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

Massimo Guidolin* Erwin Hansen† Martín Lozano-Banda‡

September, 2017

Abstract

We evaluate linear stochastic discount factor models using an ex-post 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 spanning the period 1968-2016, we find 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 they are used to inform one-month ahead portfolio selection. When we compare portfolios associated to multifactor models with mean-variance decisions implied by the single-factor CAPM, we document statistically significant differences in Sharpe ratios of up to 10 percent. Linear multifactor models that provide the best in-sample fit also yield the highest realized Sharpe ratios.

JEL *classification: G11, G12.*

*Department of Finance, CAREFIN, and IGIER, Bocconi University, Milan, Italy. Email: massimo.guidolin@unibocconi.it.

†Facultad de Economía y Negocios, Universidad de Chile. Address: Diagonal Paraguay 257, Oficina 1204, Santiago, Chile. Tel: +56(2)29772125. Email:ehansen@fen.uchile.cl

‡Instituto Tecnológico y de Estudios Superiores de Monterrey, Campus Saltillo. Address: Prolongación Juan de la Barrera N. 1241 Ote, 4118000, Saltillo, México. Tel: +52 (81) 8358-2000. Email: mlozanoqf@gmail.com.

Keywords: Linear asset pricing models, Stochastic discount factor, Portfolio selection, Out-of-sample performance.

1 Introduction

Linear factor asset pricing models are well established in finance. The CAPM model of Sharpe (1964) and Litner (1965), the three-factor model of Fama and French (F-F, 1993), and the four-factor model of Carhart (1997) are now extensively used by researchers and practitioners to compute the cost of capital and risk-adjusted returns. Previous empirical evidence evaluating the ability of linear factor models to fit the cross-section of asset returns has generally favored multi-factor models over the single-factor CAPM. Carhart's four-factor model, for example, has been particularly successful in accounting for most of the anomalies challenging the efficient market hypothesis (see Schwert, 2003). However, in this literature, models are usually evaluated by comparing measures of in-sample, statistical goodness-of-fit, such as the R^2 of cross-sectional regressions of mean excess returns on a set of factor mimicking returns, or by in-sample pricing accuracy statistics.[1] Much less is known about the actual out-of-sample (henceforth, OOS) relative performance of alternative linear asset pricing models, both in a statistical sense (e.g., their OOS predictive R^2s) and especially in an economic perspective, i.e., whether or not commonly used linear pricing models may better support financial decisions by investors when compared to standard benchmarks, that often lack of any asset pricing foundations. To fill this gap, our paper evaluates and compares linear factor models in a OOS *economic* perspective. In particular, we study the OOS realized performance of mean-variance efficient portfolios, when the test assets are the predicted excess returns generated by a range of linear factor models.

The main contribution of our work consists in investigating the (differential) power of alternative linear stochastic discount factor (SDF)-based models to yield economic value over and beyond improving the in- and out-of-sample statistical performance at explaining the cross-section of asset returns. Equivalently, we do not only care for statistical fit (albeit of a OOS type), but also for portfolio performance. We attack this issue by adopting a fairly simple

[1] See, for example, Kan et al. (2013) for a recent application.

and yet commonly used mean-variance (henceforth, MV) portfolio perspective. In a way, this represents a further contribution of our study because, even though in theory it is well understood that linear pricing models are intrinsically connected to MV efficient portfolios (see the discussion in Cochrane, 2004), it is customary for applied portfolio choice papers to ignore asset pricing models in the estimation of optimal weights.[2] On the contrary, it has become normal to assume a functional form for the joint, multivariate distribution of returns, to use some historical samples to estimate the parameters of interest, and to then proceed to compute weights, often assessing ex-post their realized performance. In the absence of solid grounding in asset pricing theory (and unless specific constraints have been imposed, as in Brandt, 1999), such MV asset allocations may even end up being grounded into empirical models that admit arbitrage opportunities, which may be thought of as unrealistic in most applications.

In our paper, we build instead MV portfolios that are exclusively grounded in asset pricing theory for which the SDF has a linear specification. Moreover, we use a natural link between the SDF and MV-efficient portfolios first discussed by Chamberlain and Rothschild (1983), to construct an alternative evaluation metric to be used in the empirical asset pricing literature. Our intuition is that different views about the sources of risk must affect the portfolio allocation of MV investors and therefore imposing such views may have economically relevant effects that we exploit to study the OOS realized performance of a set of linear factor models. Moreover, we also investigate whether a superior in-sample fit at the cross-sectional level is associated with a superior OOS realized performance. It is important to emphasize that as far as we know, neither questions admit a trivial answer. In general, it is perfectly plausible (even though unwelcome) that models performing well in-sample may show a poor

[2]See Brandt (2010) for a description of the plug-in method and the references therein. Of course, at face value only our benchmark, the CAPM, is (under some assumptions) the equilibrium asset pricing model derived from MV efficient portfolio choice; to the contrary, the other multi-factor models investiaged in the paper are not necessarily consistent with static mean-variance optimizing behavior at the investor-level, in spite of the connections between all linear SDF models and MV efficiency studied since Chamberlain and Rothschild (1983) and recently emphasized by Cochrane (2009).

performance OOS, or the other way around.

In practice, our empirical strategy consists of two steps. In the first step, we estimate linear factor models using the SDF/GMM method described by Cochrane (2009). There are two reasons to prefer the SDF/GMM method over the more standard beta method. First, the SDF representation is more 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 measures of in-sample goodness of fit as possible, this property of the SDF method is particularly appealing to us. In the second step, we estimate MV efficient portfolios using the predicted excess returns from the linear factor models. Theoretically, we use the connection between the SDF and the MV frontier established by Hansen and Richard (1987): we estimate the structure of the MV efficient portfolio using the concept of *mean-representing portfolio* introduced by Chamberlain and Rothschild (1983);[3] as shown by Peñaranda and Sentana (2011, 2012), the MV frontier can then be consistently estimated by GMM using a set of moment conditions implied by the definition of the mean-representing portfolio. Finally, we compare the OOS performance of the resulting MV efficient portfolios associated to linear factor models using the bootstrap proposed by Ledoit and Wolf (1998) to perform tests of the significance of differences in Sharpe ratios.

We conduct our empirical tests using a monthly sample of US equity portfolio returns spanning the period 1968-2016. The portfolios are both industry-sorted portfolios and the classical Fama and French's (henceforth, F-F) size- and value-sorted portfolios. The key results of

[3] We impose short sale constraints on portfolio weights estimation throughout. Jagannathan and Ma (2003) argue that a trade-off exists between specification error and estimation error when short sales constraint are imposed.

the paper are summarized in Figure 1, where we plot the in-sample root mean squared error (RMSE) against the OOS Sharpe ratio of a number of linear factor models.[4] Each dot represents a particular test asset/model pair and the number written nearby indicates the number of factors in the model. The figure implies two differences between its horizontal vs. vertical coordinates. On the vertical axis, we report one *in-sample, statistical* measure of fit, the RMSE. The vertical axis is deemed to represent the standard criterion used in the literature, where the pricing performance of alternative models is assessed on the basis of their quantitative cross-sectional fit. On the horizontal axis, we plot instead the values for one OOS economic index of performance, i.e., the realized Sharpe ratios of the portfolios constructed on the basis of the forecasts derived from alternative factor models.[5] If both our conjectures held, then we would expect that the dots should cluster by model along some imaginary minus 45-degree upward sloping line, possibly offering one clear winner, located in the rightmost lower portion of the plot, where the highest OOS Sharpe ratios are found along with the lowest RMSE. As a matter of fact, two results emerge from Figure 1. First, the estimated negative slope of the linear fit in the plot reveals that the best performing models in-sample are, at the same time, the best OOS performing models according to our portfolio metric. Second, multifactor models (numbered 3, 4 and 5) outperform the CAPM. The cloud of points of multifactor models is located in the lower right corner, whilst the cloud of points generated by the CAPM is located in the upper left region in the plot. This result is well known from an in-sample perspective but new in a OOS portfolio perspective.

