

Portfolio Performance of Linear SDF Models: An Out-of-Sample Assessment

Massimo Guidolin* Erwin Hansen† Martin Lozano‡

January 14, 2013

Abstract

In this paper, we evaluate linear stochastic discount factor models using a portfolio metric: the realized out-of-sample Sharpe ratio of mean-variance portfolios backed by alternative linear factor models. Using a sample of monthly US portfolio returns data spanning the period 1927 - 2011, we provide evidence that multifactor linear models have better empirical properties than the CAPM, not only when the cross-section of expected returns is evaluated in-sample, but also when this economic metric is used out-of-sample. When we compare a portfolio associated to a multifactor model and a portfolio associated to the CAPM, we document differences of Sharpe ratio of up to 10 percent.

EFM Codes: 310, 380.

*CAIR, Manchester Business School, and IGIER, Bocconi University. Address: MBS Crawford House, Manchester, M13 9PL, UK. Tel: +44(0)1612756406. Email: massimo.guidolin@mbs.ac.uk.

†Manchester Business School and Universidad de Chile. Address: MBS Crawford House, Manchester, M13 9PL, UK. Tel: +44(0)1612754585. Email:erwin.hansen@postgrad.mbs.ac.uk.

‡Financial Mathematics and Computation Cluster (FMC2), School of Business, University College Dublin. Address: Belfield, Dublin 4, Ireland. Tel: +353(1)7164786. Email: mlozanoqf@gmail.com. Lozano acknowledges the support of Science Foundation Ireland under Grant Number 08/SRC/FM1389.

1 Introduction

Linear factor asset pricing models are well established in finance. The CAPM model of Sharpe (1964) and Lintner (1965), the three-factor model of Fama and French (1993), and the four-factor model of Carhart (1997) are extensively used by researchers and practitioners to compute the cost of capital and risk-adjusted returns. Previous empirical evidence evaluating their ability to fit the cross-section of mean returns has favored multifactor models over the CAPM. The Carhart model, for example, has been particularly successful in accounting for most of the anomalies challenging the efficient market hypothesis.¹ In this literature, models are usually evaluated by comparing measures of in-sample goodness-of-fit, such as the R^2 of a cross-sectional regression of mean excess returns on a set of factor mimicking portfolio returns.² This paper evaluates and compares linear factor models from an out-of-sample economic perspective. In particular, we study the out-of-sample realized performance of mean-variance efficient portfolios, whose test assets are the predicted excess returns generated by different linear factor models.

We aim at shedding light on the empirical performance of linear factor models using an economic metric instead of focussing on a pure econometric perspective. Even though, theoretically, linear pricing models are intrinsically connected to mean-variance efficient portfolios, it is common to observe that portfolio choice papers ignore asset pricing models in the estimation of optimal weights.³ Usually, a functional form of the distribution of returns is assumed and then a historical sample is used to estimate the parameters of interest. Despite of our main focus laying in the evaluation of linear pricing models more than in the associated investment strategies, we use this link to construct an alternative evaluation measure in the empirical asset pricing literature. Different views about the sources of risk must affect the portfolio allocation of a mean-variance investor, therefore, the consequences of imposing such a view have economically relevant effects of which we take advantage to study a set of linear factor models. We also study to which extent there is a correspondence between the in-sample and the out-of-sample performance of the linear factor models.

A priori, the theory is silent regarding our two main point of interest: the relative performance of the models under this out-of-sample economic metric and the existence of a link between the in-sample and the out-of-sample performance of the models. It is perfectly possible that models performing well in-sample may show a poor performance out-of-sample, or the other way around. We hypothesize that both metrics are intrinsically connected, at least from an empirical point of view. A model producing lower pricing errors in the cross-section of expected returns might be associated to a mean-variance portfolio, outperforming competitive models, since the investor is as close to the true data-generating process as she can be. If this is true, we expect to observe a higher

¹See Schwert (2003) and references therein.

²Recently, Lewellen et al. (2010) argue that the GLS R^2 is a more appealing statistic than the standard OLS R^2 to evaluate and compare asset pricing models. Following the advice by Lewellen et al. (2010), we report GLS R^2 is this paper.

³See the description of the plug-in method in Brandt (2010) and references therein.

out-of-sample performance in multifactor models that previous literature identifies as models with lower pricing errors.

A first glance at the evidence reported in this paper is given in Figure 1. We plot the in-sample root mean squared error (RMSE) against the out-of-sample Sharpe ratio of several linear factor models.⁴ Each dot represents a particular model and the number attached to it indicates the number of factors in the model. Three facts emerge clearly. First, the best performing models in-sample are, at the same time, the best performing models out-of-sample, according to this economic metric (this is clear from the negative slope in the plot). Second, multifactor models outperform the CAPM. The cloud of points of multifactor models is located at the lower right side, whilst the cloud of points of CAPMs is located at the upper left side of the plot. This result is well known from an in-sample perspective but new from our out-of-sample perspective. Third, the performance of some of these mean-variance portfolios is comparable with the $1/N$ investment strategy of De Miguel et al. (2009). Some points are located to the right of the dashed line representing the highest performance achieved by the $1/N$ strategy in our sample, across models and specifications.

We conduct a formal statistical analysis using a sample of monthly US portfolios from 1927 to 2011 that corroborate our main hypothesis. We do find that the best performing models in-sample (higher GLS R^2 or lower RMSE) are, simultaneously, those with higher realized out-of-sample Sharpe ratio one period ahead. In particular, we find that multifactor models achieve consistently higher out-of-sample Sharpe ratios than the CAPM does. The estimated difference of Sharpe ratios between models may reach values as high as 10 per cent. These results are valid for alternative assets in the investment menu, first and second stage GMM estimators, centered and uncentered versions of the stochastic discount factor, a sub-sample analysis, and 5-year and 10-year windows in the rolling estimation setup. Moreover, we study the ability of the 4-factor liquidity model of Pastor and Stambaugh (2003) in explaining the cross-section of expected returns and producing high out-of-sample Sharpe ratios. Our results indicate that even though the liquidity factor model is outperformed by the Carhart model in-sample, it is able to outperform the other models, in terms of out-of-sample Sharpe ratio, in many of the cases studied. Finally, we study the effects of imposing short sales constraints on the portfolio weights estimation. Jagannathan and Ma (2003) argue that a trade-off between specification error and estimation error exist when short sales constraint are imposed. We find that short sales constraints help to achieve higher Sharpe ratios when the full sample is studied. However, this is not the case for the sub-sample covering the second half of our sample (1968-2011).

Our empirical strategy consists of two stages. In the first stage, we estimate linear factor models using the SDF/GMM method as described by Cochrane (2005).⁵ There are two reasons to prefer the SDF/GMM method over the more standard beta method. First, the SDF representation is more

⁴Models change according to the numbers of factors included, the GMM estimator used and the normalization of the stochastic discount factor considered. See more details about this characterization of the models in section 2.

⁵The SDF method consists of estimation of a linear factor model in its stochastic discount form by GMM.

general, and we want to keep our model as general as possible. Second, Jagannathan and Wang (2002) show that the SDF form is comparable to the beta form in terms of the efficiency in the estimation of risk premia and in terms of the power of the specification tests. More recently, Lozano and Rubio (2011) show that multifactor models estimated using the SDF method, in particular, the first stage GMM estimator, produce lower pricing errors than the beta method does. Because we are interested in using as reliable as possible measures of in-sample goodness of fit, this property of the SDF method is particularly appealing for our analysis.⁶ In the second stage, we estimate mean-variance efficient portfolios using the predicted excess returns from the linear factor models. Theoretically, we use the connection between the stochastic discount factor and the mean-variance frontier established by Hansen and Richard (1987). In practical terms, we estimate the weights of the mean-variance efficient portfolio using the concept of mean-representing portfolio introduced by Chamberlain and Rothschild (1983). As it is shown by Penaranda and Sentana (2009, 2011), the mean-variance frontier can be consistently estimated by GMM using a set of moment conditions implied by the mean-representing portfolio definition. Finally, we compare the out-of-sample performance of the mean-variance efficient portfolios associated to linear factor models, one period ahead, using the test of difference in Sharpe ratios of Leidot and Wolf (1998). This is a circular-block bootstrap test that takes into account both non-normalities and serial correlation, usually presented in the empirical distribution of returns.

Our paper contributes to two strands of the literature. First, we provide new empirical evidence for the literature evaluating the performance of linear factor asset pricing models, see Jagannathan and Wang (2002), Shanken and Zou (2007), Lewellen et al. (2010), Lozano and Rubio (2011), among others. This literature has developed a pure in-sample econometric perspective to evaluate the performance of alternative linear factor models. We aim to contribute to this literature by providing evidence regarding the performance of the models when they are evaluated using an out-of-sample financial metric. First, we document the existence of an empirical link between the in-sample statistical performance of the models and their out-of-sample portfolio performance: models with lower pricing errors in-sample exhibit, at the same time, higher out-of-sample Sharpe ratios. Moreover, we show that multifactor models outperform the CAPM under this metric. Second, we contribute to a less-developed literature studying how asset pricing models can provide useful insights to an investor solving a portfolio choice problem (see for example Brandt, 1999; Pastor and Stambaugh, 2000; MacKinlay and Pastor, 2000; Pastor, 2000; Chevrier and McCulloch, 2008). In this literature, our aim is to propose the use of a methodology that combines the estimation of linear factor models using the SDF method and the estimation of mean-variance efficient portfolios using a consistent GMM estimator. Previous studies only consider the estimation of the linear models under the beta representation. Finally, we show that, in some cases, the portfolios associated to multifactor models can yield a higher out-of-sample Sharpe ratio than the $1/N$ strategy of De Miguel

⁶Lozano and Rubio (2011) also document that the beta method dominates the SDF method in terms of producing more precise estimates of the risk premium when multifactor models are evaluated.

et al. (2009), indicating that these models may potentially become a worthy tool in the hands of investors looking for high risk-adjusted portfolio returns.

Just a few papers have investigated the impact of using linear asset pricing models in a portfolio choice problem.⁷ Pastor and Stambaugh (2000) and Pastor (2000) use risk-based and characteristic-based asset pricing models to center the prior beliefs of a Bayesian investor solving a mean-variance asset allocation problem. In particular, Pastor and Stambaugh (2000) study how the optimal portfolio changes depending on the model used to center the priors. They conclude that portfolio choices do not change significantly even after considering mispricing uncertainty and margin requirements. Pastor (2000) studies the home bias and the value and size effects in a portfolio choice context. He concludes that US investors should have strong priors on their beliefs (factor models) to explain observed biases in both phenomena. Chevrier and McCulloch (2008), in a Bayesian environment as well, study how economically motivated priors may help to build portfolios that outperform the equally-weighted portfolio ($1/N$). Linear factor models are used by the author as one of the set of motivated priors in the Bayesian analysis. Mackinlay and Pastor (2000) study the implications of assuming that asset returns have an exact factor structure on the estimation of the expected returns of mean-variance portfolios. The authors conclude that in a model with one unobserved factor, the covariance matrix of returns collapses to the identity matrix, and the associated mean-variance portfolio outperforms the benchmark portfolio in terms of out-of-sample returns. More recently, Tu and Zhou (2011) show that among the theoretical-motivated portfolios that they analyse, Mackinlay and Pastor's model is the only one exhibiting a solid performance in terms of Sharpe ratios and certainty equivalent.⁸ Behr et al. (2012) studies how industry momentum can be use to improved the performance of minimum variance portfolios with a parametric portfolio policy. Among several benchmark models, they study the performance of minimum variance portfolios estimated with a covariance matrix associated to linear factor models (CAPM, the Fama and French 3-factor model, and the Carhart model). The three analysed portfolios yield Sharpe ratios of approximately 13 per cent. These portfolios - the ones linked to linear factor models - have higher Sharpe ratios than most of the other portfolios under analysis.⁹ Finally, Kirby and Ostdiek (2012) argue that the success of the $1/N$ strategy in outperforming mean-variance portfolios, documented by De Miguel et al. (2009), is due the research design employed by the authors, which overweight portfolios subject to high estimation risk and extreme turnover. The authors propose two new mean-variance portfolio allocation methods - volatility timing and reward-to-risk timing strategies - which are able to outperform the $1/N$ portfolio. In this context, the authors employ the Fama and French 3-factor model, and the Carhart model, to produce conditional expected returns, which are later used in the implementation of the new proposed mean-variance methods. The authors provide evidence

⁷A general overview of the role of factor models in portfolio and asset pricing is given by Connor and Korajczyk (2010).

⁸Tu and Zhou (2011) study the performance of four portfolio models: Markowitz mean-variance, Jorion (1986), MacKinlay and Pastor (2000) and Kan and Zhou (2007) portfolios.

⁹The performance of the CAPM is slightly inferior to that of multifactor models (0.12 vs 0.13).

that the Carhart model, when used by the mean-variance reward-to-risk timing strategy, yields promising results, beating the $1/N$ portfolio.

The rest of the paper is organized as follows. In section 2, we describe our methodology. In the first subsection, we describe the GMM estimation of linear asset pricing models; in the second subsection, we introduce the concept of mean-representing portfolio of Chamberlain and Rothschild (1983) and we describe the GMM estimator proposed by Penaranda and Sentana (2009, 2011) to compute the portfolio weights; in the third subsection, we describe the test of difference in Sharpe ratios of Ledoit and Wolf (2008). In section 3, we describe our dataset. The estimation results are reported in section 4. Finally, we conclude in section 5.

