Empirical asset pricing with many test assets

Rasmus Lönn[†] and Peter C. Schotman[‡]

This draft: Dec 19, 2022
First draft: Oct 31, 2018
Revised: August 22, 2022

Abstract: We reformulate the problem of estimating risk prices in a stochastic discount factor model as an instrumental variables regression. The IV estimator allows efficient estimation for models with non-traded factors and many test assets. We implement L_2Boosting as a regularised regression technique to select optimal instruments. Optimal instruments emerge as tracking portfolios. In a simulation study the IV estimator is close to the infeasible GMM estimator in a setting with short time series and many assets. In an empirical application the tracking portfolio for consumption growth appears strongly correlated with consumption news. It implies that consumption is a priced factor for the cross-section of excess equity returns. Other factors such as inflation are tracked by asset returns but do not improve the asset pricing properties of the model. A similar regularised regression, projecting the stochastic discount factor on test assets, leads to an estimate of the Hansen-Jagannathan distance.

Keywords: L_2Boosting; Asset Pricing Tests; Hansen-Jagannathan Distance; Instrumental Variables

JEL codes: G12 (Asset Pricing); C44 (Statistical Decision Theory); C55 (Large Data Sets)

[†] Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands, lonn@ese.eur.nl.

[‡] School of Business and Economics, Maastricht University, The Netherlands, p.schotman@maastrichtuniversity.nl

[§] We thank conference participants at SoFiE (New York), ESEM (Lisbon) and the Conference on the Econometrics of Financial Markets (Stockholm), ESWC (Bocconi), IAAE (Rotterdam) for helpful comments and suggestions and seminar participants at Erasmus University Rotterdam, Lund University and Stockholm School of Economics.

Our econometric techniques are all designed for large time series and small cross sections. Our data has a large cross section and short time series. A large unsolved problem in finance is the development of appropriate large-N small-T tools for evaluating asset pricing models. (Cochrane, 2005, p 226)

1 Introduction

A standard asset pricing test evaluates whether a small number of factors can explain the differences in expected returns across a variety of test assets. Such an evaluation is statistically problematic when the number of test assets (N) is large relative to the length of the time series (T) available for the factors. The combination of large N and small T is particularly awkward for macroeconomic factors observed at low frequencies. We propose a methodology that is well-suited for such tasks.

We consider a stochastic discount factor (SDF) asset pricing model. Theory implies that the SDF should be orthogonal to the excess returns on the test assets. An overall test for the asset pricing conditions involves the N-vector of sample means of the pricing errors and its $(N \times N)$ covariance matrix. This runs into two problems. First, the test requires the inverse of the high-dimensional covariance matrix. Second, a test on all N test assets may have low power, even if a profitable trading strategy exists. Many of the test assets may be correctly priced, but carefully chosen portfolios of test assets may violate the orthogonality conditions.

We address both problems by applying machine learning methods for model selection and shrinkage. We rewrite the asset pricing test as a time series regression of the SDF on excess returns. If the model is correct, the regression should not have any explanatory power. But without regularisation that regression still suffers from the same large-N problem. We therefore impose sparsity. If the model has any value, we should not expect that many assets have large coefficients in this regression.

The fitted values of that regularised regression are excess returns on a portfolio of test assets that is most anomalous for the asset pricing model. The larger the magnitude of the fitted values, the larger the pricing errors. The average of the squared fitted values is an estimate of the Hansen-Jagannathan (HJ) distance (Hansen and Jagannathan, 1997), a well-defined metric for evaluating asset pricing models. Usually the HJ-distance cannot be computed for large N due to the need for inverting the high-dimensional second moments matrix of returns. Being able to estimate the HJ distance the evaluation of an asset pricing model can go beyond the commonly used cross-sectional R^2 (of a regression of average returns on beta's) with its well known shortcomings.