In more detail, we find that the best performing models in-sample (i.e., commanding the highest GLS R^2s and/or the lowest RMSEs) are also those that yield the highest realized one-month ahead Sharpe ratios. Multifactor models consistently achieve higher OOS Sharpe ratios than the CAPM does. The estimated difference in Sharpe ratios may reach values

[4] As explained in Section 2, in this paper, models will be distinguishable according to the numbers of factors included, the type of GMM estimator used, and the normalization of the SDF considered.

[5] Of course, collapsing two dimensiones of performance in one Cartesian plot has its limitations, even though it offers the advantage of immediacy.

as high as 10 per cent and are often statistically significant. These results are valid for alternative asset menus, one- and two-step GMM estimators, and centered and uncentered versions of the SDF. Moreover, we study the ability of the 4-factor model that includes the liquidity risk factor of Pastor and Stambaugh (2003) at explaining the cross-section of expected returns and yielding useful MV portfolio prescriptions: even though the liquidity factor model is outperformed by Carhart's model in-sample, it is also able to outperform the other models, in terms of OOS Sharpe ratios.

Our paper contributes to two strands of literature. First, we bring new evidence to the literature evaluating the empirical performance of linear factor models, see Jagannathan and Wang (2002), Shanken and Zhou (2007), Lewellen et al. (2010), and Lozano and Rubio (2011), Kan et al. (2013), among others. This literature has developed a pure in-sample econometric approach to the assessment of the performance of alternative linear factor models. We aim to contribute to this literature by also documenting the empirical, realized OOS performance of the same models when they are used to support portfolio decisions. Second, we contribute to a less-developed literature that studies how asset pricing models can provide useful insights in portfolio choice problem (see for example Brandt, 1999; Pastor and Stambaugh, 2000; MacKinlay and Pastor, 2000; Pastor, 2000; Chevrier and McCulloch, 2008; Connor and Korajczyk, 2010).[6] In this perspective, we aim at proposing a simple and

[6]Pastor and Stambaugh (2000) and Pastor (2000) have used risk-based and characteristic-based asset pricing models to center the prior beliefs of a Bayesian investor solving a MV problem. Because we adopt a frequentist GMM approach, our paper is mute on the role played by priors on the models, although we follow the spirit of Pastor (2000). Also adopting a Bayesian framework, Chevrier and McCulloch (2008) have studied how economically motivated priors may help building portfolios that outperform the equally-weighted portfolio ($1/N$). They use linear factor models as one of their sets of economically-grounded priors. Although our goals are different, we also uncover evidence that standard asset pricing models may push realized performance over the hurdle represented by the equal weighting benchmark. 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 MV portfolios. They conclude that in a model with one unobserved factor, the covariance matrix of returns collapses to the identity matrix, and the associated MV portfolio outperforms many benchmarks in terms of OOS returns. Behr et al. (2012) study 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, F-F 3-factor model, and Carhart's model). These portfolios—the ones linked to linear factor models—yield higher Sharpe ratios than most of the other

yet robust methodology that combines the estimation of linear factor models using the SDF method and the estimation of MV efficient portfolios using a consistent GMM estimator. In fact, we show that the portfolios implied by multifactor models may often yield risk-adjusted returns that outperform the simple $1/N$ strategy of De Miguel et al. (2009), indicating that these models may potentially become a worthy tool in the hands of investors.

The paper is organized as follows. In Section 2, we describe our methodology. The first subsection is devoted to GMM estimation of linear pricing models; in the second subsection, we introduce the concept of mean-representing portfolio and we describe the GMM estimator to compute the portfolio weights; in the third subsection, we explain how to test for differences in Sharpe ratios using a block bootstrap. In Section 3, we describe our data. Estimation results and key findings are reported and discussed in Section 4. Section 5 concludes.

2 Methodology

In order to estimate MV efficient portfolios backed by alternative linear factor asset pricing models, we proceed in two stages. In the first stage, we estimate by GMM linear asset pricing models written in their SDF form, following Cochrane (2009); in the second stage, we use again GMM to compute MV efficient portfolios following Peñaranda and Sentana (2011, 2012). Resulting portfolio performances are then compared using bootstrapped tests of differences in Sharpe ratios first proposed by Ledoit and Wolf (2008).

---

portfolios under analysis. Kirby and Ostdiek (2012) employ the F-F 3-factor and Carhart's models to estimate conditional expected returns. More recently, Tu and Zhou (2011) have shown 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.

2.1 GMM, SDF-based estimation of linear asset pricing models

For the case of excess returns, the fundamental pricing equation is

$$0 = E_t \left[m_{t+1} r_{t+1} \right], \tag{1}$$

where 0 is an $N \times 1$ vector of zeros, $E_t[\cdot]$ 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 r_{t+1} is the $N \times 1$ vector of excess returns in period $t+1$. This expression indicates that the conditional expected excess returns of any asset, after being discounted to the present time by the stochastic discount factor m_{t+1}, are zero. The stochastic discount factor, m_{t+1}, represents the realization of any random variable satisfying (1) between t and $t+1$. As customary, a few additional assumptions need to be imposed in order for m_{t+1} be uniquely defined and positive.[7]

In particular, in this paper we assume that m_{t+1} is characterized by the following linear functional form:

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

where b is a $K \times 1$ vector of parameters to be estimated and f_{t+1} are the realizations of K risk factors at time $t+1$. As it has been noted in previous empirical work (see Burnside, 2007), additional 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 $\widehat{m}$ is the estimated SDF. Therefore, from equation (1), it holds that $E[\widehat{m}r] = 0$. Now, for any constant c, the SDF $\widetilde{m} = c\widehat{m}$ also satisfies (1), i.e. $E\left[\widetilde{m}r\right] = 0$. From this example, it is clear that an infinite number of SDFs exist to satisfy (1) simultaneously. This problem is solved by normalizing the value of the constant a in (2). As pointed out by Cochrane

[7]As far as uniqueness is concerned, the assumption of complete markets is necessary. As for positiviness, both the abssence of arbitrage and the law of one price are required. See Chapter 4 in Cochrane (2009) for further details.

(2009), the choice of this normalization only depends on convenience. The first and simplest normalization consists of imposing $a = 1$. In this case, we say that the SDF is *uncentered*. The second normalization is $a = 1 + b'E(f)$, which corresponds to the *centered* SDF case.[8] After imposing a normalization on a, the set of parameters b is estimated by GMM using the pricing errors from (1) as ingredients to selected moment conditions, which we describe in the next section.

2.2 Mean-representing portfolio and its estimation by GMM

In this section, we introduce the concept of mean-representing portfolio and how we can use it to estimate a unique (up to an scalar) MV efficient portfolio for each linear factor model. Applying Riesz's representation theorem, Chamberlain and Rothschild (1983) have proven the existence of a unique portfolio (called the mean-representing portfolio, p^0) in the set of all possible portfolios formed for a particular vector of returns of a set of test assets, whose weights are proportional to the weights of the MV efficient portfolio. As a consequence, these two portfolios have the same Sharpe ratio. We use this property to estimate the Sharpe ratio of the MV efficient portfolio that we use to compare economically alternative linear asset pricing models.

