$ such that

$$a + \mathbf{b}'\boldsymbol{\mu} = 1/\kappa,$$

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

when $\kappa \neq 0$, then

$$\begin{bmatrix} a\mathbf{I}_n & \mathbf{b}' \otimes \mathbf{I}_n & a\mathbf{B} + \kappa(\ell_n - \mathbf{B}\ell_k)\mathbf{b}' & (\mathbf{b}' \otimes \mathbf{B})\mathcal{D} \end{bmatrix}$$
$$\times \begin{bmatrix} (\mathbf{R} - \kappa\ell_n) - \mathbf{B}(\mathbf{f} - \kappa\ell_k) \\ vec\left(((\mathbf{R} - \kappa\ell_n) - \mathbf{B}(\mathbf{f} - \kappa\ell_k))\,\mathbf{f}'\right) \\ \mathbf{f} - \boldsymbol{\mu} \\ vech\left(\mathbf{f}\mathbf{f}' - \boldsymbol{\Gamma}\right) \end{bmatrix}$$
$$= \left[\mathbf{R}\left(a + \mathbf{b}'\mathbf{f}\right)\right] - \left[\kappa(\ell_n - \mathbf{B}\ell_k)\left(a + \mathbf{b}'\boldsymbol{\mu}\right) + \mathbf{B}(\boldsymbol{\mu} a + \boldsymbol{\Gamma}\mathbf{b})\right]$$
$$= \left[\mathbf{R}\left(a + \mathbf{b}'\mathbf{f}\right)\right] - \left[(\ell_n - \mathbf{B}\ell_k) + \mathbf{B}\ell_k\right] = \mathbf{R}\left(a + \mathbf{b}'\mathbf{f}\right) - \ell_n,$$

where the matrix $\mathcal{D}$ denotes the appropriate duplication matrix, and

$$\begin{bmatrix} \mathbf{0} & \mathbf{0} & a\mathbf{I}_k & (\mathbf{b}' \otimes \mathbf{I}_k)\mathcal{D} \end{bmatrix} \begin{bmatrix} (\mathbf{R} - \kappa\ell_n) - \mathbf{B}(\mathbf{f} - \kappa\ell_k) \\ vec\left(((\mathbf{R} - \kappa\ell_n) - \mathbf{B}(\mathbf{f} - \kappa\ell_k))\,\mathbf{f}'\right) \\ \mathbf{f} - \boldsymbol{\mu} \\ vech\left(\mathbf{f}\mathbf{f}' - \boldsymbol{\Gamma}\right) \end{bmatrix} =$$

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

Given these relationships, we can apply similar arguments to those in the proof of Proposition 1. □

Counterpart to Proposition 2 with gross returns:

We can obtain the SDF influence functions as a transformation and reparametrisation of the regression influence functions augmented with exactly identified moment conditions. In particular, if we choose $(a, \mathbf{b})$ such that

$$a + \mathbf{b}'\boldsymbol{\mu} = 1/\kappa,$$
$$\boldsymbol{\mu} a + \boldsymbol{\Gamma}\mathbf{b} = -\varkappa/\kappa,$$

when $\kappa \neq 0$, then

$$\begin{aligned}
&\left[\begin{array}{cccc} a\mathbf{I}_n & \mathbf{b}' \otimes \mathbf{I}_n & a\mathbf{B} + (\kappa\ell_n + \mathbf{B}\varkappa)\,\mathbf{b}' & (\mathbf{b}' \otimes \mathbf{B})\,\mathcal{D} \end{array}\right] \\
&\qquad \times \left[\begin{array}{c} (\mathbf{R} - \kappa\ell_n) - \mathbf{B}\,(\mathbf{f} + \varkappa) \\ vec\left(\left((\mathbf{R} - \kappa\ell_n) - \mathbf{B}\,(\mathbf{f} + \varkappa)\right)\mathbf{f}'\right) \\ \mathbf{f} - \boldsymbol{\mu} \\ vech\left(\mathbf{f}\mathbf{f}' - \boldsymbol{\Gamma}\right) \end{array}\right] \\
&= \left[\mathbf{R}\left(a + \mathbf{b}'\mathbf{f}\right)\right] - \left[(\kappa\ell_n + \mathbf{B}\varkappa)\left(a + \mathbf{b}'\boldsymbol{\mu}\right) + \mathbf{B}\left(\boldsymbol{\mu} a + \boldsymbol{\Gamma}\mathbf{b}\right)\right] \\
&= \left[\mathbf{R}\left(a + \mathbf{b}'\mathbf{f}\right)\right] - \left[\ell_n + \mathbf{B}\left(\varkappa/\kappa\right) + \mathbf{B}\left(-\varkappa/\kappa\right)\right] = \mathbf{R}\left(a + \mathbf{b}'\mathbf{f}\right) - \ell_n,
\end{aligned}$$

where the matrix $\mathcal{D}$ denotes the appropriate duplication matrix. Given this relationship, we can apply similar arguments to those in the proof of Proposition 2. □