2 Methodology

The main goal of our work is to estimate mean-variance efficient portfolios which are backed by alternative linear factor asset pricing models, in order to evaluate them according to their out-of-sample Sharpe ratios. To fulfill this task, we use a methodology which is characterized by the following set of assumptions. First, we decide to employ a two-stage approach: in the first stage, linear factor models are estimated under its SDF representation, and then, in the second stage, we estimate mean-variance efficient portfolios using the fitted returns obtained from the first stage. Second, is the use in the estimation process of the SDF form of the asset pricing models instead of the more standard beta representation. Third, we use GMM as our estimation method.¹⁰ Finally, and in order to keep our estimation process as consistent as possible across the two stages, we have decided to continue using GMM in the estimation of mean-variance efficient portfolios in the second stage. As it is explained in more detail below, we make use of a set of moment conditions that deliver consistent mean-variance frontier estimates.

Before going into the details, some further motivation of our listed assumptions follows. Regarding the use of the SDF form, we have decided to use them because our main interest in this work is not the estimation of the risk premia but the accurate estimation of pricing errors. Previous literature has identified precisely this characteristic - more accurate estimates of the pricing errors of the model - as one of the main strength of the use of the SDF form.¹¹ Regarding the estimation of the mean-variance efficient portfolios, we make use of a theorem, introduced by Chamberlein and Rothschild (1983), that allow us to identify these portfolios making use of an auxiliary portfolio: the mean-representing portfolio. The weights of this portfolio are proportional to the weights of the mean-variance efficient portfolio, therefore, both have the same Sharpe ratio.¹² Intuitively, this portfolio

¹⁰Even though it is common to combine the SDF form and GMM estimation in recent papers in empirical asset pricing, we must remind the reader that this is not mandatory, it is a choice made by the econometrician. As it is well pointed out by Jagannathan and Wang (2002) and Cochrane (2005), it could be perfectly possible to estimate the SDF form by alternative estimation methods, as the maximum likelihood for instance.

¹¹See Jagannathan and Wang (2002) and Lozano and Rubio (2011).

¹²Britten-Jones (1999) develops a portfolio's weights estimator using OLS. In her methodology, an OLS regression

can be interpreted as the portfolio that "mimics" the safe asset with minimum tracking error. The estimation of this portfolio is completed by GMM using the set of moments conditions introduced by Penaranda and Sentana (2009, 2011). The main advantage of using this methodology is that these authors show that under weak assumptions, this procedure delivers a consistent estimate of the mean-variance efficient frontier. Moreover, as we mentioned above, using this procedure allow us to estimate both stages by the same method, GMM.

The two-stage methodology is implemented in a moving rolling window setup. For each 60-months window, we estimate the linear factor models and the weights of the mean-representing portfolio. For each of these portfolios, we save the realized ex-post return in the following month. We only evaluate the out-of-sample performance of the models one period in advance since the estimated mean-variance frontier is static, therefore, any analysis beyond one-period will not be consistent with this characteristic of mean-variance model. We repeat this procedure for each of the 959 windows in our sample and we obtain a time series of realized ex-post returns for each portfolio. From these time-series, we compute the out-of-sample Sharpe ratio of each model, and finally, we compare these Sharpe ratios using the test of Ledoit and Wolf (2008).

2.1 SDF/GMM estimation of linear asset pricing models

In this paper, the SDF is approximated by a linear factor model, which is later estimated by GMM. In this subsection, I briefly explain what the SDF is, how the linear structure is explicitly incorporated into the models, and how GMM estimators are defined. Our starting point is the fundamental asset pricing equation given by:

$$p_t = E_t [m_{t+1}x_{t+1}], \quad (1)$$

where p_t is the asset price at time t , E_t is the expectation operator conditional on the information up to time t , m_{t+1} is the stochastic discount factor valid between t and $t + 1$, and x_{t+1} is the asset payoff in period $t+1$. Equation (1) is valid for any asset. In particular, for the case of excess returns (r), which are the payoffs analyzed in this paper, it reduces to:

$$0 = E_t [m_{t+1}r_{t+1}]. \quad (1a)$$

This expression tells us that the expected excess returns of any asset, after being discounted to the present time by the stochastic discount factor m_{t+1} , are zero. It is worth noting that this equation is valid either for an individual asset or for a set of assets. In the first case, we will have a simple

between a vector of ones on a matrix containing the asset returns is run. The beta coefficients are the estimated weights that are also proportional to the weights of the mean-variance efficient portfolio. In this regard, our methodologies share the same spirit. Indeed, the set of moments that define the mean-representing portfolio may be the same set of moments defining the OLS estimates in Britten-Jones (1999).

equation, whereas in the second case we will have a system of equations. For example, if there are N assets in the investment menu, r and 0 are $(N \times 1)$ vectors of gross excess returns and zeros, respectively. On the other hand, m_{t+1} is the realization of any random variable satisfying (1a) between t and $t + 1$. Some additional assumptions need to be imposed in order for m_{t+1} be unique and positive.¹³

An attractive feature of the SDF model is that it is well connected with economic theory since it tracks the dynamic of the intertemporal rate of substitution of aggregate consumption across periods. For the simple case of 2 periods, from the first-order condition for consumption, it is possible to identify the SDF as $m_{t+1} \equiv \beta \frac{U'(C_{t+1})}{U'(C_t)}$, where β is the subjective discount factor faced by the representative agent, C_t denotes aggregate consumption at time t , and $U' \equiv \left(\frac{\partial U}{\partial c}\right)$ is the marginal utility of the aggregate consumption. In this paper, we avoid making an assumption on the functional form of the utility of the representative agent, we approximate the dynamic of the SDF through linear risk factors instead.¹⁴

In particular, we assume the following functional form for the SDF:

$$m_{t+1} = a - b'f_t, \tag{2}$$

where b is a $(K \times 1)$ vector of parameters to be estimated and f_t are the realizations of K risk factors at time t . As it has been noted in previous empirical work, further assumptions on the constant term a are required to identify the parameters of interest, b . The intuition of the lack of identification is as follows. Suppose that \hat{m} is the estimated SDF. Therefore, from equation (1a), it holds that $E[\hat{m}r] = 0$. Now, for any constant c , the SDF $\tilde{m} = c\hat{m}$ also satisfies (1a), i.e. $E[\tilde{m}r] = 0$. From this example, it is clear that exist an infinite number of SDFs satisfying (1a) simultaneously. This is the identification problem. This problem is solved normalizing the value of the constant a in (2). As it is pointed out by Cocharane (2005), the choice of the normalization is one of convenience. The first and most simple normalization is to impose $a = 1$. In this case we say that the SDF is uncentered. The second normalization is $a = 1 + b'E(f)$, which correspond to the centered SDF case.¹⁵

After imposing a normalization on a , the set of parameters b is estimated by GMM using the pricing errors from (1a) as the relevant moment conditions. The GMM estimation is particularly attractive because it is efficient even though the residuals of the estimated model are non-spherical errors (autocorrelation, heteroskedasticity and non-normality), which is indeed the case for excess return data.

¹³For uniqueness, the assumption of complete markets is necessary. For positiveness, both absence of arbitrage and the law of one price are required. See Chapter 4 in Cochrane (2005) for further details on these results.

¹⁴Cochrane (2005, ch. 9) discusses the use of factor models in empirical asset pricing.

¹⁵The properties of SDF asset pricing models under both normalizations have been studied by Burnside (2007) and Lozano and Rubio (2011).

Let $g_T(b)$ be the mean value of the pricing errors of a particular linear factor SDF model¹⁶:

$$g_T(b) \equiv E_T [m_t(b)r_t]. \quad (3)$$

Then the GMM estimates of b is the set of parameters minimizing the following quadratic form

$$\min_b g_T(b)'Wg_T(b), \quad (4)$$

where W is a weighting matrix of the moment conditions. First-stage GMM estimates are obtained setting W equal to the identity matrix, $W = I_N$. In this case \hat{b} minimizes the sum of squared pricing errors like in the standard OLS estimation method. In this stage, therefore, each asset in the investment menu receives equal weight in the estimation procedure. This first-stage GMM estimator is a consistent estimator of the true parameter vector and it is asymptotically normally distributed. Second stage GMM estimates are obtained setting W equal to the inverse of the spectral density matrix defined as $S \equiv \sum_{j=-\infty}^{\infty} E [u_t u'_{t-j}]$, where $u_t \equiv m_t(\hat{b})r_t$ is the vector of pricing errors of the model. In practice, u_t is obtained using \hat{b} from the first stage. For the case where $W = S^{-1}$, the GMM estimator is optimal in the sense of producing the lowest possible asymptotic variances for the estimated coefficients.

Now, we describe the GMM estimators associated to each of the two normalizations of a .

2.1.1 Uncentered SDF/GMM

Imposing the normalization $a = 1$ in equation (2), the mean values of the pricing errors of the linear factor model are

$$g_T(b) = -E_T(mr) = -E_T(r) + E_T(rf')b.$$

Let d be the derivative of $g_T(b)$ with respect to b :

$$d = -\frac{\partial g_T(b)}{\partial b'} = E_T(rf'),$$

which is equal to the second-moment matrix of excess returns and factors. Then, the first order condition from (4) is given by:

$$-d'W[E_T(r) - db] = 0.$$

¹⁶Note that we use the notation $E_T \equiv \frac{1}{T} \sum_{t=1}^T ()$ in the expression below.

Finally, the first-stage and second-stage GMM estimates of b are defined respectively as:

$$\widehat{b}_1 = (d'd)^{-1}d'E_T(r) \quad (5)$$

$$\text{and } \widehat{b}_2 = (d'S^{-1}d)^{-1}d'S^{-1}E_T(r) \quad (6)$$

From these expressions we observe that the GMM estimators are cross-sectional regressions of the mean excess return of the assets, $E_T(r)$, on the matrix of second moments of excess returns and factors, d . In fact, there is a clear parallel between these closed form solutions and the formulae of the OLS and the GLS estimators. The first-stage GMM estimator is a simple OLS regression between $E_T(r)$ and d , whereas the second-stage GMM estimator is the GLS version of the same regression, where the weighting matrix is the inverse of the spectral density matrix, S .

2.1.2 Centered SDF/GMM

The second normalisation is $a = 1 + b'E(f)$, where $E(f)$ is the expected value of the factors. The main advantage of this normalization is that the cross-sectional regressions defining the GMM estimators contain covariances between excess returns and factors, instead of second moments. Usually, researchers are more familiar with these formulas as they are closer to the definition of betas. It is worth noting that after imposing the centered normalization to the SDF, we have, that on average, the SDF equals one, $E(m) = 1$.¹⁷ For this case, the mean value of pricing errors are given by

$$g_T(b) = E_T(mr) = E_T(r) - E_T(r\tilde{f}')b,$$

where $\tilde{f} \equiv f - E(f)$. Its partial derivative with respect to b is

$$d = \frac{\partial g_T(b)}{\partial b'} = E_T(r\tilde{f}').$$

Here d is the covariance matrix between excess returns and factors. The first order condition of the GMM minimization problem is given by

$$-d'W[E_T(r) - db] = 0.$$

Using the sample counterpart of $E(f)$, we estimate d . Then, the first and second stage GMM estimators can be obtained using formulas (5) and (6) respectively. Analogously to the uncentered case, GMM estimators are OLS and GLS cross-sectional regressions of expected returns on the

¹⁷If $a = 1 + b'E(f)$, the stochastic discount factor becomes $m = 1 - b'(f - E(f))$. Taking expectations and using the law of iterative expectations, we have $E(m) = 1$.

covariance matrix between excess returns and factors.

2.1.3 GMM inference

From the standard formulae of GMM's distribution theory, the standard errors of the estimators are defined as

$$cov(\widehat{b}_1) = \frac{1}{T}(d'd)^{-1}d'Sd(d'd)^{-1} \quad (7)$$

$$\text{and } cov(\widehat{b}_2) = \frac{1}{T}(d'S^{-1}d)^{-1}, \quad (8)$$

where d will change depending on the selected normalization. The covariance matrix of the pricing errors is given by

$$Tcov(g_T(\widehat{b}_1)) = (I - d(d'd)^{-1}d')S(I - d(d'd)^{-1}d') \quad (9)$$

$$\text{and } Tcov(g_T(\widehat{b}_2)) = S - d(d'S^{-1}d)^{-1}d' \quad (10)$$

Finally, the models are validated using the J-test of over-identification restrictions. For the first and second stage estimators, the χ^2 test can be estimated from the following expressions:

$$g_T(\widehat{b}_1)'cov(g_T(\widehat{b}_1))^{-1}g_T(\widehat{b}_1) \sim \chi^2(N - K) \quad (11)$$

$$Tg_T(\widehat{b}_2)'S^{-1}g_T(\widehat{b}_2) \sim \chi^2(N - K), \quad (12)$$

where N is the number of moments used in the estimation of the parameters, and K is the number of estimated parameters. We use this test as one of the measures of in-sample fit of the linear factor models later on in the analysis.

2.2 Mean-representative portfolio and its estimation by GMM

Applying the Riesz representation theorem, Chamberlein and Rotchschild (1983), prove the existence of a unique portfolio (the mean representing portfolio, p^0) in the set of all possible portfolios formed by the vector of returns, r , which is directly linked to the mean-variance efficient portfolio, r^{MV} : the weights of p^0 are proportional to the weights of r^{MV} , therefore, both have the same Sharpe ratio. Thus, after estimating the weights of p^0 , we can identify the Sharpe ratio of the mean-variance efficient portfolio we are looking for. In practical terms, we estimate the weights of

p^0 by GMM, using a set of moment conditions proposed by Penaranda and Sentana (2009, 2011). These authors show that this method produces consistent estimates of the portfolio weights, and therefore, of the mean-variance frontier. The details of the definition of p^0 and its estimation by GMM are given in this subsection.