2.2.1 Mean-representing portfolio

Consider a set of N risky assets and one risk-free asset. Define $r = (r_1, ..., r_N)'$ to be the set of returns in excess of the risk-free rate for 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 $w = (w_1, ..., w_N)'$. The set of all possible portfolios

[8] The properties of SDF-based asset pricing models under both normalizations have been studied by Burnside (2007) and Lozano and Rubio (2011).

built starting from the excess returns in 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 also zero.[9]

In addition, 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 $\mathbb{P}$. It is well known that the set $L_2(P)$ is a Hilbert space under the mean-square inner product: $(p, q) \equiv E(pq)$ 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 Riesz's representation theorem, to 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

$$E(p) = E(p^0 p) \quad \forall p \in P. \tag{3}$$

Under this topology, the mean representing portfolio is defined as:

$$p^0 = E(r')E(rr')^{-1}r = \phi^0 r. \tag{4}$$

From this expression, we observe that the weights of the mean representing portfolio are given by the vector $\phi^0 \equiv E(r')E(rr')^{-1}$. A useful property of this construction methodology is that there exist a one-to-one mapping between p^0 and any portfolio on the MV efficient frontier, r^{MV}, as described by

$$r^{MV} = \mu \frac{1}{E(p^0)} p^0 \tag{5}$$

$$V\left[r^{MV}(\mu)\right] = \left[\frac{1 - E(p^0)}{E(p^0)}\right] \mu^2, \tag{6}$$

[9] In a more general setup, when $C(r) \neq 0$, it is anyway possible to define a costly representing portfolio that in addition to the mean representing portfolio, characterizes the efficient mean-variance frontier.

where $\mu \in \mathbb{R}$ is the expected return and $V\left[r^{MV}(\mu)\right]$ is the variance of the optimal MV portfolio.[10] In this framework, the Sharpe ratio of all portfolios on the MV frontier is the same for all values of μ, and equal to the Sharpe ratio of p^0.[11]

From (5), note that the weights of the MV portfolio, r^{MV}, are simply proportional to the weights of the *unique* 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 any MV portfolio.

2.2.2 Portfolio estimation by GMM

The portfolio weights, ϕ^0, are estimated using GMM. In particular, Peñaranda and Sentana (2011, 2012), propose to use the following system of $N + 1$ moment conditions to estimate the weights of the mean-representing portfolio:

$$E\begin{bmatrix} rr'\phi^0 - r \\ r\phi^0 - \mu^0 \end{bmatrix} = 0_{(N+1\times 1)}. \tag{7}$$

Here, the first N moment equations comes directly from the definition in (3), 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. Peñaranda and Sentana (2011) show, that under some regularity conditions, the GMM estimates of the coefficients, $\theta = (\phi^0, \mu^0)$, are consistent.

Peñaranda and Sentana (2011) propose to incorporate linear factor models in estimation of

[10] Taking expectations in (5), we have $E(r^{MV}) = \mu$. The expression of the variance in (6) uses the fact that $E(p^0) = E\left(p^0 p^0\right)$ in (3).

[11] If $r^{MV} = ap^0$, where a is a constant defined as $\mu/E(p^0)$, then

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

the MV frontier, expanding (7) to the additional moment conditions, $E[(1-bf)r]=0$, and estimating the model in one step. We use instead a two-step approach in which, in a first stage, linear factor models are estimated and fitted excess returns are obtained, and then, in the second step, the efficient MV portfolio is estimated by GMM using (7).[12] We opted in favor of using this two-step approach considering the fact that in preliminary experiments, the estimated weights turned out to be more stable under this approach than adopting a one-step method. We consider this advantage to be prevalent over the increase in estimation error implied by our two-step set up.

2.3 Portfolio performance evaluation

We evaluate the ex-post realized performance of the portfolios computed in the manners described in the previous section, by testing whether there are statistically significant differences between their OOS, realized Sharpe ratios. In particular, we perform pairwise tests for all possible combination of models. We use the robust bootstrapped test of differences in Sharpe ratios proposed by Ledoit and Wolf (2008).[13] This test uses the circular-block bootstrap of Politis and Romano (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 in Sharpe ratios between any two portfolios. This test is suitable for our exercise as it explicitly accommodates non-normality and time dependence in excess returns data through resampling.

Additionally, to increase the accuracy of our performance analysis, a small sample bias correction is applied to the estimated Sharpe ratios before testing. In particular, Opdyke (2007)

[12]For the case of excess returns (r), we have that the pricing equation under the SDF form is given by $E(r_{t+1}m_{t+1}) = 0$ where m_t is the stochastic discount factor. For linear asset pricing models, under a proper normalization, we can assume that $m_t = 1 - b'f_t$. Thus, we have $E[r_{t+1}(1 - b'f_t)] = 0$. In this context, pricing errors are defined as $\pi = E(r_{t+1}) - b'E(r_t b' f t_{+1})$. The second term in this expression are the fitted returns. After estimating b, and considering different factors in the model (f_t), we can estimate them. A similar approach is followed by Jagannathan and Wuang (2002).

[13]This test has been used in portfolio applications by De Miguel et al. (2009 , 2014), among others.

proves that, in the case of small samples, the expected value of the estimated Sharpe ratio is

$$E\left(\widehat{SR}\right) = SR\left[1 + \frac{1}{4}\frac{\left(\frac{\hat{\mu}_4}{\hat{\sigma}^4} - 1\right)}{T}\right], \tag{8}$$

where $\hat{\mu}_4/\hat{\sigma}^4$ is simply the sample kurtosis of excess returns. Regardless of the assumption made on the distribution of portfolio returns, this result is valid asymptotically.

We report a measure of portfolio turnover to quantify the amount of trading required to implement each of the evaluated portfolios. Higher turnover implies higher transaction costs, therefore, the final portfolio profitability is reduced. As is De Miguel et al. (2009), portfolio turnover is defined as the average sum of the trades across the N assets in the portfolio as follows:

$$Turnover = \frac{1}{T-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\left|\widehat{\phi}_{K,\, i,t+1} - \widehat{\phi}_{K,\, i,t^{+}}\right|\right),$$

where $\widehat{\phi}_{k,\, j,t+1}$ is the optimal portfolio weight in asset i at time $t+1$ for model K, and $\widehat{\phi}_{K,\, j,t^{+}}$ is the portfolio weight before rebalancing at $t+1$. Note that $\widehat{\phi}_{K,\, j,t^{+}}$ is different than the optimal portfolio weight at time t, $\widehat{\phi}_{k,\, j,t}$, in most of the cases because of changes in the prices of the assets in the portfolio.

Finally, note that reported OOS Sharpe ratios are built using portfolio returns net of transaction costs. Following De Miguel et al. (2009), we incorporate proportional transaction costs by computing portfolio net returns as:

$$R_{p,t+1} = \sum_{i=1}^{N}\widehat{\phi}_{\,i,t}R_{i,t+1} - c\sum_{i=1}^{N}\left|\widehat{\phi}_{i,t} - \widehat{\phi}_{i,t-1^{+}}\right|$$

where c is the proportional transaction cost of each trade in the process of rebalancing the portfolio. We assume c equals 50 basis points.

3 Data

Our data consist of monthly, value-weighted, US portfolio excess returns (over the risk-free rate), from January 1968 to December 2016. [14] The data is collected from alternative sources. Fama and French factors and the Momentum factor are retrieved from Kenneth French's online data library. Pastor and Stambaugh (2003)'s liquidity factor comes from Lubos Pastor's website. Finally, the data of the 4-factor model of Hou, Xue and Zhang (2015) was kindly provided by Professor Hou. The set of test assets considered in our main results are 10 and 17 industry portfolios, plus the 25 double-sorted size and book-to-market portfolios.[15] We consider a 49 industry portfolio and a 25 double-sorted size and momentum portfolio in additional exercises. All the test asset returns are collected from Kenneth French's library as well.