E Empirical Application: Implied Estimates

Table E1: CAPM

		CUE	Iterated	U. SDF	C. SDF	Reg.	2S	U. SDF	C. SDF	Reg.
λ		0.043		0.039	0.025	0.037		0.038	0.022	0.037
π	1	-0.929		-1.061	-1.525	-1.141		-1.084	-1.614	-1.117
	2	-0.277		-0.333	-0.529	-0.367		-0.342	-0.567	-0.356
	3	-0.592		-0.607	-0.658	-0.616		-0.609	-0.668	-0.613
	4	1.514		1.403	1.012	1.335		1.384	0.937	1.356
	5	-0.284		-0.272	-0.228	-0.264		-0.270	-0.220	-0.267
	6	-0.265		-0.260	-0.243	-0.257		-0.259	-0.240	-0.258
	7	2.624		2.658	2.778	2.679		2.664	2.801	2.673
	8	0.670		0.798	1.247	0.875		0.819	1.332	0.851
	Market	-9.216		-7.694	-2.355	-6.775		-7.435	-1.337	-7.058
τ		0.122		0.098	0.031	0.078		0.091	0.026	0.082
ψ	1	-2.622		-2.660	-1.906	-2.427		-2.562	-1.908	-2.463
	2	-0.781		-0.834	-0.662	-0.780		-0.809	-0.670	-0.786
	3	-1.671		-1.520	-0.822	-1.309		-1.440	-0.789	-1.351
	4	4.271		3.514	1.265	2.839		3.270	1.108	2.990
	5	-0.802		-0.681	-0.285	-0.562		-0.637	-0.260	-0.588
	6	-0.748		-0.652	-0.303	-0.547		-0.613	-0.283	-0.569
	7	7.404		6.660	3.472	5.696		6.297	3.311	5.894
	8	1.889		1.999	1.558	1.861		1.937	1.575	1.877
	Market	-26.002		-19.278	-2.944	-14.405		-17.574	-1.581	-15.565
μ		14.974		15.383	8.045	14.468		15.035	6.963	14.627
α	1	-1.369		-1.494	-2.179	-1.784		-1.802	-2.229	-1.817
	2	-0.421		-0.444	-0.646	-0.533		-0.507	-0.687	-0.511
	3	-0.784		-0.800	-0.988	-0.868		-0.879	-0.986	-0.882
	4	0.461		0.466	0.947	0.536		0.539	1.025	0.548
	5	-0.310		-0.318	-0.464	-0.413		-0.445	-0.478	-0.449
	6	-0.368		-0.368	-0.290	-0.347		-0.339	-0.244	-0.335
	7	2.156		2.284	3.009	2.493		2.465	3.083	2.479
	8	0.928		0.856	1.558	0.719		0.800	1.725	0.815
	Market	-7.999		-8.408	-1.070	-7.493		-8.060	0.012	-7.652

Table E2: CCAPM

		CUE	Iterated	U. SDF	C. SDF	Reg.	2S	U. SDF	C. SDF	Reg.
λ		0.495		0.488	0.411	0.496		0.488	0.530	0.491
π	1	-0.645		-0.668	-0.931	-0.641		-0.667	-0.526	-0.660
	2	-0.994		-0.992	-0.974	-0.994		-0.992	-1.003	-0.993
	3	-0.446		-0.450	-0.497	-0.445		-0.450	-0.425	-0.449
	4	-0.114		-0.108	-0.039	-0.115		-0.108	-0.145	-0.110
	5	-0.696		-0.688	-0.604	-0.697		-0.689	-0.734	-0.691
	6	-0.320		-0.319	-0.302	-0.320		-0.319	-0.328	-0.319
	7	-0.462		-0.415	0.122	-0.469		-0.416	-0.705	-0.432
	8	0.377		0.399	0.657	0.373		0.399	0.260	0.391
τ		4.388		4.092	1.154	4.474		4.099	1.201	4.171
ψ	1	-5.715		-5.596	-2.613	-5.784		-5.600	-1.191	-5.606
	2	-8.809		-8.315	-2.732	-8.966		-8.328	-2.272	-8.440
	3	-3.952		-3.771	-1.395	-4.016		-3.776	-0.963	-3.813
	4	-1.012		-0.907	-0.110	-1.038		-0.909	-0.330	-0.938
	5	-6.166		-5.768	-1.694	-6.285		-5.778	-1.664	-5.874
	6	-2.838		-2.671	-0.848	-2.890		-2.675	-0.743	-2.714
	7	-4.091		-3.476	0.343	-4.228		-3.490	-1.598	-3.669
	8	3.340		3.346	1.843	3.367		3.347	0.590	3.325
δ		5.566		5.270	2.758	5.611		5.282	2.545	5.308
α	1	-0.368		-0.389	-1.575	-0.324		-0.365	-1.256	-0.337
	2	-0.900		-0.921	-1.650	-0.828		-0.921	-1.765	-0.885
	3	-0.935		-0.935	-1.270	-0.894		-0.989	-2.118	-0.909
	4	-0.567		-0.577	-0.909	-0.526		-0.630	-1.454	-0.548
	5	-0.440		-0.443	-0.887	-0.402		-0.430	-1.076	-0.389
	6	-0.732		-0.731	-0.676	-0.707		-0.773	-1.458	-0.734
	7	-0.743		-0.738	0.043	-0.675		-0.791	0.306	-0.689
	8	-1.170		-1.114	0.100	-1.208		-1.260	-1.876	-1.230