Consider a set of N risky assets and one risk-free asset. Define $r = (r_1, \dots, r_N)'$ as the set of excess returns over the risk-free rate of the N risky assets. The payoffs are defined over an underlying probability space Ω . The first uncentered moment is $E(r)$, and the second uncentered moment, that we assume to be finite, is given by $E(rr')$. Let $p = w'r$ be the payoff of a portfolio with fixed weights $w = (w_1, \dots, w_N)'$. The set of all possible portfolios built using r is denoted by P . More formally, P is the linear span of r . The mean value of any portfolio $p \in P$ is given by $E(p) = w'E(r)$, and its cost by $C(p) = w'C(r)$. For the case of excess returns, the cost of the payoffs is zero, $C(r) = 0$, and as a consequence, the cost of the portfolio p is zero as well. In a more general setup, when $C(r) \neq 0$, it is possible to define a cost representing portfolio that, with the mean representing portfolio defined below, characterized the efficient mean-variance frontier.

In order to define what a mean representing portfolio is, we need to introduce additional notation and concepts. The set P is a linear subspace of $L_2(P)$, which is the collection of all random variables with finite variance, defined on the underlying probability space of P . It is known that the set $L_2(P)$ is a Hilbert space under the mean-square inner product: $(p, q) = E(pq)$ for any $p, q \in P$, with the associated norm $\|p\| = \sqrt{E(p^2)}$. Since $E(\cdot)$ is a continuous function in $L_2(P)$, Chamberlain and Rothschild (1983), invoking the Riesz representation theorem, prove that there is a unique portfolio in P , p^0 , representing the mean value of any portfolio in P . Thus, the (uncentered) mean representing portfolio p^0 is such that

$$E(p) = E(p^0 p) \quad \forall p \in P. \quad (13)$$

Under the topology just introduced, it is possible to show that this portfolio is defined as

$$p^0 = E(r')E(rr')^{-1}r = \phi^0 r. \quad (14)$$

From this expression, we observe that the weights of the mean representing portfolio are given by the vector $\phi^0 \equiv E(r')E(rr')^{-1}$. The mean-variance efficient frontier can be characterized either as a function in the mean-variance space, $\{\mu, V(r^{MV}(\mu))\}$, or as a function in the mean-standard deviation space, $\{\mu, \sqrt{V(r^{MV}(\mu))}\}$. The former function will be a parabola tangent to the origin in the mean-variance space, whereas the later will be a straight line in the mean-standard deviation space. A nice property of this methodology is that there exist a one-to-one mapping between p^0

and any portfolio on the mean-variance efficient frontier, r^{MV} . This link is given by

$$r^{MV} = \mu \frac{1}{E(p^0)} p^0 \quad (15)$$

$$\text{and } V[r^{MV}(\mu)] = \left[\frac{1 - E(p^0)}{E(p^0)} \right] \mu^2, \quad (16)$$

where $\mu \in \mathbb{R}$ is the expected return and $V[r^{MV}(\mu)]$ is the variance of the optimal portfolio.¹⁸ It is worth mentioning at this point that in this setup the Sharpe ratio of the portfolios in the mean-variance frontier is the same for any value of μ , equal to the Sharpe ratio of p^0 .¹⁹

From (15), we observe that the weights of the mean-variance portfolio, r^{MV} , are proportional to the weights of the mean representing portfolio, p^0 . Hence, after estimating the weights of p^0 , it is straightforward to estimate the weights of r^{MV} , and then to compute the ex-post realized Sharpe ratio of the portfolio, one period in advance.

Penaranda and Sentana (2009, 2011), propose to use the following system of $N + 1$ moments to estimate the weights of the mean representing portfolio:

$$E \begin{bmatrix} rr' \phi^0 - r \\ x \phi^0 - \mu^0 \end{bmatrix} = 0_{(N+1 \times 1)}. \quad (17)$$

Here, the first N moment equations comes directly from the definition in (13), valid for each of N test assets in the investment menu. The last moment condition identifies the expected return of the mean representing portfolio, μ^0 . Penaranda and Sentana (2011) show, that under some regularity conditions, the GMM estimator of the set of coefficients, $\theta = \{\phi^0, \mu^0\}$, is consistent, and hence, the estimated mean-variance efficient frontier converges in probability to its population value. Despite of this, Penaranda and Sentana (2011) also show, using simulations, that the estimated frontier may suffer of severe problems of sampling uncertainty. In fact, this problem is not specific to this particular methodology, and it has been well studied in the portfolio choice literature. See, for example, Britten-Jones (1999). If our assumption that excess returns are governed by a linear factor structure is valid, then this problem should ameliorate.²⁰

As we mentioned above, we are using predicted excess returns in the second stage of estimation.

¹⁸Taking expectations in (15), we have $E(r^{MV}) = \mu$. Furthermore, the expression of the variance in (16) uses the fact that $E(p^0) = E(p^0 p^0)$ by (13).

¹⁹If $r^{MV} = a p^0$, where a is the constant $\mu/E(p^0)$, then

$$SR(r^{MV}) = \frac{aE(p^0)}{\sqrt{a^2 V(p^0)}} = \frac{E(p^0)}{\sigma(p^0)} = SR(p^0) \text{ for any } \mu \in \mathbb{R}.$$

²⁰Peñaranda and Sentana (2011) propose to incorporate linear factor models in estimation of the mean-variance frontier, expanding (17) with the following additional moment conditions: $E[(1 - bf)x] = 0$, and estimate the full model in one stage. We estimate both sets in different stages.

These returns are computed as follows: replacing equation (2) in (1a) and imposing the normalization $a = 1$, we obtain the fundamental asset pricing equation for the case of excess returns, $0 = E[(1 - b'f)r]$. This expression can be re-written as $0 = E[r - r b'f]$. We call predicted returns to the term $r^e b'f$, i.e., we are using the predicted discounted excess returns. Of course, each linear factor model produces estimates of b , and therefore, estimates of the predicted excess returns.

2.3 Portfolio evaluation and the test of differences of Sharpe ratios.

We evaluate the ex-post realized performance of the portfolios testing whether there are statistically significant differences between their out-of-sample Sharpe ratios. In particular, we perform pairwise tests for all possible combination of models. Previous literature (see for example De Miguel et al., 2009) has relied on the test of differences of Sharpe ratios of Jobson and Korkie (1981), and its small-sample corrected version by Memmel (2003). This test, however, assumes that asset returns are i.i.d. normal. However, it is well documented that asset return distributions are far from being normal, exhibiting also significant levels of autocorrelation. Lo (2002) and Opdyke (2007) propose a test of difference in Sharpe ratios under the less restricted assumption of stationarity of the return distribution. Nevertheless, we discard them since they do not take into account the time-series nature of the excess return data.

We use the robust bootstrapped test of differences in Sharpe ratios proposed by Ledoit and Wolf (2008). This test uses the circular-block bootstrap of Politis and Romanos (1992) to build a two-sided confidence interval for the null hypothesis $H_0 : \Delta = 0$, where $\Delta \equiv SR_1 - SR_2$ is the difference of the Sharpe ratios between any two portfolios. This test is suitable for our exercise as it explicitly accommodates normality deviations and time dependency in excess returns data through the resampling technique.

Before describing the test and its implementation, some notation and definitions are necessary to be introduced.²¹

2.3.1 Preliminaries

Suppose we observe a time-series of length T of pairs of excess returns, $\{r_{it}, r_{nt}\}_{t=1}^T$, for any two assets i and n . The asset returns are characterized by a bivariate, strictly stationarity distribution function with mean $\mu = \begin{pmatrix} \mu_i \\ \mu_n \end{pmatrix}$ and variance-covariance matrix $\Sigma = \begin{pmatrix} \sigma_i^2 & \sigma_{in} \\ \sigma_{in} & \sigma_n^2 \end{pmatrix}$. The second uncentered moments of the returns are denoted by $\gamma_i = E(r_i^2)$ and $\gamma_n = E(r_n^2)$. Let's define the vector $v = (\mu_i, \mu_n, \gamma_i, \gamma_n)'$ and its sample counterpart as $\hat{v} = (\hat{\mu}_i, \hat{\mu}_n, \hat{\gamma}_i, \hat{\gamma}_n)'$. The difference of the Sharpe ratios between these two portfolios is given by $\Delta = f(v)$ with $f(a, b, c, d) = \frac{a}{\sqrt{c-a^2}} - \frac{b}{\sqrt{d-b^2}}$.

²¹We closely follow the notation in Ledoit and Wolf (2008).

In this expression, a and b are the mean values of the returns, and c and d are the second uncentered moments of the returns. Moreover, $f(v)$ has a sample counterpart, which is denoted by $\widehat{\Delta} \equiv f(\widehat{v})$.

It is assumed that \widehat{v} is normally distributed with a mean v and covariance matrix $T^{-1}\Psi$ as follows

$$\sqrt{T}(\widehat{v} - v) \xrightarrow{d} N(0, \Psi), \quad (18)$$

where Ψ is an unknown, symmetric, positive semi-definite matrix.²² Using the delta method, we have that the estimated difference of Sharpe ratios, $\widehat{\Delta}$, is also normally distributed:

$$\sqrt{T}(\widehat{\Delta} - \Delta) \xrightarrow{d} N(0, \nabla' f(v)\Psi \nabla' f(v)), \quad (19)$$

where the gradient vector ∇f is defined as

$$\nabla' f(a, b, c, d) = \left(\frac{c}{(c - a^2)^{1.5}}, -\frac{d}{(d - b^2)^{1.5}}, -\frac{1}{2} \frac{c}{(c - a^2)^{1.5}}, \frac{1}{2} \frac{b}{(d - b^2)^{1.5}} \right). \quad (20)$$

From (19), the estimated standard error for $\widehat{\Delta}$ is given by

$$s(\widehat{\Delta}) = \sqrt{\frac{\nabla' f(\widehat{v})\widehat{\Psi} \nabla' f(\widehat{v})}{T}}, \quad (21)$$

where $\widehat{\Psi}$ is a consistent estimator of Ψ . Lo (2002) uses a HAC estimator of Ψ in his test whereas Ledoit and Wolf (2008) utilize a kernel estimator of Ψ using the bootstrapped data. Romano and Wolf (2006) show that when the sample size is small, the inference based on the kernel/bootstrap estimator is more precise, compared with the inference based on the HAC estimator. In particular, the HAC estimator tends to overreject the true null hypothesis.

2.3.2 The robust bootstrapped test

Ledoit and Wolf (2008) propose to build a two-sided, studentized, bootstrap confidence interval, with nominal level $(1 - \alpha)$ to test the null hypothesis, $H_0 : \Delta = 0$. As usual, if the interval does not contain the zero, the null hypothesis is rejected, otherwise, it is not rejected. Initially, the distribution of the studentized variable, $\frac{\widehat{\Delta} - \Delta}{s(\widehat{\Delta})}$, is approximated by the bootstrapped return data in the following way

$$\mathcal{L} \left(\frac{\widehat{\Delta} - \Delta}{s(\widehat{\Delta})} \right) \approx \mathcal{L} \left(\frac{\widehat{\Delta}^* - \widehat{\Delta}}{s(\widehat{\Delta}^*)} \right), \quad (22)$$

where $\mathcal{L}(x)$ is the distribution function of the random variable x , $\widehat{\Delta}$ is the difference of Sharpe ratios

²²For time-series data, Ledoit and Wolf (2008) point out that this assumption is valid, as long as there exist $4 + \delta$ moments, where δ is some small positive constant, and a mixing condition holds.

estimated from the observed returns, Δ is the true (unobservable) difference of Sharpe ratios, $s(\widehat{\Delta})$ is the standard error of $\widehat{\Delta}$, $\widehat{\Delta}^*$ is the difference of Sharpe ratios computed with the bootstrapped data, and $s(\widehat{\Delta}^*)$ is the standard error of $\widehat{\Delta}^*$. The bootstrap $(1 - \alpha)$ confidence interval for Δ is defined by

$$\widehat{\Delta} \pm z_{|\cdot|, 1-\alpha}^* s(\widehat{\Delta}), \quad (23)$$

where $z_{|\cdot|, 1-\alpha}^*$ is the λ quantile of $\mathcal{L}\left(\frac{\widehat{\Delta}^* - \widehat{\Delta}}{s(\widehat{\Delta}^*)}\right)$. If the distribution of the returns is non-normal and/or exhibits significant levels of autocorrelation, $z_{|\cdot|, 1-\alpha}^*$ will be larger than its normal counterpart, $z_{|\cdot|, 1-\alpha}$. As a consequence, the rejection of the null hypothesis is harder to achieve. In order to build the confidence interval using (23), we need to decide which bootstrap procedure and what estimator of Ψ will be used. Regarding the bootstrap procedure, the test uses the circular block bootstrap procedure of Politis and Romano (1992), where blocks of pairs from the observed data are resampled with replacement. The block size is of fixed length, $b \geq 1$.²³ Resampling blocks, instead of single observations, is required to preserve the properties of the empirical distribution of excess returns. Regarding the estimator of Ψ , Ledoit and Wolf (2008) propose to use the pre-whitened QS kernel estimator of Andrews and Monohan (1992).