Our main contribution takes this regression approach one step further. The SDF usually depends on unknown parameters δ. For a linear SDF, $m = 1 - f'\delta$, with traded factors f, estimation comes down to a regression of the constant '1' on the factor returns,

$$1 = f'\delta + u, \tag{1}$$

For a small set of traded factors, this does not present any complications (Britten-Jones, 1999). Kozak, Nagel, and Santosh (2020) consider the case of many traded factors. We focus on the case of non-traded factors. In this case, minimising the HJ-distance is equivalent to estimating δ in the same regression model (1) but now by instrumental variables using the excess returns of the test assets as instruments. That again is a large-N problem for which we need sparsity conditions to select the optimal instruments. The optimal instruments emerge as portfolios of test assets that minimise the overall pricing error and can be interpreted as returns on mimicking portfolios for the factors. Mimicking portfolios, for which we use tracking portfolios as a synonym, contain all the essential information about the stochastic discount factor (Cochrane, 2005, ch 7). With the estimate $\hat{\delta}$ we construct $m = 1 - f'\hat{\delta}$ and proceed with the projection of m on the excess returns to identify anomalies and estimate the HJ distance as discussed before.

In short, our proposed methodology has two elements. The first is an instrumental variables regression to estimate the parameters of the asset pricing model. Although the IV regression is a straightforward implication of basic asset pricing moment conditions, it has, to the best of our knowledge, never been implemented empirically. With large N it relies on regularised regressions to construct tracking portfolios. The second element is a regularised regression to identify an anomaly portfolio.

The proposed IV estimator is an alternative to Fama-MacBeth (FM) regressions (Fama and MacBeth, 1973). In an FM regression the δ parameters are estimated through a cross-sectional OLS regression of average returns on covariances between factors and returns. It avoids the high-dimensional matrix inversion by giving equal weight to all test assets and by exploiting the time series variation for statistical tests. The IV estimator can be interpreted as a method for finding the optimal N-dimensional weighting matrix to perform a GLS regression. The IV estimator will be more efficient than FM in situations with a lot of heteroskedasticity and correlations in the cross section, which is typical in applications with macro-economic factors that can only explain part of the factor structure in returns. Balduzzi and Robotti (2008) indicate that the risk premium estimated from tracking portfolios can only agree with the FM regression if the FM regression would be done in GLS style, which is exactly what we aim to approximate by constructing the optimal instrument as a tracking portfolio.

Our approach complements Feng, Giglio, and Xiu (2020). Feng, Giglio, and Xiu (2020) estimate the risk prices δ using FM regressions, but apply machine learning to the selection of factors, allowing for a large set of candidate factors. We assume that the number of factors is small, while the number of test assets can be large. Instead of FM regressions we use the machine learning tools to select optimal instruments for efficient estimation of risk prices and to identify anomaly portfolios that are most informative on mispricing.

The first stage in the IV estimator is the construction of a tracking portfolio as an optimal instrument. Factor mimicking portfolios for non-traded factors have a long history in empirical asset pricing (Breeden, Gibbons, and Litzenberger, 1989). Lamont (2001) developed the econometric methodology for the construction of tracking portfolios for macro-economic factors. Examples of applications in an SDF asset pricing model with macro factors are Vassalou (2003) and Aretz, Bartram, and Pope (2010). Most distinguishing in our approach is the use of machine learning tools in constructing tracking portfolios from a large cross-section of assets.

Since a portfolio return is a *linear* combination of asset returns, we need regularised *linear* regressions, both to construct tracking portfolios, as well as to identify anomaly portfolios relative to the pricing model. The various model selection and shrinkage estimators in the literature differ by their implicit priors on the coefficients (Murphy, 2012). We choose to work with L_2Boosting. Bai and Ng (2009) earlier suggested L_2Boosting for instrument selection, while Belloni, Chen, Chernozhukov, and Hansen (2012) propose it as one of the methods for selecting optimal instruments. Belloni et al. (2012) also provide the statistical theory for inference for the IV estimator with optimal instrument selection.