The set of factors considered in the analysis includes the excess return on the market portfolio over the risk-free rate as proxied by 1-month T-bill returns (MKT), the size portfolio (SMB-small minus big), the value portfolio (HML - high minus low), the momentum portfolio (MOM), the liquidity factor (LIQ) of Pastor and Stambaugh (2003), and the profitability portfolio (RMW - robust minus weak) and the investment portfolio (CMA - conservative minus aggressive) incorporated in Fama and French (2015). Hou, Xue and Zhang (2015) also propose to consider a profitability factor and an investment factor that we use when required. 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. The liquidity factor corresponds to the value-weighted return on the 10-1 portfolio from a sort on historical liquidity betas.[16] The profitability and investment factors

[14] The sample period is defined by the availability of data of the whole set of tested models and test assets. Both, the liquidity factor of Pastor and Stambaugh (2003) and the factors in Hou, Xue and Zhang (2015) are available from 1968:01 onwards.

[15] Abhakrom et al. (2013) also use 25 size & value portfolios and 10 industries portfolios to evaluate the value added by Fama-French factors to the C-CAPM.

[16] In 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

introduced by Fama and French (2015) are double-sorted portfolios on size and operating profitability[17] and size and investment, measured as the change in total assets, respectively. Hou, Xue and Zhang (2015) uses ROE as profitability measure and the change in total assets as investment measure as well. Our final sample (1968:01-2016:12) contains 588 months.

In summary, we study six linear factor models: the CAPM ($K = 1$), the three-factor model ($K = 3$) of Fama and French (1993), the four-factor model ($K = 4M$) of Carhart (1997), the four-factor liquidity model ($K = 4L$) of Pastor and Stambaugh (2003), the four-factor model ($K = 4H$) of Hou, Xue and Zhang (2015), and finally, the five-factor model ($K = 5$) of Fama and French (2015).

4 Empirical Results

In this Section, we describe and analyze our empirical results.

4.1 Empirical Setup

Our empirical strategy relies on a 5-year moving rolling window set up.[18] Our first assessment window starts in January 1968 and ends in December 1972. For this window, we estimate the set of linear asset pricing models under examination and the associated MV efficient portfolios. Then, using the excess returns actually observed in January 1973, we compute the realized ex-post portfolio return on the basis of the weights computed as of the end of December 1972. Next, we move the estimation window one period forward, covering now

webpage.

[17] Operating profitability is measured as revenues minus cost of goods sold, minus selling, general, and administrative expenses, minus interest expense all divided by book equity.

[18] De Miguel et al. (2009) use a similar rolling window framework in applied analysis of realized portfolio performances. We have performed robustness checks and repeated our tests for alternative window lengths in the rolling estimation setup finding essentially unchanged results that remain available from Author(s) upon request.

the sample February 1968 - January 1972. We re-estimate the models, the implied MV portfolio weights, and compute realized ex-post portfolio return realized during February 1973. We repeat this process for the 528 windows that can be built from our data. The time series of realized ex-post portfolio returns $\{r_w^p\}_{w=1}^{528}$ is used to compute the OOS Sharpe ratio as $\mu(r_w^p)/\sigma(r_w^p)$. Finally, we apply the small-sample bias correction using (8) and compare the portfolio performances generated by alternative SDFs using the test of Sharpe ratios differences by Ledoit and Wolf (2008).

We use five sets of test assets: 10 industries,17 industries, and the 25 size-value portfolios for our main results, and 25 size-momentum and 49 industries for additional results. Thus, we report and comment five different sets of results (tables), one for each asset menu. Each table contains four panels proving results for first and second stage GMM estimators, and uncentered and centered SDF models. Following Jagannathan and Ma (2003), we impose short-sale constraints in the estimation of our portfolios as a way to increase the robustness of MV portfolios and protect them against excessive variation induced by sampling error afflicting the sample moments that represent their ingredients.[19]

In each panel, we report in-sample measures of fit (GLS R^2, the p-value of the J-Test, and the root mean squared error), the OOS Sharpe ratio accounting for transaction costs, a measure of portfolio turnover, and finally, the p-value of the test of differences in Sharpe ratios for pairs of models. The GLS R^2 is the Generalized Least Squared cross-sectional R^2 of linear factor models. Lewellen et al. (2010) suggest reporting this measure, instead of the standard OLS R^2 for example, to evaluate alternative asset pricing models. The higher the GLS R^2, the better 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(p-val) statistic is the p-value of the chi-squared J-test of overidentifying restrictions in GMM estimation. The null

[19] In robustness checks we have studied the role of imposing short sales constraint on the portfolio optimization problem. We find that short sale constraints help to increase the out-of-sample Sharpe ratio of the portfolios by reducing sampling variability and estimation error in the portfolio problem. Detailed results are available from Author(s) upon request.

hypothesis of the test is that the moment conditions included in GMM estimation are valid. We also report the in-sample root mean square error (RMSE)[20]. As a measure of performance of the portfolios, we discuss the ex-post realized OOS Sharpe ratio (OOS-SR_{TC}), which is commonly used by academics and practitioners alike for performance evaluation purposes (e.g., see De Miguel et al., 2009). The subscript TC refers to the fact portfolio returns are adjusted by transaction costs. Following De Miguel et al. (2009), we assume proportional transaction costs of 50 basis points for each trade during portfolio rebalancing. We also report a portfolio turnover measure defined as the average sum of the trades across the assets in the portfolio.

The p-values of the test of difference in Sharpe ratios are reported in a triangular matrix in the last columns of each panel. The first column reports the p-values comparing the CAPM with the other models, the second column reports the p-values comparing the FF3 factor model with the other models, etc. The null hypothesis of the test is $H_0 : SR_1 - SR_2 = 0$, and the alternative hypothesis is $H_1 : SR_1 - SR_2 \neq 0$ for any two portfolios. Finally, at the bottom of each table, we report OOS-SR for three benchmark models: the equally-weighted portfolio ($1/N$), the mean-variance tangency portfolio (MV) and the global minimum variance portfolio (MinV). The last two models are computed using historical returns.[21]

4.2 Main results

In Table 1, we report estimates for the case of 10 industry portfolios. Looking at the in-sample performance measures in the first panel, we observe that, in general, multifactor models outperform the CAPM, as the 4-factor Liquidity model is the model producing the

[20]Similar results are obtained using the Mean Absolute Error (MAE). We do not report those results to save space.

[21]The weights of the mean-variance tangency portfolio are given by $w^T = \widehat{\Sigma}^{-1}\widehat{\mu}/(\imath'\widehat{\Sigma}^{-1}\widehat{\mu})$ and the weights of the global minimum variance portfolio are given by $w^{GMVP} = \widehat{\Sigma}^{-1}\imath/(\imath'\widehat{\Sigma}^{-1}\widehat{\mu})$, where $\widehat{\Sigma}$ is the sample $(N \times N)$ covariance matrix of excess returns, $\widehat{\mu}$ is the N-vector of sample mean excess returns and $\imath$ is a N-vector of ones.

best fit. For instance, the CAPM implies a GLS R^2 of 0.14 to be contrasted with 0.58 for both the 4-factor Liquidity model and the 4-factor Momentum model , and 0.57 for the 5-factor model. This result is consistent with the bulk of the asset pricing literature. The results concerning the OOS performance measures are however more intriguing. In the case of the 10 industry portfolios, we observe a similar pattern as for the in-sample performance measures: multifactor models outperform the CAPM by producing higher realized OOS Sharpe ratios. In the first panel, the estimated OOS-SR are 0.030, 0.127, 0.156, 0.158, 0.132 and 0.116 for the CAPM, the FF3 model, the Carhart model, the liquidity model, the HXZ model and the 5-factor model respectively. The estimated differences between the CAPM and the Sharpe ratios of the multifactor models are about 12 percent and are strongly statistically significant. The difference between the 3-factor and the 4-factor models does not appear to be statistically significant, though. The 5-factor model is dominated in terms of OOS-SR by both the 4-factor liquidity model and the 4-factor momentum model. When the performance of the models is compared with the benchmark portfolios, we observe that most of the multifactor models produces an OOS-SR higher than the $1/N$ portfolio (0.115) , the historical tangency MV portfolios (0.097) and the minimum variance portfolio (0.13). This evidence supports the idea that in some cases, formally incorporating linear factor models for the SDF in the asset allocation problem—hence enforcing from the start the absence of arbitrage opportunities—produces a better performance than traditional portfolio models. Across the remaining three panels in Table 1, we find results similar to the ones found in the first panel. Hence we conclude that our results are robust to the use of first and second stage GMM estimators and uncentered and centered SDF normalizations.