The kernel estimator of Ψ is computed as follows. Let l be the integer part of $\frac{T}{b}$ and let $\widehat{v}^* = (\widehat{\mu}_i^*, \widehat{\mu}_n^*, \widehat{\gamma}_i^*, \widehat{\gamma}_n^*)'$ be the estimators of $v = (\mu_i, \mu_n, \gamma_i, \gamma_n)'$ obtained from the bootstrapped data. A consistent estimate of $\widehat{\Psi}^*$ is given by

$$\widehat{\Psi}^* = \frac{1}{l} \sum_{j=1}^l \zeta_j \zeta_j' \quad (24)$$

where $\zeta_j = \frac{1}{\sqrt{b}} \sum_{t=1}^b y_{(j-1)b+t}^*$ for $t = 1, \dots, l$ and $y_t^* = (\widehat{r}_{it}^* - \widehat{\mu}_i^*, \widehat{r}_{nt}^* - \widehat{\mu}_n^*, \widehat{r}_{it}^{*2} - \widehat{\gamma}_i^*, \widehat{r}_{nt}^{*2} - \widehat{\gamma}_n^*)$. The standard error of the estimated difference of Sharpe ratios, $s(\widehat{\Delta}^*)$, is computed replacing the estimator $\widehat{\Psi}^*$ into the (21). Using a similar procedure $s(\widehat{\Delta})$ can be estimated using the observed excess returns instead of the resampled data.

2.3.3 Small Sample Bias

Previous literature has shown that estimated Sharpe ratios are biased when the sample size is small. Opdyke (2007) shows that in this particular case the bias is given by

$$E(\widehat{SR}) = SR \left[1 + \frac{1}{4} \frac{\left(\frac{\mu_4}{\sigma^4} - 1\right)}{T} \right] \quad (25)$$

where $\frac{\mu_4}{\sigma^4}$ is the kurtosis of excess returns. Regardless of the assumption made on the distribution of

²³Even though Ledoit and Wolf propose an iterative procedure to estimate the optimal block length b , we just fix it at a value of 6.

portfolio returns, this result is valid asymptotically. Thus, before to applying the test of differences we correct the Sharpe ratios accordingly.

3 Data

Our dataset consists of monthly value-weighted US portfolio excess returns, over the risk-free rate, from January, 1927 to December, 2011. The data, but the liquidity factor, is collected from Kenneth French’s online data library. More details on how the portfolios are build can be found on his website. The set of test assets evaluated are: 5, 10 and 17 industry portfolios, and a 25 double-sorted size-value portfolio. We deliberately exclude industry portfolios with a higher number of sectors (49, 100 industries for example) because we find in our estimates that the linear factor models analyzed produce poor fits when these test assets are used, therefore, it make no sense to move to a second stage (the estimation of mean-variance efficient portfolios) when the results of the first stage are not satisfactory. Thus, our exercise of estimating mean-variance efficient portfolios is conditional on the fact that the linear factor models under study are able to produce reasonable in-sample fits of the cross-section of expected returns in the first stage. This point is relevant, since Lewellen et al. (2010) make a suggestion in the opposite direction: to consider larger set of test assets (for example, 30 industries plus 25 size-value portfolios) in the validation of asset pricing models. However, our main interest is not to find the best asset pricing model in-sample but to compare and evaluate three well used linear factor asset pricing models, using an out-of-sample financial metric.

The set of factors considered in the analysis includes the excess return of the market portfolio over the risk-free rate (MKT), the size portfolio SMB (small minus big), the value portfolio HML (high minus low), the momentum portfolio (MOM) and the liquidity factor (LIQ) of Pastor and Stambaugh (2003). The first four factors have been extensively used in the empirical finance literature, therefore, we refer the reader to Fama and French (1993) and Carhart (1997) for details of how they are built. Regarding the liquidity factor, it corresponds to the value-weighted return on the 10-1 portfolio from a sort on historical liquidity betas.²⁴ The data is available in Lubos Pastor’s webpage from January, 1968 to December, 2010. Since this time-series is shorter than the others, in those cases where the liquidity factor is included in the analysis, we estimate the liquidity factor model and its competitors using this small sample. Thus, we reported results for the full sample (1927:01-2011:12) containing 1,020 observations, and for the short sample (1968:01-2010:12) containing 516 observations.

We study four linear factor models: the CAPM ($K = 1$), the Fama-French 3 factor model ($K = 3$), the Carhart model ($K = 4$ or $4M$) and a 4 factor model, where the liquidity factor is added to the 3 Fama-French factor ($K = 4L$).

²⁴In Pastor and Stambaugh (2003), the liquidity factor was built sorting the portfolios on predicted betas instead of historical liquidity betas. However, the use the last series as it is the one available in Lubos Pastor’s webpage.

4 Empirical Results

In this section, we describe and analyse our empirical results. Our empirical strategy relies on a 5-year moving rolling window setup.²⁵ Our first window starts in January, 1927 and ends in December, 1931. For this window we estimate the set of linear asset pricing models under study, using the formulae in subsection 2.1, and estimate the mean-variance efficient portfolio weights using the procedure described in subsection 2.2. Then, using the excess returns observed in January, 1932, we compute the realized ex-post portfolio return one period in advance, and we save the result. Later, we move the estimation window one period forward, covering now February, 1927 to January, 1932. We re-estimate the models, the portfolio weights and compute the realized ex-post portfolio return in February, 1932. We repeat this process in the 960 windows available in our dataset. The time-series of realized ex-post portfolio returns $\{r_w^p\}_{w=1}^{960}$ is used to compute the out-of-sample Sharpe ratio as $\frac{\mu(r_w^p)}{\sigma(r_w^p)}$. Then, we apply the correction for small-sample bias using (25). Each model generates a Sharpe ratio that we use to perform the test of difference in Sharpe ratios described in subsection 2.3.

4.1 Basic Results

In tables (1-8) we report our basic results, where portfolio weights are unrestricted in the estimation process. Each table reports in-sample measures of fit (GLS R^2 , J Test (p-val), RMSE and MAE), the out-of-sample Sharpe ratio, and the p-value of the test of difference in Sharpe ratios for the set of linear factor models under study, for a particular set of test assets. Each linear factor model is estimated using first and second stage GMM estimators, and uncentered and centered SDF models. Tables (1-4) report the results for the full-sample, whereas tables (5-8) report the results for the short-sample.²⁶

The GLS R^2 is the Generalised Least Squared cross-sectional R^2 of linear factor models. Recently, Lewellen et al. (2010) have suggested to report this measure, instead of the standard OLS R^2 for example, to evaluate the validity of alternative asset pricing models. The higher the GLS R^2 the better is the in-sample fit of the models, which in this particular case, can be interpreted as the maximum Sharpe ratio obtainable from the set of test assets. The J-test (pval) is the p-value of the chi-squared J-test of overidentification restrictions in the GMM estimation, see (11) and (12). The null hypothesis of the test is that the moment conditions included in the GMM estimation are valid. We also report the in-sample root mean square error (RMSE) and the mean absolute error (MAE). As a measure of performance of the portfolios, we report the ex-post realized out-of-sample Sharpe ratio (OOS-SR), which is widely utilized for performance evaluation purposes.

²⁵This approach has been followed by De Miguel et al. (2009) among others. As a robustness test, we also consider a 10 year window. Those results are not reported in this version of the paper though.

²⁶1968:01-2010:12.

Regarding the test of difference in Sharpe ratios, for tables (1-4), in the last two columns we report the p-values of the tests. In particular, we report a lower triangular matrix where the elements in the first column of the matrix are, in descending order, the p-values of the tests comparing the Sharpe ratios of the CAPM with the FF3 model and the Carhart model, respectively. The second column has a unique element: the p-value of the test comparing the Sharpe ratio of the FF3 model and the Carhart model. In tables (5-8) the p-value matrix contains an additional column because we are comparing four models instead of three. Again, the values in the first column are the p-values of the test comparing the Sharpe ratios of the CAPM with the other three models, the second column reports the p-values of the test comparing the FF3 model with the two 4 factor models, and finally, the last column reports the p-value of the test comparing the two 4 factor models. The null hypothesis of the test is $H_0 : SR_1 - SR_2 = 0$ versus $H_1 : SR_1 - SR_2 \neq 0$ for any two models and portfolios.

At the bottom of each table we report the out-of-sample Sharpe ratio for 3 benchmark models: the $\frac{1}{N}$ strategy of De Miguel et al. (2009), the mean-variance tangency portfolio (MV) and the global minimum variance portfolio (MinV). The last two models are computed using historical returns.²⁷

4.1.1 Full-Sample

In table 1, we report the results when the test assets are the 5 industries portfolios. The GLS R^2 , which is the same across GMM stages and normalizations, show that the FF3 has a superior in-sample cross-sectional fit to that of the CAPM. The GLS R^2 for the FF3 model is 78 per cent versus 29 per cent for the CAPM. An awkward results is obtained for the Carhart model. The GLS R^2 is -71 per cent. As it is pointed out by Lewellen et al. (2010), the GLS R^2 not need to be limited to the (0,1) interval, therefore, the negative sign is possible, however, the magnitude of the estimate is surprising. Thus, according to this statistic, when the Carhart model is evaluated using as test asset 5 industries portfolios, it does a very poor job fitting the (weighted) expected returns in the cross-section. For the other test assets considered, this is not the case indeed. The J-test clearly shows that the three models are valid: the p-values range between 0.40 and 0.59 across the models. Looking at the RMSE and the MAE, we observe that the Carhart model yields lower pricing errors than the FF3 model, which in turn, produces lower pricing errors than the CAPM. Their RMSE are 0.26, 0.14 and 0.07, respectively, and their MAE are 0.22, 0.11 and 0.06. Looking at the out-of-sample Sharpe ratios associated to the models, for example in the first block, the out-of-sample Sharpe ratios are -0.028, -0.007 and 0.007 for the CAPM, FF3 and the Carhart model, respectively. In the other three blocks we observe a similar pattern, with the CAPM having a lower out-of-sample Sharpe ratio than the FF3 and the Carhart model. In any of the cases,

²⁷The weights of the mean-variance tangency portfolio are given by $w = \frac{\widehat{\Sigma}^{-1}\widehat{\mu}}{i'\widehat{\Sigma}^{-1}\widehat{\mu}}$ and the weights of the global minimum variance portfolio are given by $w = \frac{\widehat{\Sigma}^{-1}i}{i'\widehat{\Sigma}^{-1}i}$, where $\widehat{\Sigma}$ is the $(N \times N)$ covariance matrix of excess returns, $\widehat{\mu}$ is the $(N \times 1)$ vector of mean excess returns and i is a $(N \times 1)$ vector of ones.

however, this difference is statistically not significant according to the Ledoit and Wolf test. For example, the p-values of the test between the CAPM and the Carhart model are 0.45, 0.70 and 0.20 and 0.37 across GMM estimators and normalizations. Regarding the benchmarks, in general, the out-of-sample Sharpe ratio is smaller than both the $1/N$ and global minimum variance portfolios, but higher than mean-variance tangent portfolio computed with historical data.

In table 2, we report the results for 10 industry test assets. GLS R^2 are now in line with previous empirical evidence: the 4 factor model has a better fit than the FF3 model and the CAPM, whereas the FF3 model is superior to the CAPM. The GLS R^2 are 0.14, 0.35 and 0.52, respectively. The three models are validated by the J-test: the p-values of the test are 0.36, 0.45 and 0.61, respectively in the second block for example. Looking at RMSE and MAE, we observe that across specifications, the Carhart model produces lower pricing errors when compared with the other two models. For example, in the second block, the RMSE are 0.36, 0.25 and 0.19 and the MAE are 0.29, 0.21 and 0.16 for the CAPM, FF3 model and Carhart model respectively. Looking at the out-of-sample Sharpe ratios of the models, we observe the same path as in table 1, but now even clearer since the magnitude of the Sharpe ratios is bigger, and more importantly, some differences are statistically significant. For example, in the second block, the out-of-sample Sharpe ratios are -0.006, -0.016 and 0.04 for the CAPM, FF3 model and Carhart model, respectively. In this case, the Sharpe ratio of the Carhart model is statistically higher than the one of the other two models (p-values of 0.05 and 0.02, respectively). In the fourth block, we also observe that the Sharpe ratio of the Carhart model is statistically higher than that of the CAPM (p-value of 0.02). In this case, the difference between the Sharpe ratios is 0.09 (0.033 - (-0.056)). In comparison with the benchmarks models, the Carhart model is well below the $1/N$ and the global mean variance portfolios, but still above the historical mean-variance portfolio.

For the cases of 17 industries portfolios, in table 3, and 25 size-value portfolios, in table 4, we observe a similar pattern as before: models with better performance in-sample, i.e. producing lower pricing errors, are those with higher performance out-of-sample. These results hold regardless of the selection of either the GMM estimator or the normalization considered. The statistical significance of the test of difference is mixed. For the case of 17 industry portfolios, we do not find any statistically significance different, but we do for the case of 25 portfolios. In fact, the results for the 25 test assets are stronger than the previous ones, since now the differences (in the first and third blocks) are significant at the 99 per cent confidence level. For example, in the first block, the out-of-sample Sharpe ratios are -0.005, 0.083 and 0.144 for the CAPM, the FF3 model and the Carhart model, respectively. The test of differences in Sharpe ratios indicates that the Sharpe ratio of the Carhart model is larger than that of the FF3 model at 99 per cent confidence level, and taht of the CAPM, at 90 per cent confidence level. The best performing model in the case of 25 test assets (uncentered, first-stage GMM) produces an out-of-sample Sharpe ratio of 0.144. This number is competitive when compared with the performance of the $1/N$ investing strategy, indicating that, to some extent, linear factor models, in particular the Carhart model, are useful at

hands of investors looking at a profitable investment strategy.