L_2Boosting is related to Lasso (Hastie, Tibshirani, and Friedman, 2009, §16.2). Both perform simultaneous model selection and shrinkage, while increasing model complexity step by step, along different paths, until some optimal stopping time. Lasso is known to be less effective in regressions with strong multicollinearity. This is a real concern, since return data typically have a factor structure, and thus feature strong multicollinearity. Lasso and L_2Boosting differ in the form of regularisation and type of sparsity that they produce. Lasso assumes that just a few assets have non-zero weights, whereas L_2Boosting allows for many small coefficients with a bound on their sum of absolute values when N grows large (Bühlmann and Van de Geer, 2011). The latter is a natural form of sparsity in portfolio applications. If returns follow a factor model, the tracking portfolios will generally have weights on all assets with a loading on the factor. But when the number of test assets grows, the weight of each single asset decreases such that sum of (absolute values) of weight remains bounded.

Another related regularisation device is an Elastic Net, which also does simultaneous model selection and shrinkage and which is designed to address multicollinearity (Zou and Hastie, 2005). In the asset pricing literature, Kozak, Nagel, and Santosh (2020) apply an Elastic Net for portfolio construction from a large set of correlated returns. For the implementation they develop an elaborate informative prior. In contrast, L_2Boosting has minimal tuning parameters.

In a large-N setting Giglio and Xiu (2021) construct a tracking portfolio using the first few principal components of the excess returns as base assets. Their aim is very different from ours, though. Important in their application is that the base assets span the space of excess returns. The principal components serve that purpose, since they are the linear combinations of excess returns that explain as much as possible of the return covariance matrix. But the first few PCs do not necessarily provide the best fit for a factor. Especially in the second stage, when we search for anomalies relative to the fitted SDF model, the interest is not on the PCs of the excess returns, but on the portfolios of test assets that violate the pricing conditions most. Likewise, in the first step, our approach finds the linear combination of excess returns that is maximally correlated with the factor. This reduces the impact that assets with weak association to the factor have on the estimation of the risk price. Giglio, Xiu, and Zhang (2022) also stress the importance of emphasizing the assets with the strongest association to the factors. They highlight that whether a factor is strong or weak should be assessed within the context of a given set of test assets, and propose a supervised version of PCA to ensure reliable inference.

Other methods, such as ridge estimators, solely operate on the second moment (or covariance) matrix. For example, the shrinkage estimators developed by Ledoit and Wolf (2003, 2004, 2017) or estimators based on high-frequency data such as Bollerslev, Meddahi, and Nyawa (2019), would produce tracking portfolios as well as anomaly portfolios that have non-zero weights for all test assets. While we prefer L_2Boosting based on our reading of the literature, we have not empirically tested which one works best for our application. With proper calibration and tuning other methods may outperform L_2Boosting. But the simplicity of L_2Boosting provides robustness to overfitting and attenuate the skepticism that macro factors are only weakly related to equity returns.

Concerns about weak or useless instruments have been raised in, *e.g.*, Kan and Zhang (1999) and Kleibergen and Zhan (2020). Prominent macro factors such as consumption have very low correlation with financial returns. Reasons for the weak correlation are plenty and well-known (Breeden, Gibbons, and Litzenberger, 1989; Kroencke, 2017). For macroeconomic factors that only exhibit weak correlation with returns Kleibergen (2009) shows that the cross-sectional FM two-pass regression can provide misleading inference. A similar problem exists for IV estimators (Staiger and Stock, 1997). As a partial solution Belloni et al. (2012) suggest a split-sample estimator, as a large N version of ideas in Angrist and Krueger (1995). The split-sample estimates of the tracking portfolio returns produce out-of-sample fitted values for the tracking portfolios. Lacking easy solutions to rule out weak instrument problems, Kroencke (2021) argues that the best one can do is have a large enough correlation between factor and excess returns. In our empirical application the out-of-sample correlation between consumption and its tracking portfolio is 0.45.