In Table 2, we report estimates for the case of the 17 industry portfolios. These results are qualitatively the same as those obtained in the case of 10 industry portfolios. We note again that multifactor models outperform the CAPM, both in sample and OOS. For example, in the first panel, the GLS R^2 are 0.07, 0.27, 0.36, 0.32, 0.32 and 0.34 for the CAPM, the FF3 model, the 4-factor momentum model, the 4-factor liquidity model, and the 5-factor model, respectively. The OOS-SR are 0.033, 0.131, 0.125, 0.105, 0.130 and 0.108, respectively. Again, the differences between multifactor models and the CAPM are positive (around 9 percent) and significant in most of the cases. As before, we do not find statistical significant differences between the multifactor models at standard confidence levels. When we examine the remaining three panels, we find similar results. However, it is worth mentioning that in the case of the uncentered SDF applied to two-stage estimators in panel three, we find some significant differences between multifactor models: the 4-factor liquidity model outperforms both the 4-factor momentum model and the 5-factor FF model. The latter also produces a higher OOS-SR than the 4-factor model of Hou, Xue and Zhang (2015). Therefore how finely the CRSP universe stock return data are disaggregated in terms of industries does not seem to affect the superior ability of relatively rich linear SDF recommended in the asset pricing literature to yield portfolio weights that outperform either the CAPM or classical benchmarks that have shaker rooting in no-arbitrage pricing models.

In Table 3, we report estimates for the case of the 25 size- and value- (double) sorted portfolios. The results become slightly weaker in terms of the OOS-SR point estimates and significance across models. However, the pattern remains qualitatively the same in the sense that multifactor models outperform the CAPM both in-sample and OOS. For example, in the first panel, the GLS R^2 are now only 0.04, 0.12, 0.16, 0.16, 0.16 and 0.21, whereas the OOS-SR are 0.005, 0.103, 0.127, 0.094, 0.127 and 0.137 for the CAMP, FF3 model, 4-factor models, and the 5-factor model of Fama and French (2015). Just to mention them at least once, the RMSEs are 0.41, 0.22, 0.19, 0.20, 0.18, and 0.16 and they clearly decline as one increases the number of factors K to include also SMB, HML, momentum, RMW and CMA as additional factors in the SDF.

When we use the level of turnover to evaluate how costly the implementation of the portfolios is for different factor models, we find a similar pattern than the one given by the OOS-SR across test assets: the level of turnover for the CAPM is considerable higher than the one of multifactor models. This implies that not only the multifactor models outperform the CAPM by delivering higher risk-adjusted out-of-sample returns, but also by involving much less trades triggered by portfolio rebalancing. Less clear is the comparison among multifactor models where turnover estimates are of similar magnitude. Overall, it seems that the higher is the number of factors in the model, the lower is the portfolio turnover. In comparison with Benchmarks models, for this set of test assets, we find that only in one case, the 5-factor model in the second panel, the performance of the portfolio backed by a linear factor model is superior to the $1/N$ portfolio (0.149 vs 0.142).

4.3 Additional results for alternative test assets

When studying the relationship between the in-sample and the out-of-sample performance of the portfolios backed by the models, one concern is whether the analyzed linear factor models are good asset pricing models in the first place or not. In other words, whether the CAPM and the multifactor models are able to reasonable explain the cross-section of stock returns. In this regard, the results presented so far show that the vast majority of the analyzed models are valid in fitting the cross-section of stock returns, when the 10 and 17 industries portfolios and the 25 size-book-to-market portfolios are used as test assets. According to the GMM´s J-test of overidentifying restrictions, most of the models are valid at the 5 percent of significance. In this sense, the reported out-of-sample results so far are conditional to using a valid model, at least under this metric.

What happen if the considered linear asset pricing model is not valid then? Would they still be able to produce portfolios exploiting potential differences among models? To address this concern, we estimate the models considering two additional test assets: 49 industries portfolios and 25 double-sorted size and momentum portfolios. We select these two additional test assets because it is well known that standard linear factor asset pricing models show limited ability to fit large dimensional (industry) portfolios and portfolios capturing momentum.[22]

In Table 4, we report our estimates considering as test asset 49 industries portfolios, and in Table 5, we report our estimates considering as test assets 25 size-mometum portfolios. Consistent with prior literature studying the in-sample fit of linear asset pricing model, we find that the considered models show a poor fit as compared with the previous results. For the case of 49 industries portfolios, the J-Test's null hypothesis of valid moment conditions is rejected in all the cases; whereas for the case of 25 double-sorted size and momentum portfolios only the 4-factor liquidity model and the 5-factor model are marginally valid as the null hypothesis is not rejected at the 5 percent of significance. When we look at the OOS-SR performance of the models, the documented differences between the CAPM and multifactor models are reduced significantly or they disappear. Thus, this evidence seems to be consistent with the idea that the use of linear factor asset pricing models to build portfolios delivering significant risk-adjusted returns is conditional to the ability of the model to fit the cross-section of returns reasonably well in a first stage.

5 Conclusions

This paper evaluates linear stochastic discount factor models based on the out-of-sample realized performance of MV efficient portfolios backed by the models. In particular, we test whether the well-documented superior ability of multifactor models to fit the cross-section of expected returns in sample over the CAPM survives the test of well-crafted OOS tests based on a commonly employed portfolio metric. Moreover, we test whether there is any connection or even correlation between the in-sample statistical performance of the SDF linear factor models and their OOS portfolio performance. Our methodology consists of two steps. In the first stage, the linear factor models are estimated under their SDF representation by

[22] We thank a referee for suggesting this analysis.

GMM, as described by Cochrane (2009). In the second step, we use the predicted excess returns from the estimated models to compute MV efficient portfolios using the concept of mean-representing portfolio introduced by Chamberlain and Rothschild (1983). As it is shown by Peñaranda and Sentana (2011, 2012), the mean-representing portfolio delivers a set of moments that allow us to consistently estimate by GMM a (no arbitrage) MV frontier. Finally, we compute realized OOS Sharpe ratios based on the recursive portfolio weights obtained from each model.

Using several samples of test assets consisting of monthly US portfolio equity returns spanning the period 1968-2016, 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 OOS. Besides, we document that there is an empirical link between the in-sample statistical performance and the OOS performance of linear factor SDF models: the models exhibiting the best in-sample performance are also the models with the best OOS one. This result is consistent with the idea that asset pricing models provide useful information to an investor solving a MV problem and that the standard in-sample fit recorded in the literature contains reliable information on the underlying DGP driving the SDF. We also report that multifactor models outperform the CAPM yielding monthly OOS Sharpe ratios that are higher by as much as 10 per cent. These results are robust to extending the exercise to alternative asset menus, to adopting first- vs. two-step GMM estimators, and to centered and uncentered SDF specifications.