4.1.2 Short-Sample

As we mentioned before, we include in the analysis a four factor liquidity model, where the FF3 model is expanded by a liquidity factor. Pastor and Stambaugh (2003) build this liquidity factor and study whether this factor is priced in a sample of US stocks. They find that the liquidity factor is priced and that it accounts for a dimension of risk, which is not captured by the other factors considered. We estimate this model using a shorter sample because the time series of the liquidity factor are only available since January, 1968. Thus, our short sample covers January 1968 to December 2011. Tables (5-8) report our estimates of the four linear factor models under study.

First, we observe that there exist an empirical link between the fit of the models in-sample and the financial performance of the associated portfolios out-of-sample: those models producing lower pricing errors in-sample generate portfolios with the higher out-of-sample Sharpe ratios. In particular, the CAPM even though is not rejected by the J-test, produces the lower GLS R^2 and the higher RMSE and MAE among the four models. For example, in the first block in table 6, the CAMP produces a GLS R^2 of 0.15, whereas the four factor liquidity model produces 0.55. The RMSE of the CAPM is 0.35, whereas the liquidity model has a much lower RMSE of 0.15. In terms of its out-sample performance, we find that in most of the cases the CAPM is outperformed by multifactor models. Again, looking at the first block of table 6, the CAPM generates an out-of-sample Sharpe ratio of 0.008 whereas the liquidity factor model generates a Sharpe ratio of 0.099. In the second block of table 6, the out-of-sample Sharpe ratios are 0.028 for the CAPM and 0.128 for the liquidity factor model. This difference of 10 percent is statistically significant at 95 percent.

Second, comparing the two four factor models, we observe that the Carhart model has slightly better in-sample fit than the liquidity factor model, but when we look at their out-of-sample Sharpe ratios, we find that the liquidity factor model produces superior performance. For example, in the fourth block in table 7, the Carhart model has a GLS R^2 of 0.37 and the liquidity factor model has a value of 0.33. The out-of-sample Sharpe ratios are 0.005 and 0.079, respectively. It is worth noting that two exceptions occur when the 25 size-value portfolios are used as test assets. Here, the Carhart model outperforms the liquidity factor model in terms of the out-of-sample Sharpe ratio. See for example the first block of table 8, where the out-of-sample Sharpe ratios are 0.167 and 0.118, respectively.

Third, in general terms, both four factor models exhibit better in-sample adjustment and superior out-of-sample performance than both the CAPM and the FF3 model. For example, in the first block of table 6, the GLS R^2 are 0.15 and 0.36 for the CAPM and the FF3 model whereas the four factor models yield values of 0.57 and 0.55, respectively. In the same block, the out-of-sample Sharpe ratios of the CAPM and the FF3 models are lower than the ones of the two four factor

models (0.008 and 0.031 versus 0.057 and 0.099).

Fourth, in this sub-sample, we observe more often that the differences in Sharpe ratios are statistically significant. In each of the 4 test assets considered, we find that at least in one block a four factor model has an out-of-sample Sharpe ratio statistically higher than the CAPM. In the case of the 25 size-value portfolios, this difference is statistically significant in each of the four blocks. In most of the cases it is the liquidity factor model the one outperforming the CAPM. Cases where the Carhart model or both four-factor models simultaneously outperform the CAPM are less frequent.

Finally, it is interesting to point out that in some cases the liquidity factor model is able to produce out-of-sample Sharpe ratios a bit above one of the benchmark $1/N$ strategies. For example, for the case of 10 industries portfolios, in the second block the out-of-sample Sharpe ratio is 0.128 versus 0.115 of the $1/N$ strategy. In the case of 25 size-value portfolios, in the last two blocks the out-of-sample Sharpe ratios of the liquidity factor model are above the $1/N$ strategy (0.136 and 0.162 for the liquidity factor model versus 0.13 of the $1/N$ strategy). These results indicate that, in some cases, the use of linear factor models in the building of investment portfolios may help to achieve competitive rewards in terms of performance.

4.2 Short-Sales Constraints

Imposing short sales restrictions is common practice among practitioners and researchers dealing with portfolio allocation problems. Sometimes the restriction takes the form of an upper bound, in others, short sales are completely ruled out. At least two reasons explain why short sales constraints are so popular among practitioners and researchers: first, it avoids the occurrence of extreme positions in the portfolio, which can be unfeasible to implement, due either legal restrictions or liquidity issues ; and second, previous evidence has shown that the performance of the restricted portfolios is superior to the one of the unrestricted ones. Jagannathan and Ma (2003) rationalize this phenomenon arguing that there exist a trade-off between specification errors and estimation errors at solving the portfolio problem: short sales constraints would reduce the estimation errors but increasing the specification errors. Considering these elements we explore the effect of imposing short-sales constraints in our portfolio exercise. In tables (9-12), we report the results for the full sample, and in tables (13-16), the results for the short sample.

4.2.1 Full Sample

Our first observation is that our results become stronger after imposing short sales constraints. We clearly observe than the reported link between in-sample and out-of-sample performance of the models do exist, and it is highly significant across test assets, linear models, GMM specifications and normalizations. It is clear that multifactor models are able to better explain the cross-section of

expected returns than the CAPM, a well-known result in the literature, but these results document that multifactor models yield a superior out-of-sample Sharpe ratio as well.

In the case of 5 industries test assets in table 9, the GLS R^2 of the FF3 model is 0.78, whereas the one for the CAPM is 0.29. As we mentioned above, the puzzling result here is that the GLS R^2 of the Carhart model is -0.71. Based on this evidence, the Carhart model would do a poor job fitting the cross-section of expected returns, however, we observe the opposite in all the other cases under consideration. In particular, we find that the GLS R^2 of the Carhart model is always higher than the one of the FF3 model and the CAPM. The other measures of in-sample fit are consistent with this pattern: the RMSE and the MAE systematically shows that multifactor models yield lower pricing errors than the CAPM. For example, in the first block of table 9, the RMSE are 0.26, 0.14 and 0.07 and the MAE are 0.22, 0.11 and 0.06 for the three models under study, respectively. The out-of-sample Sharpe ratio of the multifactor models is always statistically larger than the performance of the CAPM. In the first block, the out-of-sample Sharpe ratios are 0.030, 0.139 and 0.127. The differences of Sharpe ratios are sizeable, reaching values of 10 percent. It is interesting to note that, for these test assets, we do not only observe that multifactor models exhibit superior performance to the CAPM but that the Carhart model is superior to the FF3 model, which at the same time, is superior to the CAPM. These results resemble the standard in-sample results in the asset pricing literature. See Carhart (1997) for example. The same pattern is observed in the third and fourth blocks. The test of difference of Sharpe ratios are highly significant across blocks. The Carhart model produces higher out-of-sample Sharpe ratios than the CAPM, regardless of the GMM estimator and normalization used. These results indicates that imposing short sales constraint improve significantly the performance of the portfolios. In fact, when we compare these results with those in table 1 (with unrestricted weights), we observe similar results but less significant.

In table 10, we report the results for 10 industries portfolios. These results are even stronger than the ones for 5 industries. Here, not only the Carhart model exhibits a higher out-of-sample Sharpe ratio than the CAPM, the FF3 model also outperforms the CAPM. The Ledoit and Wolf test shows that the difference of Sharpe ratios between multifactor models and the CAPM are all statistically significant at 95 per cent. The reported differences are larger than previous results, averaging 12 per cent across blocks. A remarkable result here is that the portfolio associated to the Carhart model outperforms both the $1/N$ investment strategy and the historical mean-variance portfolio in three out of four blocks, and it outperforms the global minimum-variance once.²⁸ This evidence reinforces the idea that linear factor models may provide useful information to investors dealing with asset allocation problems. In fact, it shows that when additional economic structure is embedded in the portfolio analysis, the performance of this portfolio is superior to several competitive investment strategies.

²⁸In order to make a fair comparison with our models, we also restrict short sales in the benchmark models when required.

Table 11 reports the results for 17 industries portfolios. Similarly to the case of 5 industries portfolios, after imposing the short sales restrictions, we obtain statistically significant differences between the out-of-sample Sharpe ratio of multifactor models and the CAPM. These results would indicate that short sales reduce the estimation error, more than compensating any increment in the specification error of the portfolio optimization process. These results are quite similar to the case of 10 industries in terms of the in-sample measures of fit. The GLS R^2 are 0.08, 0.21 and 0.34 for the CAPM, the FF3 model and the Carhart model, respectively. The RMSE and the MAE also show that the Carhart model produces lower pricing errors than the two other models. Even though, the differences in Sharpe ratios between multifactor models reach eleven per cent, we observe that in three out of four blocks it is in fact the FF3 model is the one with a slightly higher out-of-sample Sharpe ratio than the Carhart model.

Finally, in table 12 we report the results for the 25 size-value portfolios. As compared with the results in table 4 (with unrestricted weights), just in two blocks, the first and the third one, the test of difference of Sharpe ratios are significant. Nevertheless, in general, the results are still consistent with the previous evidence: multifactor models are estimated with less error in-sample and they have higher out-of-sample Sharpe ratio than the CAPM. We observe that the ability of the linear models in explaining the cross-section of the expected returns declines, since the GLS R^2 are now only 0.04, 0.12 and 0.16. Also, it is worth noting that the J-test rejects the CAMP model at 95 percent confidence level. The RMSE and the MAE show that multifactor models produces lower pricing errors than the CAPM. Regarding the out-of-sample Sharpe ratio we observe that even though the difference between multifactor models and the CAPM do exist and it is statistically significant in two blocks according to the Ledoit and Wolf test, they are weaker now, reaching values of 9 per cent in the first block and 5 per cent in the third block approximately.

4.2.2 Short Sample

As before, we add to the analysis the liquidity factor model. Its inclusion forces us to use a shorter sample, due to data availability. Again, in order to fairly compare the models we re-estimate the other factor models using the same sample. In general, the results in this subsection are quite similar to those reported in subsection 4.2.1 in terms of following the trends we have been documenting so far. Multifactor models tend to better explain the cross-section of expected returns than the CAPM and also, tend to exhibit a stronger out-of-sample performance one period in advance. The statistical significance of the Ledoit and Wolf test decays as compared with the estimation of the same models using the full sample reported in 4.1.2. Now, for the case of 5 industries portfolios in table 13, most of the times the test of difference is not significant. Just for two cases, the liquidity factor model has an statistical superior out-of-sample Sharpe ratio than the FF3 model.

In table 14, the difference of Sharpe ratios are statistically significant in 3 blocks. For example, in the third block the liquidity factor model is superior to the CAPM at the 90 per cent confidence

level. In this case, the out-of-sample Sharpe ratios are 0.152 for the liquidity factor model and 0.054 for the CAPM. In the fourth block, it is the Carhart model that has a superior out-of-sample Sharpe ratio at the 90 percent confidence level. In this case, the out-of-sample Sharpe ratios are 0.136 for the Carhart model and 0.05 for the CAPM.

In table 15, when 17 industry portfolios are used as test assets, in the first two blocks four factor models outperform the CAPM, in this case the momentum factor being more important than the liquidity factor. In the third block, we do not find any statistical significant difference between Sharpe ratios, however, multifactor models still exhibit a superior in-sample fit and out-of-sample portfolio performance. In the fourth block, only the FF3 model produces a higher out-of-sample Sharpe ratio than the CAPM (0.068 versus 0.037, respectively). In comparison with the benchmark models, we observe that a multifactor model, in this case the Carhart model, outperforms the three benchmark portfolios. In particular, in the first two blocks, the out-of-sample Sharpe ratio of the Carhart model are 0.138 and 0.123 respectively whereas the highest out-of-sample Sharpe ratio reached by a benchmark models is 0.118 (the historical mean-variance portfolio). The $1/N$ investment strategy generates an even lower out-of-sample Sharpe ratio of 0.109.

In table 16 we report the results for the 25 size-value portfolios. We find that only in two blocks the Ledoit and Wolf test finds statistically significant differences of Sharpe ratios between the four factor models and the CAPM. In the first block, the out-of-sample Sharpe ratios are 0.009 and 0.122 for the CAPM and the Carhart model, respectively. In the third block, these numbers are 0.033 and 0.038. However, in this case, the liquidity factor model has a out-of-sample Sharpe ratio of 0.101. For this set of test assets, our portfolios underperform the benchmark portfolios.

Finally, when we compare the performance of the two four factor models, we find that the liquidity factor model has a worse in-sample fit than the Carhart model. This is the same result we obtained with the full sample estimates. The results change when we compare the out-of-sample performance of the models. In this case, we do not observe that a model dominates the other as it was the case with the full sample estimates. Here, we find that in half of the blocks the Carhart model yields higher out-of-sample Sharpe ratios, and in the other half, it is the liquidity factor model that outperforms the Carhart model.