Figure 1: CU-GMM criterion function for the CAPM

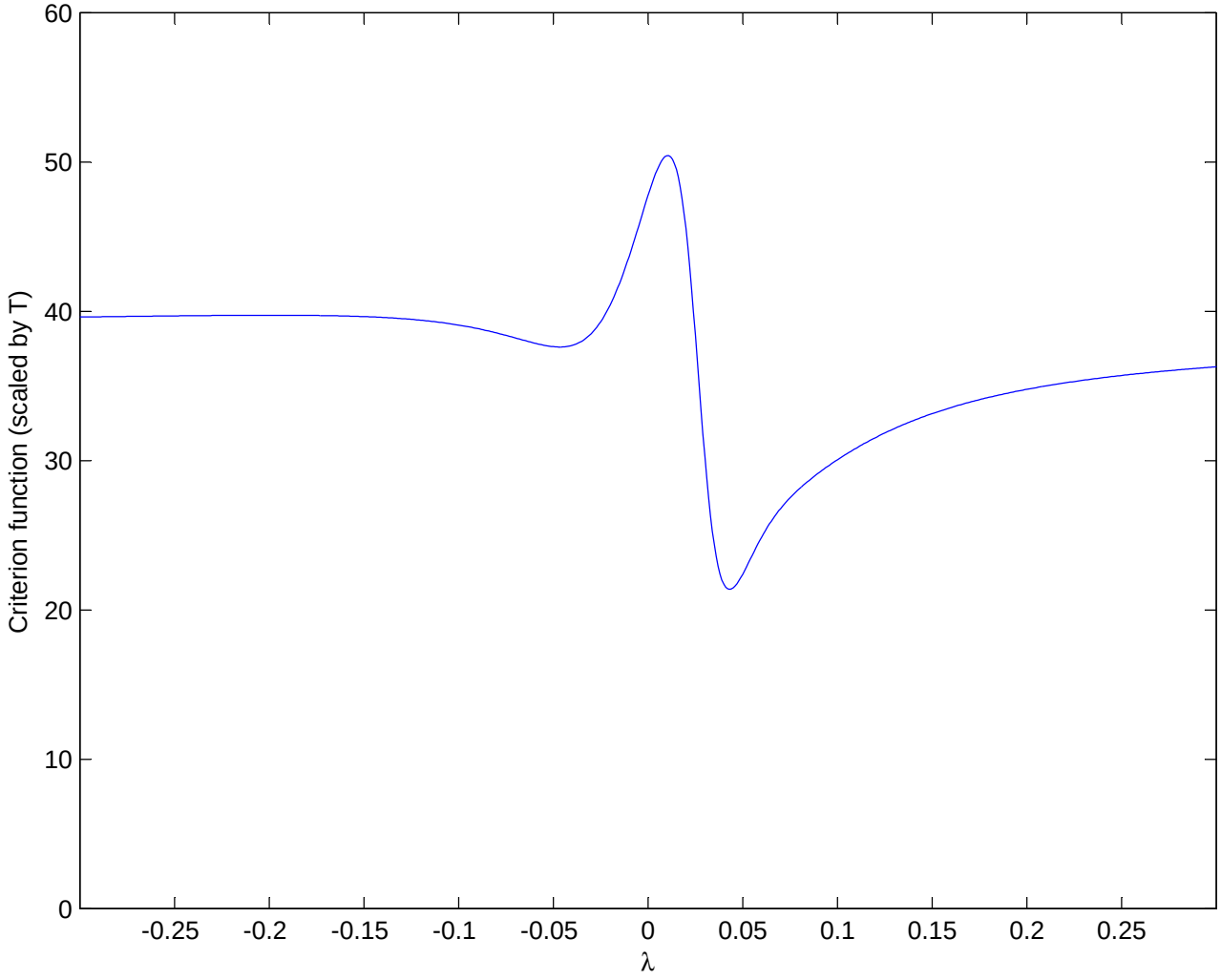

Figure 2: CU-GMM criterion function for the CCAPM