We left for future research to explore the economic forces driving the documented connection between the in-sample and the OOS performance of linear factor models. In this regard, Morana (2014) provides an interesting insight by showing that the SMB, HML, MOM and LIQ factors reflect compensation for macroeconomic and financial risk. Thus, the link between factors and macroeconomic conditions may potentially account for the documented correlation between in-sample and OOS performance measures. Additionally, our key result may be considered natural when the subspace of returns spanned by the multi-factor models generates more efficient mean-variance portfolios than the returns on the market alone, which is the subspace generated by the CAPM, and this fact gets in principle stronger, the larger the number of factors driving the dynamics of cross section of returns not spanned by the market portfolio, that multi-factor models correctly capture. When in the paper we have expanded the size of the test asset menus, the evidence has failed to reveal that as the number of factors grow, the OOS performances of richer SDF models improves. It is not clear why this may occur, even though one cannot rule out that the samples may still be too short to reveal the true data generating process governing the SDF and/or that the true but unknown SDF governing the cross section of US equity returns may be of a nonlinear type, for instance as in Dittmar, 2002). Therefore it would be interesting to explicitly map our results, say a notion of a “ratio” between in-sample RMSE and OOS Sharpe ratios improvements vs. the CAPM to formal spanning tests, along the lines of Peñaranda and Sentana (2012).

References

[1] Abhakorn, P., Smith, P. N., and Wickens, M. R. What do the fama–french factors add to c-capm? *Journal of Empirical Finance 22* (2013), 113–127.

[2] Andrews, D., and Monahan, J. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica: Journal of the Econometric Society* (1992), 953–966.

[3] Behr, P., Guettler, A., and Truebenbach, F. Using industry momentum to improve portfolio performance. *Journal of Banking & Finance* (2012).

[4] Brandt, M. Portfolio choice problems. *Handbook of financial econometrics 1* (2010), 269–336.

[5] Brandt, M. W. Estimating portfolio and consumption choice: A conditional euler equations approach. *The Journal of Finance 54*, 5 (1999), 1609–1645.

[6] Britten-Jones, M. The sampling error in estimates of mean-variance efficient portfolio weights. *The Journal of Finance 54*, 2 (1999), 655–671.

[7] Carhart, M. On persistence in mutual fund performance. *Journal of Finance* (1997), 57–82.

[8] Chamberlain, G., and Rothschild, M. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica: Journal of the Econometric Society* (1983), 1281–1304.

[9] Chevrier, T., and McCulloch, R. Using economic theory to build optimal portfolios. *Available at SSRN 1126596* (2008).

[10] Cochrane, J. H. *Asset Pricing:(Revised Edition).* Princeton university press, 2009.

[11] Connor, G., and Korajczyk, R. Factor models in portfolio and asset pricing theory. *Handbook of Portfolio Construction* (2010), 401–418.

[12] Daniel, K., and Titman, S. Evidence on the characteristics of cross sectional variation in stock returns. *Journal of Finance 52*, 1 (1997).

[13] 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.

[14] DeMiguel, V., Nogales, F. J., and Uppal, R. Stock return serial dependence and out-of-sample portfolio performance. *Review of Financial Studies 27*, 4 (2014), 1031–1073.

[15] Fama, E., and French, K. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*, 1 (1993), 3–56.

[16] Fama, E. F., and French, K. R. A five-factor asset pricing model. *Journal of Financial Economics 116*, 1 (2015), 1 – 22.

[17] 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.

[18] Hou, K., Xue, C., and Zhang, L. Digesting anomalies: An investment approach. *The Review of Financial Studies 28*, 3 (2015), 650–705.

[19] Jagannathan, R., and Ma, T. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance 58*, 4 (2003), 1651–1684.

[20] Jagannathan, R., and Wang, Z. Empirical evaluation of asset pricing models: A comparison of the sdf and beta methods. *Journal of Finance 57*, 5 (2002).

[21] Jobson, J., and Korkie, B. Performance hypothesis testing with the sharpe and treynor measures. *Journal of Finance 36* (1981), 889–908.

[22] Kan, R., Robotti, C., and Shanken, J. Pricing model performance and the two-pass cross-sectional regression methodology. *The Journal of Finance 68*, 6 (2013), 2617–2649.

[23] Kirby, C., and Ostdiek, B. Its all in the timing: Simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis 47*, 2 (2012), 437–467.

[24] Ledoit, O., and Wolf, M. Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance 15*, 5 (2008), 850–859.

[25] Lewellen, J., Nagel, S., and Shanken, J. A skeptical appraisal of asset pricing tests. *Journal of Financial Economics 96*, 2 (2010), 175–194.

[26] Lintner, J. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics 47*, 1 (1965), 13–37.

[27] Lo, A. The statistics of sharpe ratios. *Financial Analysts Journal* (2002), 36–52.

[28] 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.

[29] MacKinlay, A., and Pástor, L. Asset pricing models: Implications for expected returns and portfolio selection. *Review of Financial Studies 13*, 4 (2000), 883–916.

[30] Markowitz, H. Portfolio selection. *Journal of Finance 7*, 1 (1952), 77–91.

[31] Memmel, C. Performance hypothesis testing with the sharpe ratio. *Finance Letters 1* (2003), 21–23.

[32] Morana, C. Insights on the global macro-finance interface: Structural sources of risk factor fluctuations and the cross-section of expected stock returns. *Journal of Empirical Finance 29* (2014), 64–79.

[33] Opdyke, J. Comparing sharpe ratios: So where are the p-values? *Journal of Asset Management 8*, 5 (2007), 308–336.

[34] Pástor, L. Portfolio selection and asset pricing models. *Journal of Finance 55*, 1 (2000), 179–223.

[35] Pastor, L., and Stambaugh, R. Comparing asset pricing models: an investment perspective. *Journal of Financial Economics 56*, 3 (2000), 335–381.

[36] Pastor, U., and Stambaugh, R. Liquidity risk and expected stock returns. *Journal of Political Economy 111*, 3 (2003), 642–685.

[37] Peñaranda, F., and Sentana, E. Inferences about portfolio and stochastic discount factor mean variance frontiers. *Working Paper Cemfi* (2011).

[38] Peñaranda, F., and Sentana, E. Spanning tests in return and stochastic discount factor mean–variance frontiers: A unifying approach. *Journal of Econometrics 170*, 2 (2012), 303–324.

[39] Politis, D., and Romano, J. A circular block-resampling procedure for stationary data. *Exploring the Limits of Bootstrap* (1992), 263–270.

[40] Romano, J., and Wolf, M. Improved nonparametric confidence intervals in time series regressions. *Journal of Nonparametric Statistics 18*, 2 (2006), 199–214.

[41] Schwert, G. Anomalies and market efficiency. *Handbook of the Economics of Finance 1* (2003), 939–974.

[42] 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.

[43] Sharpe, W. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance 19*, 3 (1964), 425–442.

[44] 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.