The optimal instrument selection may not solve all weak instrument problems, but will not be worse than the FM two-pass estimator in the 'small T, large N' setting. By selecting the excess returns most correlated with the factor we avoid one of the problem cases in Kleibergen (2009), where only a finite number of useful instruments is available along with a large number of useless assets. Adding all, including the many useless assets with a zero beta, in an OLS cross-sectional regression, leads to a strong bias. Our model selection alleviates the problem, since the useless assets will be ignored in the tracking portfolio. The tracking portfolio aggregates the returns that are most highly correlated with the factors. As Bryzgalova (2016) we put more weight on instruments that correlate stronger with a factor.

The first stage model selection also serves as a form of pre-test. Risk price estimates are essentially the sample mean of the tracking portfolio scaled by the covariance of the tracking portfolio returns with the factor. If tracking portfolios have a zero mean, or average returns not significantly different from zero, the price of risk will also be zero, except in the problematic 'zero divide by zero' useless factor case that the covariance is zero as well. In the empirical analysis we find a reliably non-zero mean for the consumption mimicking portfolio, but not for some of the other macro factors. Tracking portfolios with close to zero average returns are most subject to the weak instrument problems. For our data, the results for consumption appear insensitive to their inclusion.

One of the suggestions in Kleibergen and Zhan (2020) is to conduct robust inference by inverting a test statistic of the model fit in order to find the risk prices that would be consistent with the SDF model. In our case it would mean finding δ such that the HJ distance is bounded by a critical value under correct specification, or otherwise conclude that an admissible δ does not exist.[1] But the entire motivation of our approach stems from

[1] Kleibergen and Zhan (2020) propose the Anderson-Rubin test statistic, which in our model is a scaled version of the Hansen-Jagannathan distance. For large N the AR statistic is infeasible, but a version derived in Belloni et al. (2012) is valid for large N.

allowing pricing errors and thus not assuming that the population value of HJ equals zero. Estimating δ is one goal, but identifying anomaly portfolios from a large set of test assets is the main motivation for applying a statistical learning algorithm.

Before delving into the empirics, we conduct an extensive Monte Carlo study. In a setting that is typical for a macroeconomic factor, *i.e.* low correlation with returns and a strong factor structure not explained by the factor, the IV estimator performs as one would expect from its asymptotic properties. With data resembling T equal to 10 years of monthly data and $N = 200$ test assets, the IV estimator appears almost unbiased and has a standard error that is less than half that of the two-pass FM estimator. When the true HJ distance equals zero, the L_2Boosting algorithm will correctly set it to zero in the majority of cases. With the same N and T the sampling distribution under the null of correct pricing has a 60% probability of an exact zero. When there is mispricing, the sampling distribution of the HJ distance quickly becomes median unbiased. Estimation uncertainty decreases with $\sqrt{T}$, but is more or less constant in N. The mispricing hardly affects risk price estimates. The sampling distribution for δ, and also its asymptotic standard error, are robust with respect to misspecification in the form of omitted factors. The latter is similar to Giglio and Xiu (2021). Finally, adding substantial noise to the factors to approximate a weak factor, the IV estimator becomes biased. But it still performs better than the FM two-pass estimator.

We apply the estimator to revisit asset pricing with four well-known macroeconomic variables: consumption, inflation, term spread and credit spread. As test assets we select 80 managed portfolio based on anomalies and industry sorts. The industry portfolios have very different business cycle exposures and thus provide an interesting set of assets for constructing different macroeconomic tracking portfolios. We would expect that the algorithm heavily loads on some of the industries for constructing tracking portfolios, but avoids them when constructing a portfolio of mispriced assets. The anomalies may be redundant return series that just complicate the task for the statistical learning algorithm in constructing a tracking portfolio. Conversely, they may be important for the mispricing portfolio. That corresponds with what we see in the empirical results.