4.3 Summary

To sum up, the main empirical facts reported in this section are the following. First, we document that there is a link connecting the in-sample performance of linear asset pricing models and their out-of-sample financial performance. In fact, we observe that models having lower in-sample pricing errors, have, at the same time, higher out-of-sample Sharpe ratios. Second, we document that linear multifactor asset pricing models have a consistently higher performance than the CAPM, when they are evaluated out-of-sample and under a financial performance metric. Third, when an appropriate

bootstrap test is used to statistically compare the performance of the models, we find that in many cases these differences are statistically significant. Fourth, we observe that the 4-factor model, that include the liquidity factor of Pastor and Stambaugh (2003), shows a worst in-sample fit than the Carhart model in our sample period, however, it is able to beat, in some cases, out-of-sample, all the other models. Fifth, we verify that short sales constrains are important reducing sampling variability in portfolio estimates, when the full sample is used. This is not necessarily the case when we use the shorter sample, covering the second half of the full sample. This indicates that most of the uncertainty observed in the estimation of the portfolio weights is explained by the observed returns in the first half of our sample.

5 Conclusions

This paper evaluates linear stochastic discount factor models based on the out-of-sample realized performance of mean-variance efficient portfolios backed by the models. In particular, we test whether the well-documented superior ability of multifactor models, over the CAPM, of fitting the cross-section of expected returns in sample remains or not, when an out-of-sample portfolio metric is considered instead. Moreover, we test whether there is an empirical link between the in-sample statistical performance of the factor models and their out-of-sample portfolio performance.

Our methodology consists of two stages. In the first stage, the linear factor models are estimated under their SDF representation by GMM, as described by Cochrane (2005). Then, in the second stage, we use the predicted excess returns of models to estimate a mean-variance efficient portfolio using the concept of mean-representing portfolio introduced by Chamberlain and Rothschild (1983). As it is shown by Penaranda and Sentana (2009, 2011), the mean-representing portfolio delivers a set of moments, which allow to estimate by GMM a (arbitrage) mean-variance frontier consistently. Finally, we compute the realized out-of-sample Sharpe ratio, one-period in advance, for each model. The performance of the linear models is evaluated with the circular-block bootstrapped test of difference of Sharpe ratios of Ledoit and Wolf (2008).

Using a sample of monthly US portfolio returns data spanning the period from 1927 to 2011, we provide evidence that multifactor linear models have better empirical properties than the CAPM, not only when the cross-section of expected returns is evaluated in-sample, but also when a portfolio metric is used out-of-sample. Besides, we document that there exist an empirical link between the in-sample statistical performance and the out-of-sample portfolio performance of linear factor asset pricing models: those models exhibiting the best in-sample performance are also the models with the best out-of-sample performance. This result is consistent with the idea that asset pricing models provide useful information to an investor solving a mean-variance portfolio problem. A model with lower pricing errors is the best approximation to the data generating process of the returns that an investor could have at hand, therefore, he/she might incorporate this information into the

optimization process in order to reduce estimation errors.

We find that multifactor models outperform the CAPM yielding out-of-sample Sharpe ratios by as much as 10 per cent. These results are robust for considering alternative assets in the investment menu, first and second stage GMM estimators, center and uncentered SDF specifications, a sub-sample analysis, and alternative window lengths in the rolling estimation setup. We also study the performance of a 4-factor model, where the Fama-French 3-factor model is expanded with the liquidity factor of Pastor and Stambaugh (2003). We find that in-sample, this model is beaten by the Carhart model, however, in some cases, it is able to outperform, out-of-sample, its competitor models. Finally, we study the role of imposing short sales constraint on the portfolio optimization problem, as suggested by the work of Jagannathan and Ma (2003). Our results show that short sales constraint may help to increase the out-of-sample Sharpe ratio of the portfolios when the full sample is considered, however, this is not the case when the second half of the sample (1968-2011) is used instead.

The empirical evidence in this paper adds to two strands of the literature. First, it adds to the literature evaluating the empirical performance of linear asset pricing models. As we mentioned above, (i) we provide evidence that multifactor models exhibit better empirical properties than the CAPM, not only when the cross-section of expected returns is evaluated in-sample, but also when a financial metric is used out-of-sample; (ii) we document that there is a empirical link between the in-sample statistical performance and the out-of-sample financial performance of linear factor models: models producing lower pricing errors in-sample are also those exhibiting higher out-of-sample Sharpe ratios. This empirical fact has not been documented in the literature so far. Second, we add to the small body of literature aiming to evaluate the use of asset pricing models in the asset allocation problem. Our contribution here is: (i) to assemble a methodology combining the SDF form of the linear asset pricing model and the consistent estimation of mean-variance efficient portfolios by GMM, (ii) to report evidence that, in some cases, portfolios associated with multifactor models may deliver higher out-of-sample Sharpe ratios than the $1/N$ investment strategy of De Miguel et al. (2009), indicating that linear factor models may deliver economic value to an investor who is maximizing his/her wealth.

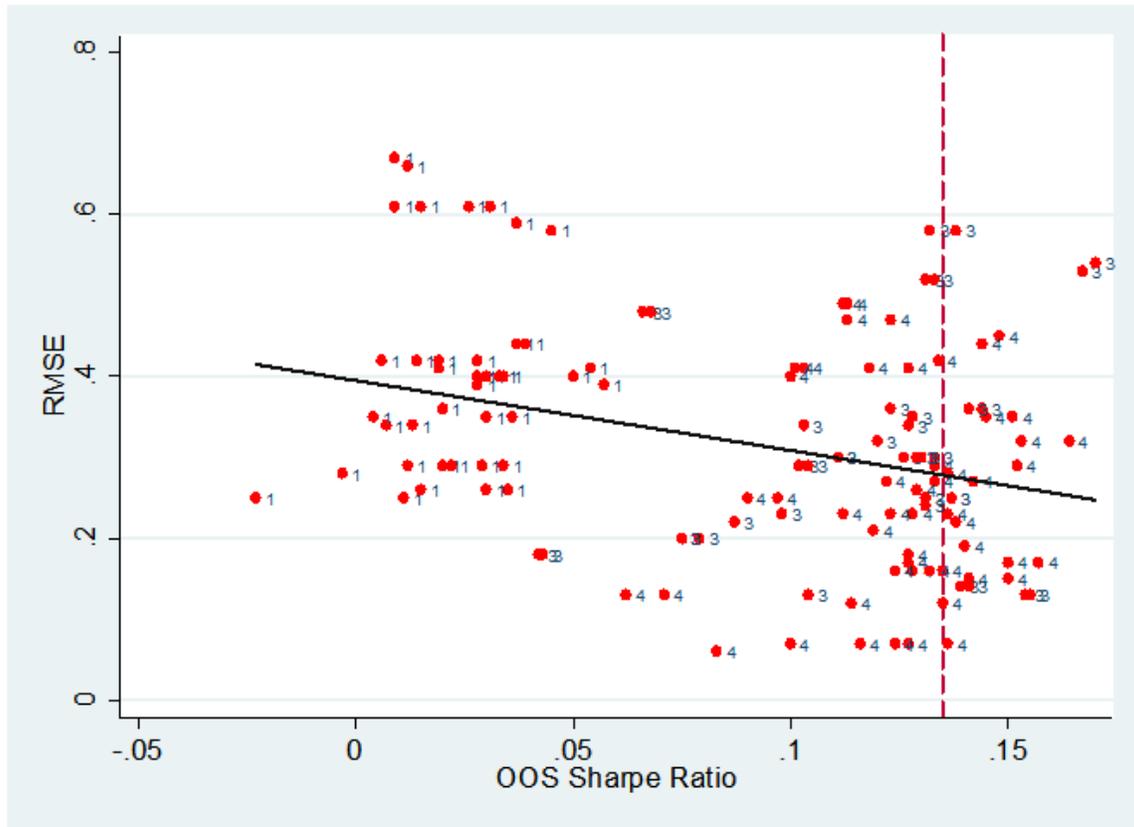
References

- [1] ANDREWS, D., AND MONAHAN, J. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica: Journal of the Econometric Society* (1992), 953–966.
- [2] BEHR, P., GUETTLER, A., AND TRUEBENBACH, F. Using industry momentum to improve portfolio performance. *Journal of Banking & Finance* (2012).
- [3] BRITTEN-JONES, M. The sampling error in estimates of mean-variance efficient portfolio weights. *The Journal of Finance* 54, 2 (1999), 655–671.
- [4] CARHART, M. On persistence in mutual fund performance. *Journal of finance* (1997), 57–82.
- [5] CHAMBERLAIN, G., AND ROTHSCILD, M. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica: Journal of the Econometric Society* (1983), 1281–1304.
- [6] CHEVRIER, T., AND MCCULLOCH, R. *Using economic theory to build optimal portfolios*. PhD thesis, 2008.
- [7] COCHRANE, J. *Asset pricing*, vol. 30. Princeton university press, 2005.
- [8] CONNOR, G., AND KORAJCZYK, R. Factor models in portfolio and asset pricing theory. *Handbook of Portfolio Construction* (2010), 401–418.
- [9] DANIEL, K., AND TITMAN, S. Evidence on the characteristics of cross sectional variation in stock returns. *The Journal of Finance* 52, 1 (1997).
- [10] DEMIGUEL, V., GARLAPPI, L., AND UPPAL, R. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies* 22, 5 (2009), 1915–1953.
- [11] FAMA, E., AND FRENCH, K. Common risk factors in the returns on stocks and bonds* 1. *Journal of financial economics* 33, 1 (1993), 3–56.
- [12] HANSEN, L., AND RICHARD, S. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica: Journal of the Econometric Society* (1987), 587–613.
- [13] JAGANNATHAN, R., AND MA, T. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance* 58, 4 (2003), 1651–1684.
- [14] JAGANNATHAN, R., AND WANG, Z. Empirical evaluation of asset-pricing models: A comparison of the sdf and beta methods. *The Journal of Finance* 57, 5 (2002).

- [15] JOBSON, J., AND KORKIE, B. Performance hypothesis testing with the sharpe and treynor measures. *Journal of finance* (1981), 889–908.
- [16] JORION, P. Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis* 21, 03 (1986), 279–292.
- [17] KAN, R., AND ZHOU, G. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis* 42, 3 (2007), 621.
- [18] KIRBY, C., AND OSTDIEK, B. It s all in the timing: Simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis* (2012).
- [19] LEDOIT, O., AND WOLF, M. Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance* 15, 5 (2008), 850–859.
- [20] LEWELLEN, J., NAGEL, S., AND SHANKEN, J. A skeptical appraisal of asset pricing tests. *Journal of Financial Economics* 96, 2 (2010), 175–194.
- [21] LINTNER, J. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The review of economics and statistics* 47, 1 (1965), 13–37.
- [22] LO, A. The statistics of sharpe ratios. *Financial Analysts Journal* (2002), 36–52.
- [23] LOZANO, M., AND RUBIO, G. Evaluating alternative methods for testing asset pricing models with historical data. *Journal of Empirical Finance* 18, 1 (2011), 136–146.
- [24] MACKINLAY, A., AND PÁSTOR, L. Asset pricing models: Implications for expected returns and portfolio selection. *The Review of Financial Studies* 13, 4 (2000), 883–916.
- [25] MARKOWITZ, H. Portfolio selection. *The journal of finance* 7, 1 (1952), 77–91.
- [26] MEMMEL, C. Performance hypothesis testing with the sharpe ratio.
- [27] OPDYKE, J. Comparing sharpe ratios: So where are the p-values? *Journal of Asset Management* 8, 5 (2007), 308–336.
- [28] PÁSTOR, L. Portfolio selection and asset pricing models. *The Journal of Finance* 55, 1 (2000), 179–223.
- [29] PASTOR, L., AND STAMBAUGH, R. Comparing asset pricing models: an investment perspective. *Journal of Financial Economics* 56, 3 (2000), 335–381.
- [30] PASTOR, U., AND STAMBAUGH, R. Liquidity risk and expected stock returns. *Journal of Political Economy* 111, 3 (2003), 642–685.
- [31] PEÑARANDA, F., AND SENTANA, E. Spanning tests in return and stochastic discount factor mean-variance frontiers: a unifying approach.

- [32] PEÑARANDA, F., AND SENTANA, E. Inferences about portfolio and stochastic discount factor mean variance frontiers.
- [33] POLITIS, D., AND ROMANO, J. A circular block-resampling procedure for stationary data. *Exploring the limits of bootstrap* (1992), 263–270.
- [34] ROMANO, J., AND WOLF, M. Improved nonparametric confidence intervals in time series regressions. *Journal of Nonparametric Statistics* 18, 2 (2006), 199–214.
- [35] SCHWERT, G. Anomalies and market efficiency. *Handbook of the Economics of Finance* 1 (2003), 939–974.
- [36] SHANKEN, J., AND ZHOU, G. Estimating and testing beta pricing models: Alternative methods and their performance in simulations. *Journal of Financial Economics* 84, 1 (2007), 40–86.
- [37] SHARPE, W. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of finance* 19, 3 (1964), 425–442.
- [38] TU, J., AND ZHOU, G. Markowitz meets talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics* 99, 1 (2011), 204–215.

Figure 1: Scatter plot of RMSE vs out-of-sample Sharpe ratios.



$$Y = \underset{(0.02)}{0.39} - \underset{(0.24)}{0.86} X \quad (N = 132, R^2 = 0.10)$$

Each dot represents a particular model and the number attached to it indicates the number of factors included in that model. The dashed vertical line indicates the average performance achieved by the $1/N$ strategy in our sample, across models and specifications.