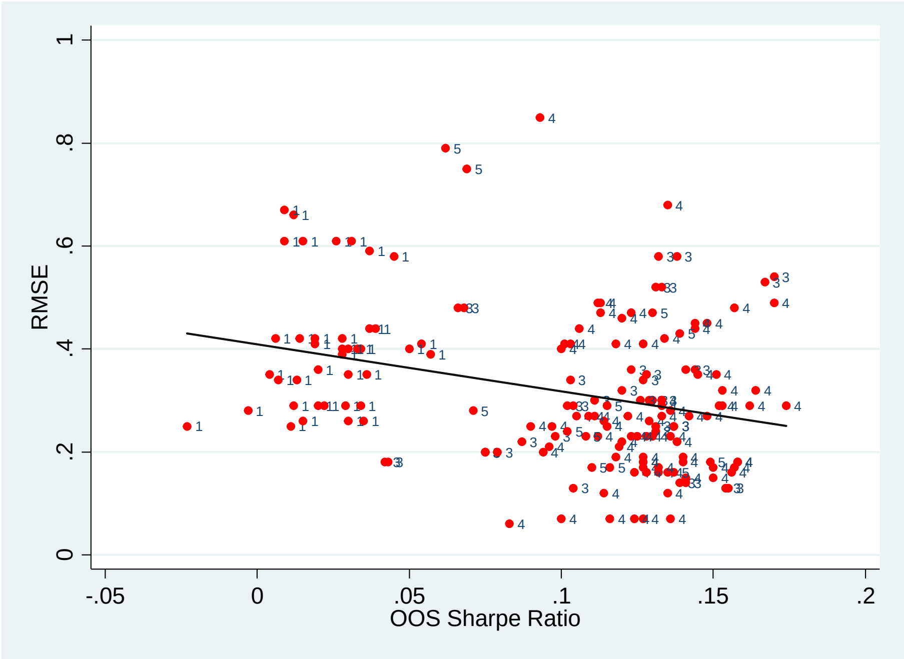

Figure 1: Scatter plot of RMSE vs. out-of-sample Sharpe ratios.

$$Y = \underset{(0.03)}{0.33} - \underset{(0.23)}{0.91}X$$

Each dot represents a particular model and the number attached to it indicates the number of factors included in that model.

Table 1: In-Sample Asset Pricing Model Evaluation and Out-of-Sample Portfolio Performance

(Test Assets: 10 Industries)

First Stage Estimators										
Uncentered SDF ($a = 1$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.14	0.38	0.36	0.026	7.63					
3	0.38	0.49	0.23	0.127	0.67	0.17				
4M	0.58	0.66	0.16	0.156	0.50	0.03**	0.18			
4L	0.58	0.57	0.18	0.158	0.49	0.03**	0.15	0.99		
4H	0.53	0.64	0.17	0.132	0.54	0.08*	0.63	0.25	0.22	
5	0.57	0.56	0.17	0.116	0.62	0.29	0.46	0.02**	0.06*	0.20
Centered SDF ($a = 1 - \lambda' E(f)$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.14	0.38	0.37	0.019	13.57					
3	0.38	0.49	0.24	0.131	0.70	0.00***				
4M	0.58	0.65	0.18	0.140	0.52	0.00***	0.60			
4L	0.58	0.58	0.18	0.158	0.53	0.01***	0.43	0.60		
4H	0.54	0.62	0.19	0.118	0.57	0.00***	0.86	0.71	0.39	
5	0.57	0.57	0.17	0.110	0.64	0.00***	0.58	0.32	0.17	0.32
Second Stage Estimators										
Uncentered SDF ($a = 1$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.14	0.35	0.43	0.046	4.88					
3	0.38	0.50	0.34	0.141	0.69	0.23				
4M	0.58	0.67	0.27	0.148	0.62	0.11	0.77			
4L	0.58	0.57	0.29	0.162	0.50	0.10*	0.59	0.70		
4H	0.54	0.64	0.27	0.111	0.57	0.55	0.33	0.13	0.21	
5	0.57	0.56	0.28	0.071	0.75	0.73	0.15	0.15	0.15	0.63
Centered SDF ($a = 1 - \lambda' E(f)$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.14	0.38	0.42	-0.012	6.04					
3	0.38	0.49	0.34	0.135	0.77	0.00***				
4M	0.58	0.65	0.29	0.153	0.63	0.00***	0.61			
4L	0.58	0.58	0.29	0.174	0.53	0.01***	0.43	0.65		
4H	0.54	0.52	0.30	0.133	0.67	0.00***	0.84	0.72	0.41	
5	0.57	0.57	0.29	0.115	0.61	0.01***	0.57	0.32	0.17	0.31
Benchmarks										
	1/N	MV	MinV							
OOS SR	0.136	0.109	0.136							

The table reports In-sample measures of goodness of fit for a set of linear asset pricing models and Out-of-Sample portfolio performance backed by the same set of models. The models consider are the CAPM ($K = 1$), the 3-factor ($K = 3$) model of Fama-French (1993), the 4-factor ($K = 4M$) model of Carhart (1997), the 4-factor ($K = 4L$) model of Pastor and Stambaugh (2003), the 4-factor ($K = 4H$) model of Hou, Xue and Zhang (2015) and the five-factor ($K = 5$) model of Fama and French (2015). The in-sample measures reported are the R^2 of the estimated linear factor model by Generalized Least Squares under its beta representation (see Lewellen et al., 2010), the p-value of the GMM's overidentification test (J-test), and the root mean squared error (RMSE). The out-of-sample portfolio performace measures are the Sharpe

Ratio, taking into account transaction costs, and portfolio turnover. Finally, we report a matrix containig 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%. 1/N is the equally weighted portfolio, MV is the historical mean-variance tangent portfolio and MinV is the historical global minimum variance portfolio. The OOS-SR of the benchmark models are also computed within a rolling window setup for comparability purposes. First and second stage estimators refer to weighted matrix used in the GMM estimation. Uncentered and centered SDF differ in the normalization imposes on the constant term in the linear SDF model.

Table 2: In-Sample Asset Pricing Model Evaluation and Out-of-Sample Portfolio Performance

(Test Assets: 17 Industries)

First Stage Estimators										
Uncentered SDF ($a = 1$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.07	0.33	0.44	0.033	6.82					
3	0.22	0.38	0.32	0.131	0.73	0.05**				
4M	0.36	0.52	0.23	0.125	0.68	0.03**	0.59			
4L	0.32	0.44	0.27	0.105	0.66	0.23	0.40	0.29		
4H	0.32	0.55	0.23	0.130	0.56	0.06*	0.99	0.59	0.43	
5	0.34	0.52	0.23	0.108	0.52	0.21	0.40	0.25	0.95	0.38
Centered SDF ($a = 1 - \lambda' E(f)$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.07	0.32	0.45	0.022	3.66					
3	0.22	0.37	0.32	0.126	0.76	0.28				
4M	0.36	0.51	0.25	0.115	0.68	0.33	0.69			
4L	0.32	0.44	0.27	0.109	0.62	0.06*	0.19	0.36		
4H	0.32	0.50	0.26	0.114	0.60	0.02**	0.14	0.20	0.81	
5	0.34	0.49	0.24	0.102	0.59	0.05**	0.21	0.35	0.92	0.66
Second Stage Estimators										
Uncentered SDF ($a = 1$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.07	0.33	0.67	0.049	1.99					
3	0.22	0.38	0.57	0.159	0.52	0.08*				
4M	0.36	0.52	0.46	0.120	0.59	0.20	0.30			
4L	0.32	0.44	0.49	0.170	0.52	0.03**	0.68	0.09*		
4H	0.32	0.55	0.46	0.056	0.81	0.57	0.39	0.32	0.09*	
5	0.34	0.52	0.47	0.130	0.44	0.26	0.39	0.98	0.05**	0.41
Centered SDF ($a = 1 - \lambda' E(f)$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.07	0.32	0.53	0.028	18.31					
3	0.22	0.37	0.51	0.036	0.89	0.28				
4M	0.36	0.51	0.44	0.106	0.66	0.34	0.67			
4L	0.32	0.44	0.45	0.144	0.59	0.07*	0.18	0.36		
4H	0.32	0.50	0.48	0.157	0.70	0.04**	0.15	0.17	0.78	
5	0.34	0.49	0.43	0.139	0.55	0.07*	0.22	0.38	0.90	0.66
Benchmarks										
	1/N	MV	MinV							
OOS SR	0.142	0.122	0.141							

See notes in table 1.