The tracking portfolio for consumption growth loads heavily on various industries, has a market beta of one, and a Sharpe ratio close to that of the market portfolio. Still it differs substantially from the market portfolio, while only half of its variation can be explained by the five Fama-French factors. Using the optimal tracking portfolio as an instrument, the risk price of the consumption factor is statistically significant. This result is robust across model specifications. This stands in contrast to the other macroeconomic tracking portfolios. The inflation tracking portfolio has a similar out-of-sample fit as the consumption mimicking portfolio. However, the portfolio returns do not reveal a sizeable risk premium.

Using the IV estimates, the macroeconomic models produce average pricing errors of similar magnitude as the five Fama-French factors. The pricing errors stem from a wide variety of anomalies and industry sorts. The largest pricing errors are due to the anomaly test assets, in particular *industry relative reversals*. None of the models is able to accurately price these assets. Pricing errors for consumption model differ from those resulting from the Fama-French five factor models, most notably due to the *asset growth* anomaly.

The remainder of the paper is structured as follows. Section 2 lays out the methodological framework. In section 3 we present implementation details of regularised regressions using L_2Boosting. Sections 4 and 5 report the results of a Monte Carlo simulation and the empirical application. Section 6 concludes.

2 Stochastic discount factor projections

Let x be a vector of excess returns on N different assets or portfolios of assets. The stochastic discount factor model states that the excess returns satisfy the N moment conditions

$$\mathrm{E}[mx] = 0, \tag{2}$$

where m is a stochastic discount factor (SDF). The model is given economic content by specifying a functional form for the discount factor. We will consider linear models of the form

$$m = 1 - \delta' f, \tag{3}$$

for M-vectors of factors f and parameters δ. For a model of excess returns we can set factor means to zero, *i.e.* $f = \tilde{f} - \mathrm{E}[\tilde{f}]$, and take the intercept "1" as an arbitrary normalisation. The two interesting questions are how well the discount factor model can explain the cross-section of expected returns, and which factors are priced. The two questions are related, as the vector of risk prices δ is estimated to maximise the model fit. We discuss both questions separately, beginning with the model fit conditional on δ, and then estimation of δ.

2.1 HJ distance

Ideally the factors explain the entire cross-section of expected returns, in which case all the moment conditions (2) hold exactly. In practice anomalies exist, either due to mispricing, omitted factors, or measurement error in observed factors. If the moment conditions are only approximate, the deviations $\mathrm{E}[mx]$ are pricing errors. Hansen and Jagannathan (1997) propose a distance measure to evaluate the fit of the stochastic discount factor model. It is defined as

$$\text{HJ} = \text{E}[mx]' \, \text{E}[xx']^{-1} \, \text{E}[mx], \tag{4}$$

which is a quadratic form in the pricing errors with weighting matrix $\text{E}[xx']^{-1}$. Hansen and Jagannathan (1997) discuss the difference between the distance measure HJ and a general optimal GMM weighting matrix. Using $\text{E}[xx']^{-1}$ as a weighting matrix assures that results are invariant to repackaging of the assets, and independent of the model for m. Forming portfolios of the original assets does not change the HJ distance. Independence of m facilitates model comparisons using the HJ distance. Moreover, if squared pricing errors are independent of the cross-products of returns, the weighting matrix is optimal for GMM.

In applications the population moments in (4) are replaced by sample moments assuming that we have a sample of T observations for both x and f. When N is large relative to T, the weighting matrix $\text{E}[xx']^{-1}$ contains $O(N^2)$ elements to be estimated. Finding any quantity depending on $\text{E}[xx']^{-1}$ involves a huge matrix inversion that can be very sensitive to estimation error. In the really large-N case, when $N > T$, the sample second moment matrix of excess returns will even be singular.[2,3]