Table 1: Linear Asset Pricing Model Estimation and MV Portfolio Performance

(Full Sample, Test Assets: 5 Industries)

First Stage Estimators						
Uncentered SDF ($a = 1$)						
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test
1	0.29	0.40	0.26	0.22	-0.028	
3	0.78	0.46	0.14	0.11	-0.007	0.62
4	-0.71	0.59	0.07	0.06	0.007	0.45 0.67
Centered SDF ($a = 1 - \lambda'E(f)$)						
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test
1	0.29	0.40	0.26	0.22	0.002	
3	0.78	0.46	0.14	0.11	0.008	0.83
4	-0.71	0.59	0.08	0.06	0.044	0.70 0.84
Second Stage Estimators						
Uncentered SDF ($a = 1$)						
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test
1	0.29	0.40	0.29	0.24	-0.019	
3	0.78	0.46	0.20	0.17	-0.009	0.90
4	-0.71	0.59	0.15	0.13	-0.002	0.20 0.24
Centered SDF ($a = 1 - \lambda'E(f)$)						
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test
1	0.29	0.40	0.29	0.24	-0.039	
3	0.78	0.46	0.20	0.17	0.000	0.35
4	-0.71	0.59	0.16	0.14	-0.004	0.37 0.90
Benchmarks						
OOS SR	1/N	MV	MinV			
	0.143	-0.025	0.149			

The table reports in-sample measures of goodness of fit for a set of linear asset pricing models ($K = 1$ is the CAPM, $K = 3$ is the Fama-French 3 factor model and $K = 4$ is the Carhart model) and the out-of-sample Sharpe ratio (OOS-SR) of the associated portfolios. R^2 GLS is the R^2 of the estimated linear factor model by Generalised Least Squares under its beta representation (see Lewellen et al., 2010), J-test (p-val) is the p-value of the GMM's overidentification test, RMSE is the root mean squared error, MAE is the mean absolute error. SR-Diff Test is a matrix containing pairwise p-values for the test of difference in Sharpe ratios of Ledoit and Wolf (2008). * significant at 90%, ** significant at 95% and *** significant at 99%. MV is the historical mean-variance tangent portfolio and MinV is the historical global minimum variance portfolio. First and second stage estimators refer to the GMM estimator. Uncentered and centered SDF differ in the

normalization imposes on the constant term in the linear SDF model.

Table 2: Linear Asset Pricing Model Estimation and MV Portfolio Performance
(Full Sample, Test Assets: 10 Industries)

First Stage Estimators							
Uncentered SDF ($a = 1$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.14	0.36	0.35	0.28	-0.009		
3	0.35	0.44	0.25	0.20	-0.023	0.77	
4	0.52	0.61	0.17	0.14	0.053	0.22 0.06*	
Centered SDF ($a = 1 - \lambda'E(f)$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.14	0.36	0.36	0.29	-0.006		
3	0.35	0.45	0.25	0.21	-0.016	0.69	
4	0.52	0.61	0.19	0.16	0.040	0.05** 0.02**	
Second Stage Estimators							
Uncentered SDF ($a = 1$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.14	0.36	0.44	0.36	-0.029		
3	0.35	0.44	0.36	0.30	-0.018	0.86	
4	0.52	0.61	0.32	0.27	0.058	0.31 0.16	
Centered SDF ($a = 1 - \lambda'E(f)$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.14	0.36	0.42	0.34	-0.056		
3	0.35	0.45	0.36	0.29	-0.013	0.23	
4	0.52	0.61	0.35	0.30	0.033	0.02** 0.11	
Benchmarks							
OOS SR	1/N	MV	MinV				
	0.148	-0.021	0.141				

See notes in table 1.

Table 3: Linear Asset Pricing Model Estimation and MV Portfolio Performance
(Full Sample, Test Assets: 17 Industries)

First Stage Estimators							
Uncentered SDF ($a = 1$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.08	0.27	0.40	0.32	-0.006		
3	0.21	0.31	0.30	0.24	0.008	0.79	
4	0.34	0.47	0.23	0.19	0.018	0.64	0.78
Centered SDF ($a = 1 - \lambda'E(f)$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.08	0.26	0.40	0.32	0.001		
3	0.21	0.31	0.30	0.24	0.018	0.74	
4	0.34	0.43	0.26	0.21	0.012	0.44	0.53
Second Stage Estimators							
Uncentered SDF ($a = 1$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.08	0.27	0.66	0.57	-0.034		
3	0.21	0.31	0.58	0.50	-0.017	0.75	
4	0.34	0.47	0.47	0.40	0.005	0.83	0.86
Centered SDF ($a = 1 - \lambda'E(f)$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.08	0.26	0.61	0.52	0.010		
3	0.21	0.31	0.52	0.44	-0.043	0.24	
4	0.34	0.43	0.49	0.42	-0.026	0.48	0.47
Benchmarks							
OOS SR	1/N	MV	MinV				
	0.141	0.051	0.105				

See notes in table 1.

Table 4: Linear Asset Pricing Model Estimation and MV Portfolio Performance
(Full Sample, Test Assets: 25 Size-Value)

First Stage Estimators							
Uncentered SDF ($a = 1$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.04	0.02	0.55	0.36	-0.005		
3	0.12	0.06	0.41	0.26	0.083	0.06*	
4	0.16	0.13	0.33	0.24	0.144	0.00*** 0.07*	
Centered SDF ($a = 1 - \lambda'E(f)$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.04	0.02	0.56	0.37	-0.008		
3	0.12	0.05	0.42	0.27	0.056	0.32	
4	0.16	0.10	0.31	0.24	0.121	0.25 0.59	
Second Stage Estimators							
Uncentered SDF ($a = 1$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.04	0.02	1.41	1.27	-0.003		
3	0.12	0.06	1.16	1.02	0.044	0.03**	
4	0.16	0.13	1.07	0.91	0.051	0.00*** 0.05**	
Centered SDF ($a = 1 - \lambda'E(f)$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.04	0.02	1.26	1.12	0.018		
3	0.12	0.05	1.02	0.88	0.043	0.66	
4	0.16	0.10	1.03	0.88	0.065	0.34 0.49	
Benchmarks							
OOS SR	1/N	MV	MinV				
	0.146	-0.017	0.176				

See notes in table 1.

Table 5: Linear Asset Pricing Model Estimation and MV Portfolio Performance
(Short Sample, Test Assets: 5 Industries)

First Stage Estimators								
Uncentered SDF ($a = 1$)								
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test		
1	0.29	0.40	0.27	0.22	-0.027			
3	0.76	0.48	0.13	0.11	-0.040	0.62		
4M	-0.70	0.64	0.06	0.05	0.003	0.16	0.34	
4L	-0.70	0.51	0.07	0.06	0.078	0.00***	0.02**	0.06*
Centered SDF ($a = 1 - \lambda'E(f)$)								
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test		
1	0.29	0.40	0.27	0.23	-0.055			
3	0.76	0.48	0.13	0.11	-0.020	0.87		
4M	-0.70	0.63	0.07	0.05	0.027	0.55	0.68	
4L	-0.70	0.51	0.07	0.06	0.042	0.42	0.21	0.54
Second Stage Estimators								
Uncentered SDF ($a = 1$)								
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test		
1	0.29	0.40	0.30	0.25	-0.017			
3	0.76	0.48	0.19	0.15	-0.012	0.21		
4M	-0.70	0.64	0.12	0.10	0.008	0.04**	0.30	
4L	-0.70	0.51	0.12	0.10	0.050	0.01***	0.15	0.66
Centered SDF ($a = 1 - \lambda'E(f)$)								
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test		
1	0.29	0.40	0.30	0.24	-0.012			
3	0.76	0.48	0.19	0.15	-0.019	0.97		
4M	-0.70	0.63	0.13	0.11	0.029	0.48	0.31	
4L	-0.70	0.51	0.12	0.10	0.063	0.25	0.10*	0.33
Benchmarks								
OOS SR	1/N	MV	MinV					
	0.106	0.025	0.122					

The table reports in-sample measures of goodness of fit for a set of linear asset pricing models ($K = 1$ is the CAPM, $K = 3$ is the Fama-French 3-factor model and $K = 4M$ is the Carhart model and $K = 4L$ is 4-factor liquidity model) and the out-of-sample Sharpe ratio (OOS-SR) of the associated portfolios. R^2 GLS is the R^2 of the estimated linear factor model by generalized least squares under its beta representation (see Lewellen et al., 2010), J-test (p-val) is the p-value of the GMM's overidentification test respectively, RMSE is the root mean squared error, MAE is the mean absolute error. SR-Diff Test is a matrix containing pairwise p-values

for the test of difference in Sharpe ratios of Ledoit and Wolf (2008). * significant at 90%, ** significant at 95% and *** significant at 99%. MV is the historical mean-variance tangent portfolio and MinV is the historical global minimum variance portfolio. First and second stage estimators refer to the GMM estimator. Uncentered and centered SDF differ in the normalization imposed on the constant term in the linear SDF model.

Table 6: Linear Asset Pricing Model Estimation and MV Portfolio Performance
(Short Sample, Test Assets: 10 Industries)

First Stage Estimators									
Uncentered SDF ($a = 1$)									
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test			
1	0.15	0.42	0.35	0.29	0.008				
3	0.36	0.52	0.23	0.19	0.031	0.77			
4M	0.57	0.67	0.16	0.13	0.057	0.51	0.54		
4L	0.55	0.57	0.18	0.15	0.099	0.24	0.04**	0.28	
Centered SDF ($a = 1 - \lambda'E(f)$)									
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test			
1	0.15	0.42	0.36	0.29	0.028				
3	0.36	0.52	0.23	0.19	0.032	0.50			
4M	0.57	0.66	0.18	0.14	0.074	0.24	0.39		
4L	0.55	0.58	0.19	0.15	0.128	0.03**	0.04**	0.10*	
Second Stage Estimators									
Uncentered SDF ($a = 1$)									
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test			
1	0.15	0.42	0.41	0.33	-0.024				
3	0.36	0.52	0.34	0.27	-0.012	0.96			
4M	0.57	0.67	0.27	0.22	0.026	0.55	0.36		
4L	0.55	0.57	0.29	0.23	0.083	0.20	0.02**	0.12	
Centered SDF ($a = 1 - \lambda'E(f)$)									
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test			
1	0.15	0.42	0.40	0.33	0.017				
3	0.36	0.52	0.34	0.27	-0.018	0.19			
4M	0.57	0.66	0.28	0.23	0.022	0.63	0.29		
4L	0.55	0.58	0.29	0.23	0.084	0.54	0.03**	0.07*	
Benchmarks									
OOS SR	1/N	MV	MinV						
	0.115	0.029	0.137						

See notes in table 5.

Table 7: Linear Asset Pricing Model Estimation and MV Portfolio Performance
(Short Sample, Test Assets: 17 Industries)

First Stage Estimators									
Uncentered SDF ($a = 1$)									
K	R^2	GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test		
1	0.08		0.36	0.42	0.33	-0.012			
3	0.23		0.41	0.30	0.23	-0.034	0.90		
4M	0.37		0.54	0.22	0.17	-0.034	0.90	0.99	
4L	0.33		0.48	0.25	0.20	0.040	0.40	0.09*	0.15
Centered SDF ($a = 1 - \lambda'E(f)$)									
K	R^2	GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test		
1	0.08		0.35	0.42	0.34	-0.013			
3	0.23		0.41	0.30	0.23	-0.031	0.52		
4M	0.37		0.19	0.23	0.19	-0.042	0.68	0.53	
4L	0.33		0.48	0.25	0.20	0.042	0.17	0.28	0.14
Second Stage Estimators									
Uncentered SDF ($a = 1$)									
K	R^2	GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test		
1	0.08		0.36	0.61	0.51	-0.017			
3	0.23		0.41	0.54	0.45	0.023	0.92		
4M	0.37		0.54	0.41	0.35	0.003	0.79	0.82	
4L	0.33		0.48	0.45	0.38	0.071	0.40	0.10*	0.11
Centered SDF ($a = 1 - \lambda'E(f)$)									
K	R^2	GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test		
1	0.08		0.35	0.59	0.49	-0.024			
3	0.23		0.41	0.48	0.40	0.018	0.34		
4M	0.37		0.53	0.41	0.34	0.005	0.70	0.59	
4L	0.33		0.48	0.42	0.35	0.079	0.06*	0.30	0.12
Benchmarks									
OOS SR	1/N	MV	MinV						
	0.109	0.059	0.040						

See notes in table 5.

Table 8: Linear Asset Pricing Model Estimation and MV Portfolio Performance
(Short Sample, Test Assets: 25 Size-Value)

First Stage Estimators									
Uncentered SDF ($a = 1$)									
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test			
1	0.05	0.01	0.42	0.34	-0.026				
3	0.14	0.06	0.22	0.17	0.071	0.13			
4M	0.18	0.10	0.18	0.14	0.167	0.00***	0.07*		
4L	0.18	0.07	0.20	0.15	0.115	0.01***	0.31	0.32	
Centered SDF ($a = 1 - \lambda'E(f)$)									
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test			
1	0.05	0.01	0.43	0.34	-0.034				
3	0.14	0.04	0.23	0.18	0.061	0.12			
4M	0.18	0.07	0.21	0.17	0.101	0.03**	0.22		
4L	0.18	0.05	0.22	0.17	0.118	0.07*	0.64	0.37	
Second Stage Estimators									
Uncentered SDF ($a = 1$)									
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test			
1	0.05	0.01	1.24	1.14	-0.009				
3	0.14	0.06	1.05	0.97	0.116	0.07*			
4M	0.18	0.10	0.88	0.81	0.181	0.02**	0.43		
4L	0.18	0.07	0.96	0.89	0.136	0.01***	0.22	0.66	
Centered SDF ($a = 1 - \lambda'E(f)$)									
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test			
1	0.05	0.01	1.15	1.04	0.030				
3	0.14	0.04	0.95	0.85	0.060	0.76			
4M	0.18	0.07	0.93	0.84	0.131	0.16	0.26		
4L	0.18	0.05	0.84	0.76	0.162	0.07*	0.20	0.52	
Benchmarks									
OOS SR	1/N	MV	MinV						
	0.130	0.049	0.219						

See notes in table 5.