Table 3: In-Sample Asset Pricing Model Evaluation and Out-of-Sample Portfolio Performance

(Test Assets: 25 Size / Boolk-to-Market)

First Stage Estimators										
Uncentered SDF ($a = 1$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.04	0.01	0.41	0.005	8.12					
3	0.12	0.06	0.22	0.103	0.66	0.10*				
4M	0.16	0.10	0.19	0.127	0.53	0.11	0.91			
4L	0.16	0.07	0.20	0.094	0.46	0.19	0.25	0.29		
4H	0.16	0.06	0.18	0.127	0.44	0.09*	0.97	0.94	0.25	
5	0.21	0.07	0.16	0.137	0.41	0.08*	0.77	0.68	0.16	0.55
Centered SDF ($a = 1 - \lambda' E(f)$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.04	0.01	0.42	0.033	4.36					
3	0.12	0.05	0.23	0.097	0.62	0.24				
4M	0.16	0.07	0.21	0.096	0.53	0.14	0.45			
4L	0.16	0.06	0.22	0.070	0.55	0.22	0.68	0.74		
4H	0.16	0.06	0.22	0.120	0.56	0.27	0.93	0.57	0.73	
5	0.21	0.08	0.18	0.149	0.44	0.18	0.33	0.88	0.68	0.44
Second Stage Estimators										
Uncentered SDF ($a = 1$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.04	0.01	1.31	0.056	1.34					
3	0.12	0.06	1.05	0.059	0.29	0.74				
4M	0.16	0.10	0.85	0.093	0.41	0.26	0.20			
4L	0.16	0.07	0.95	0.043	0.44	0.81	0.62	0.18		
4H	0.16	0.06	0.68	0.135	0.59	0.06*	0.01***	0.16	0.11	
5	0.21	0.07	0.75	0.069	0.43	0.32	0.10*	0.73	0.30	0.09*
Centered SDF ($a = 1 - \lambda' E(f)$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.04	0.01	1.18	-0.003	1.69					
3	0.12	0.05	0.93	0.042	0.39	0.27				
4M	0.16	0.07	0.91	0.056	0.60	0.14	0.43			
4L	0.16	0.06	0.82	0.049	0.62	0.22	0.71	0.70		
4H	0.16	0.06	0.78	0.041	0.62	0.26	0.95	0.54	0.67	
5	0.21	0.08	0.79	0.062	0.65	0.17	0.33	0.88	0.64	0.43
Benchmarks										
	1/N	MV	MinV							
OOS SR	0.142	0.122	0.142							

See notes in table 1.

Table 4: In-Sample Asset Pricing Model Evaluation and Out-of-Sample Portfolio Performance

(Test Assets: 49 Industries)

First Stage Estimators										
Uncentered SDF ($a = 1$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.01	0.00	0.63	-0.004	2.73					
3	0.03	0.00	0.49	0.011	1.19	0.60				
4M	0.06	0.00	0.41	0.065	0.68	0.20	0.34			
4L	0.05	0.00	0.46	0.104	0.74	0.15	0.30	0.54		
4H	0.05	0.00	0.41	0.050	0.77	0.40	0.55	0.37	0.13	
5	0.06	0.00	0.41	0.067	0.61	0.27	0.39	0.59	0.26	0.45
Centered SDF ($a = 1 - \lambda' E(f)$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.01	0.00	0.64	-0.011	2.55					
3	0.03	0.00	0.49	0.112	0.97	0.04**				
4M	0.05	0.00	0.45	0.053	0.75	0.26	0.06			
4L	0.05	0.00	0.46	0.103	0.73	0.46	0.11	0.55		
4H	0.05	0.00	0.46	0.029	1.10	0.23	0.40	0.78	0.50	
5	0.06	0.00	0.43	0.072	0.66	0.17	0.43	0.63	0.36	0.83
Second Stage Estimators										
Uncentered SDF ($a = 1$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.01	0.00	3.31	0.021	2.27					
3	0.03	0.00	1.94	0.055	0.54	0.54				
4M	0.06	0.00	1.64	0.043	0.57	0.90	0.43			
4L	0.05	0.00	1.87	0.092	0.42	0.26	0.32	0.14		
4H	0.05	0.00	1.37	0.037	0.64	0.86	0.65	0.93	0.27	
5	0.06	0.00	1.58	0.053	0.58	0.46	0.90	0.42	0.33	0.58
Centered SDF ($a = 1 - \lambda' E(f)$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.01	0.00	3.04	0.090	2.32					
3	0.03	0.00	1.66	0.022	0.84	0.04**				
4M	0.06	0.00	1.55	0.066	0.83	0.23	0.05**			
4L	0.05	0.00	1.51	0.085	0.74	0.45	0.12	0.61		
4H	0.05	0.00	1.53	0.057	0.85	0.24	0.37	0.75	0.51	
5	0.06	0.00	1.40	0.048	1.15	0.17	0.41	0.61	0.36	0.88
Benchmarks										
	1/N	MV	MinV							
OOS SR	0.126	0.106	0.124							

See notes in Table 1.

Table 5: In-Sample Asset Pricing Model Evaluation and Out-of-Sample Portfolio Performance

(Test Assets: 25 Size/Momentum)

First Stage Estimators										
Uncentered SDF ($a = 1$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.03	0.00	0.49	0.112	0.85					
3	0.11	0.01	0.32	0.121	0.62	0.47				
4M	0.15	0.03	0.24	0.126	0.42	0.48	0.97			
4L	0.14	0.07	0.29	0.066	0.57	0.78	0.14	0.17		
4H	0.13	0.03	0.23	0.128	0.40	0.42	0.95	0.87	0.17	
5	0.19	0.15	0.20	0.104	0.44	0.63	0.37	0.39	0.25	0.27
Centered SDF ($a = 1 - \lambda' E(f)$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.03	0.00	0.50	0.109	0.85					
3	0.11	0.01	0.33	0.134	0.65	0.87				
4M	0.15	0.02	0.26	0.118	0.38	0.36	0.04**			
4L	0.13	0.07	0.30	0.072	0.60	0.75	0.80	0.04**		
4H	0.13	0.02	0.27	0.099	0.52	0.60	0.27	0.58	0.19	
5	0.19	0.08	0.25	0.064	0.49	0.94	0.92	0.44	0.84	0.62
Second Stage Estimators										
Uncentered SDF ($a = 1$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.03	0.00	1.69	0.016	4.33					
3	0.11	0.01	1.12	0.060	0.44	0.72				
4M	0.15	0.03	0.98	0.062	0.54	0.58	0.80			
4L	0.14	0.07	0.99	0.049	0.57	0.83	0.67	0.63		
4H	0.13	0.03	0.85	0.068	0.49	0.25	0.23	0.35	0.21	
5	0.19	0.15	0.77	0.062	0.56	0.66	0.81	0.89	0.55	0.34
Centered SDF ($a = 1 - \lambda' E(f)$)										
K	R^2 GLS	J-Test	RMSE	OOS SR_{TC}	Turnover	SR Diff.Test (p-value)				
1	0.03	0.00	1.47	0.036	1.70					
3	0.11	0.01	1.01	0.075	0.55	0.85				
4M	0.15	0.02	1.04	0.047	0.77	0.37	0.04**			
4L	0.13	0.07	0.84	0.079	0.85	0.75	0.79	0.05**		
4H	0.13	0.02	1.04	0.056	0.72	0.62	0.27	0.57	0.23	
5	0.19	0.08	0.85	0.069	0.69	0.93	0.92	0.41	0.85	0.63
Benchmarks										
	1/N	MV	MinV							
OOS SR	0.129	0.104	0.128							

See notes in Table 1.