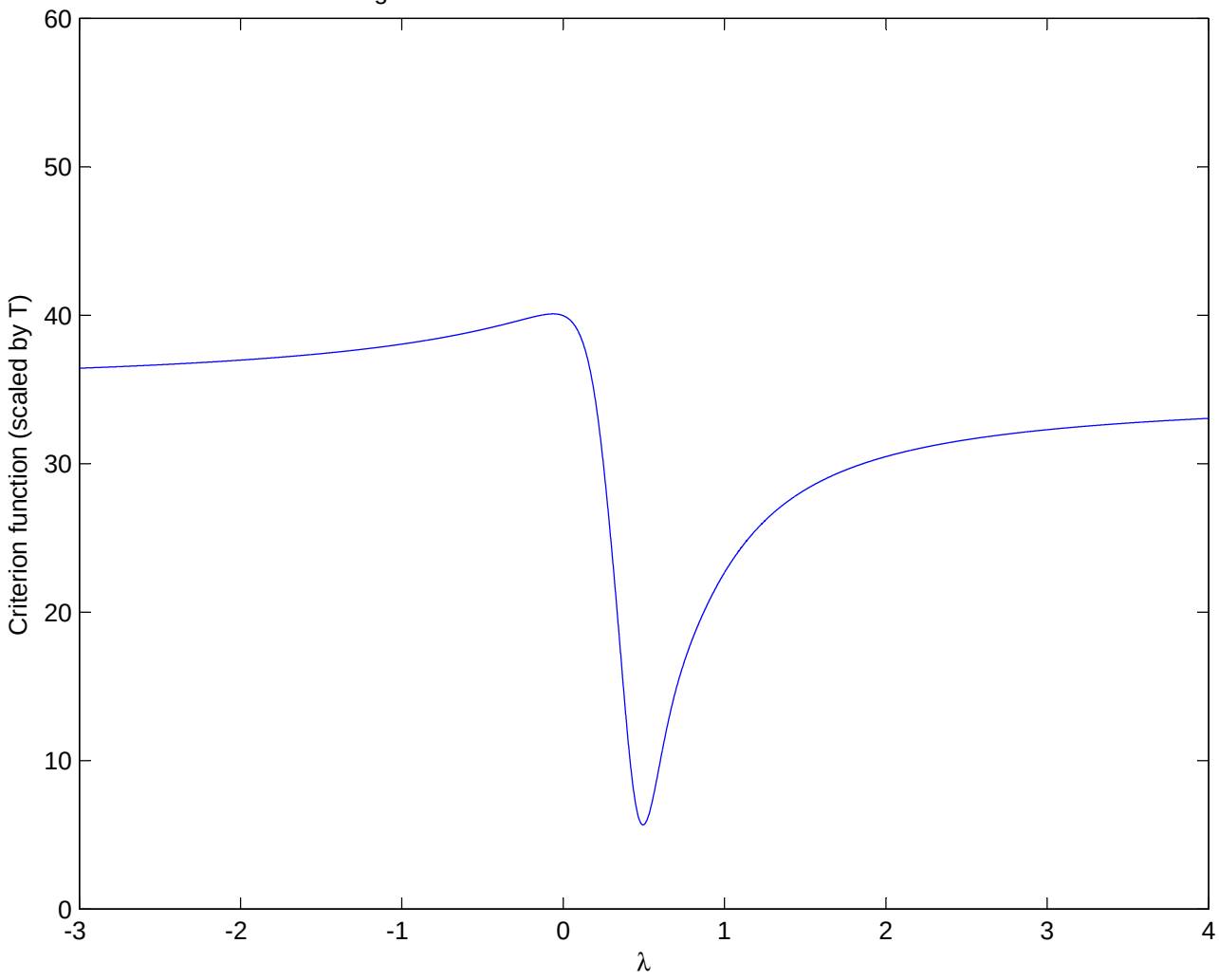

Figure 3: Contours of the CU-GMM criterion function (scaled by T) for the Epstein-Zin model