Table 9: Linear Asset Pricing Model Estimation and MV Portfolio Performance
(Full Sample, Test Assets: 5 Industries, Short-Sales Constraints)

First Stage Estimators							
Uncentered SDF ($a = 1$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.29	0.40	0.26	0.22	0.030		
3	0.78	0.46	0.14	0.11	0.139	0.02**	
4	-0.71	0.59	0.07	0.06	0.127	0.05**	0.56
Centered SDF ($a = 1 - \lambda'E(f)$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.29	0.40	0.26	0.22	0.035		
3	0.78	0.46	0.14	0.11	0.141	0.32	
4	-0.71	0.59	0.07	0.06	0.116	0.01***	0.42
Second Stage Estimators							
Uncentered SDF ($a = 1$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.29	0.40	0.29	0.24	0.029		
3	0.78	0.46	0.20	0.17	0.075	0.03**	
4	-0.71	0.59	0.15	0.13	0.141	0.08*	0.37
Centered SDF ($a = 1 - \lambda'E(f)$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.29	0.40	0.29	0.24	0.022		
3	0.78	0.46	0.20	0.17	0.075	0.15	
4	-0.71	0.59	0.16	0.14	0.124	0.02**	0.41
Benchmarks							
OOS SR	1/N	MVc	MinVc				
	0.143	0.149	0.149				

See notes in table 1.

Table 10: Linear Asset Pricing Model Estimation and MV Portfolio Performance
(Full Sample, Test Assets: 10 Industries, Short-Sales Constraints)

First Stage Estimators							
Uncentered SDF ($a = 1$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.14	0.36	0.35	0.28	0.030		
3	0.35	0.44	0.25	0.20	0.137	0.00***	
4	0.52	0.61	0.17	0.14	0.157	0.00***	0.24
Centered SDF ($a = 1 - \lambda'E(f)$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.14	0.36	0.36	0.29	0.020		
3	0.35	0.45	0.25	0.21	0.137	0.04**	
4	0.52	0.61	0.19	0.16	0.140	0.00***	0.40
Second Stage Estimators							
Uncentered SDF ($a = 1$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.14	0.36	0.44	0.36	0.039		
3	0.35	0.44	0.36	0.30	0.144	0.00***	
4	0.52	0.61	0.32	0.27	0.164	0.00***	0.79
Centered SDF ($a = 1 - \lambda'E(f)$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.14	0.36	0.42	0.34	0.014		
3	0.35	0.45	0.36	0.29	0.123	0.02**	
4	0.52	0.61	0.35	0.30	0.151	0.00***	0.30
Benchmarks							
OOS SR	1/N	MVc	MinVc				
	0.148	0.146	0.162				

See notes in table 1.

Table 11: Linear Asset Pricing Model Estimation and MV Portfolio Performance
(Full Sample, Test Assets: 17 Industries, Short-Sales Constraints)

First Stage Estimators							
Uncentered SDF ($a = 1$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.08	0.27	0.40	0.32	0.028		
3	0.21	0.31	0.30	0.24	0.133	0.01***	
4	0.34	0.47	0.23	0.19	0.136	0.00***	0.52
Centered SDF ($a = 1 - \lambda'E(f)$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.08	0.26	0.40	0.32	0.034		
3	0.21	0.32	0.30	0.24	0.130	0.02**	
4	0.34	0.43	0.26	0.21	0.129	0.01***	0.47
Second Stage Estimators							
Uncentered SDF ($a = 1$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.08	0.27	0.66	0.57	0.012		
3	0.21	0.31	0.58	0.31	0.138	0.02**	
4	0.34	0.47	0.47	0.40	0.123	0.01***	0.47
Centered SDF ($a = 1 - \lambda'E(f)$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.08	0.26	0.61	0.52	0.026		
3	0.21	0.31	0.52	0.43	0.133	0.04**	
4	0.34	0.43	0.49	0.42	0.113	0.05**	0.18
Benchmarks							
OOS SR	1/N	MVc	MinVc				
	0.141	0.148	0.159				

See notes in table 1.

Table 12: Linear Asset Pricing Model Estimation and MV Portfolio Performance
(Full Sample, Test Assets: 25 Size-Value, Short-Sales Constraints)

First Stage Estimators							
Uncentered SDF ($a = 1$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.04	0.02	0.55	0.36	0.019		
3	0.12	0.06	0.41	0.26	0.106	0.03**	
4	0.16	0.13	0.33	0.24	0.100	0.04**	0.50
Centered SDF ($a = 1 - \lambda'E(f)$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.04	0.02	0.56	0.37	0.014		
3	0.12	0.05	0.42	0.27	0.103	0.34	
4	0.16	0.10	0.31	0.24	0.097	0.89	0.21
Second Stage Estimators							
Uncentered SDF ($a = 1$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.04	0.02	1.41	1.27	0.034		
3	0.12	0.06	1.16	1.02	0.083	0.03**	
4	0.16	0.13	1.07	0.91	0.055	0.03**	0.55
Centered SDF ($a = 1 - \lambda'E(f)$)							
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.04	0.02	1.26	1.12	0.045		
3	0.12	0.05	1.02	0.88	0.097	0.11	
4	0.16	0.10	1.03	0.88	0.056	0.99	0.26
Benchmarks							
OOS SR	1/N	MVc	MinVc				
	0.145	0.151	0.153				

See notes in table 1.

Table 13: Linear Asset Pricing Model Estimation and MV Portfolio Performance
 (Short Sample, Test Assets: 5 Industries, Short-Sales Constraints)

First Stage Estimators									
Uncentered SDF ($a = 1$)									
K	R^2	GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test		
1	0.29		0.40	0.27	0.22	0.003			
3	0.76		0.48	0.13	0.11	0.088	0.18		
4M	-0.70		0.63	0.06	0.05	0.075	0.36	0.68	
4L	-0.70		0.51	0.07	0.06	0.086	0.22	0.92	0.61
Centered SDF ($a = 1 - \lambda'E(f)$)									
K	R^2	GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test		
1	0.29		0.40	0.27	0.23	0.007			
3	0.76		0.48	0.13	0.11	0.090	0.88		
4M	-0.70		0.63	0.07	0.05	0.090	0.21	0.28	
4L	-0.70		0.51	0.07	0.06	0.097	0.87	0.72	0.06*
Second Stage Estimators									
Uncentered SDF ($a = 1$)									
K	R^2	GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test		
1	0.29		0.40	0.30	0.25	0.041			
3	0.76		0.48	0.19	0.15	0.044	0.22		
4M	-0.70		0.64	0.12	0.10	0.122	0.24	0.93	
4L	-0.70		0.51	0.12	0.10	0.071	0.18	0.79	0.87
Centered SDF ($a = 1 - \lambda'E(f)$)									
K	R^2	GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test		
1	0.29		0.40	0.30	0.24	0.038			
3	0.76		0.48	0.19	0.15	0.039	0.92		
4M	-0.70		0.63	0.13	0.11	0.103	0.25	0.23	
4L	-0.70		0.51	0.12	0.10	0.063	0.70	0.72	0.06*
Benchmarks									
OOS SR	1/N	MVc	MinVc						
0.106		0.122	0.120						

See notes in table 5.

Table 14: Linear Asset Pricing Model Estimation and MV Portfolio Performance
(Short Sample, Test Assets: 10 Industries, Short-Sales Constraints)

First Stage Estimators									
Uncentered SDF ($a = 1$)									
K	R^2	GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test		
1	0.15		0.42	0.35	0.29	0.036			
3	0.36		0.52	0.23	0.19	0.098	0.24		
4M	0.57		0.67	0.16	0.13	0.135	0.08* 0.25		
4L	0.55		0.57	0.18	0.15	0.127	0.08* 0.30 0.77		
Centered SDF ($a = 1 - \lambda'E(f)$)									
K	R^2	GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test		
1	0.15		0.42	0.36	0.29	0.035			
3	0.36		0.52	0.23	0.19	0.097	0.41		
4M	0.57		0.66	0.18	0.14	0.106	0.24 0.89		
4L	0.55		0.58	0.19	0.15	0.126	0.14 0.59 0.47		
Second Stage Estimators									
Uncentered SDF ($a = 1$)									
K	R^2	GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test		
1	0.15		0.42	0.41	0.33	0.054			
3	0.36		0.52	0.34	0.27	0.127	0.23		
4M	0.57		0.67	0.27	0.22	0.133	0.13 0.61		
4L	0.55		0.57	0.29	0.23	0.152	0.08* 0.36 0.52		
Centered SDF ($a = 1 - \lambda'E(f)$)									
K	R^2	GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test		
1	0.15		0.42	0.40	0.33	0.050			
3	0.36		0.52	0.34	0.27	0.103	0.46		
4M	0.57		0.66	0.28	0.23	0.136	0.08* 0.31		
4L	0.55		0.58	0.29	0.23	0.133	0.25 0.57 0.68		
Benchmarks									
OOS SR	1/N	MVc	MinVc						
	0.115	0.097	0.130						

See notes in table 5.

Table 15: Linear Asset Pricing Model Estimation and MV Portfolio Performance
 (Short Sample, Test Assets: 17 Industries, Short-Sales Constraints)

First Stage Estimators									
Uncentered SDF ($a = 1$)									
K	R^2 GLS	J-Test (p-value)	RMSE	MAE	OOS SR	SR Diff.Test			
1	0.08	0.40	0.42	0.33	0.028				
3	0.23	0.41	0.30	0.23	0.111	0.14			
4M	0.37	0.54	0.22	0.17	0.138	0.03** 0.12			
4L	0.33	0.48	0.25	0.20	0.113	0.19 0.91 0.31			
Centered SDF ($a = 1 - \lambda'E(f)$)									
K	R^2 GLS	J-Test (p-value)	RMSE	MAE	OOS SR	SR Diff.Test			
1	0.08	0.35	0.42	0.34	0.019				
3	0.23	0.41	0.30	0.23	0.111	0.02**			
4M	0.37	0.53	0.23	0.19	0.123	0.06* 0.20			
4L	0.33	0.48	0.25	0.20	0.106	0.08* 0.49 0.59			
Second Stage Estimators									
Uncentered SDF ($a = 1$)									
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test			
1	0.08	0.36	0.61	0.51	0.015				
3	0.23	0.41	0.54	0.45	0.170	0.13			
4M	0.37	0.54	0.41	0.35	0.118	0.05 0.23			
4L	0.33	0.48	0.45	0.38	0.148	0.17 0.83 0.32			
Centered SDF ($a = 1 - \lambda'E(f)$)									
K	R^2 GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test			
1	0.08	0.35	0.59	0.49	0.037				
3	0.23	0.41	0.48	0.40	0.068	0.08*			
4M	0.37	0.53	0.41	0.34	0.101	0.31 0.56			
4L	0.33	0.48	0.42	0.35	0.134	0.16 0.69 0.32			
Benchmarks									
OOS SR	1/N	MVc	MinVc						
	0.109	0.118	0.117						

See notes in table 5.

Table 16: Linear Asset Pricing Model Estimation and MV Portfolio Performance
 (Short Sample, Test Assets: 25 Size-Value, Short-Sales Constraints)

First Stage Estimators								
Uncentered SDF ($a = 1$)								
K	R^2	GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.05		0.01	0.42	0.34	0.009		
3	0.14		0.06	0.22	0.17	0.112	0.11	
4M	0.18		0.10	0.18	0.14	0.122	0.06*	0.78
4L	0.18		0.07	0.20	0.15	0.105	0.11	0.22 0.20
Centered SDF ($a = 1 - \lambda'E(f)$)								
K	R^2	GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.05		0.01	0.43	0.34	0.009		
3	0.14		0.04	0.23	0.18	0.117	0.37	
4M	0.18		0.07	0.21	0.17	0.123	0.95	0.27
4L	0.18		0.05	0.22	0.17	0.109	0.43	0.99 0.23
Second Stage Estimators								
Uncentered SDF ($a = 1$)								
K	R^2	GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.05		0.007	1.24	1.14	0.033		
3	0.14		0.06	1.05	0.97	0.100	0.17	
4M	0.18		0.10	0.88	0.81	0.038	0.06**	0.99
4L	0.18		0.07	0.96	0.88	0.101	0.10*	0.44 0.37
Centered SDF ($a = 1 - \lambda'E(f)$)								
K	R^2	GLS	J-Test (p-val)	RMSE	MAE	OOS SR	SR Diff.Test	
1	0.05		0.01	1.15	1.04	0.044		
3	0.14		0.04	0.95	0.85	0.104	0.30	
4M	0.18		0.07	0.93	0.84	0.021	0.78	0.22
4L	0.18		0.05	0.84	0.76	0.104	0.31	0.98 0.15
Benchmarks								
OOS SR	1/N	MVc	MinVc					
	0.130	0.159	0.124					

See notes in table 5.