Our approach transforms the problem of estimating a weighting matrix to a model selection problem that constructs interesting portfolios from the N test assets. Let

$$\hat{m} = \text{E}[mx]' \, \text{E}[xx']^{-1} x \tag{5}$$

be the projection of m onto the excess returns x. Then the HJ distance can be rewritten as

$$\text{HJ} = \text{E}\left[\hat{m}^2\right], \tag{6}$$

a result that follows by direct calculation using the definition of $\hat{m}$ in (4). This expresses the distance as the expected magnitude of the squared projection of the discount factor on the excess returns. To avoid the explicit need for $\text{E}[xx']^{-1}$ we estimate $\hat{m}$ as the fitted values from the regression model

$$m = \xi' x + w, \tag{7}$$

with ξ a vector of regression parameters. Without restrictions on ξ this regression still requires the same large-N matrix $\text{E}[xx']^{-1}$. It becomes feasible by imposing some form of

[2] Hansen and Jagannathan (1997) provide asymptotic distribution theory for the sample HJ distance for fixed N and a true HJ distance strictly greater than zero. For fixed small N the asymptotic distribution of the HJ distance, under the null that all assets are correctly priced, and assuming homoskedasticity is chi-squared with $N - K$ degrees of freedom. With large N neither is a good approximation.

[3] A partial solution is the use of high-frequency data to obtain a much more accurate estimate of the quadratic variation. When not all assets are traded on a sufficiently high frequency, this restricts the space of test assets.

sparsity on the parameters ξ. The sparsity constraint implies that many of the elements in ξ should be small or equal to zero. The estimation algorithm selects the combination of test assets that provide the most flagrant violations of the pricing conditions. Implementation details, and the form of sparsity, will be discussed in section 3 below.

2.2 Risk price estimator

When the risk prices are unknown, the HJ distance can be minimised with respect to δ. The solution is

$$\delta = \left(\mathrm{E}[fx']\,\mathrm{E}[xx']^{-1}\,\mathrm{E}[xf']\right)^{-1}\mathrm{E}[fx']\,\mathrm{E}[xx']^{-1}\,\mathrm{E}[x] \tag{8}$$

Again we will transform the problem such that we avoid the explicit estimation of the high-dimensional weighting matrix $\mathrm{E}[xx']^{-1}$. We interpret the optimal value for δ in (8) as the instrumental variables (IV) estimator for δ in the regression model,

$$1 = \delta' f + u, \tag{9}$$

using x as instruments. To see the equivalence, note that the first stage regression

$$f = \Pi' x + v, \tag{10}$$

with $(N \times M)$ parameter matrix Π implies instruments

$$\hat{f} = \mathrm{Proj}(f|x) = \mathrm{E}[fx']\,\mathrm{E}[xx']^{-1}x\ . \tag{11}$$

Given $\hat{f}$, the second stage IV estimator then becomes

$$\delta = \mathrm{E}[\hat{f}f']^{-1}\,\mathrm{E}[\hat{f}1], \tag{12}$$

which is identical to the original expression (8).

The large-N challenge is in the tracking portfolios $\hat{f}$, as they depend on $\mathrm{E}[xx']^{-1}$. Analogously to the pricing error regression (7) before, we construct these portfolios using a regularised linear regression imposing a sparsity constraint on the elements of the matrix of portfolio weights Π. The fitted values from this regression, $\hat{f}$, are the instruments in (12).

Our model setup fits directly within the framework of Belloni, Chen, Chernozhukov, and Hansen (2012). They consider optimal instrument selection in a regression model with a fixed small number of endogenous regressors f, for which we have many instruments x. All x_j are potential instruments, but including too many will lead to overfitting in finite samples and thus create a bias. Under sparsity conditions for the projections, Belloni *et al* (2012) prove that various machine learning methods lead to asymptotically optimal instruments, one of them being L_2Boosting that we use in the empirical work.

Let $\boldsymbol{F}$ and $\boldsymbol{X}$ be the data matrices containing T rows of observations on the M factors and N excess returns on the test assets, respectively. In the empirical analysis factors are always demeaned, *i.e.* $\boldsymbol{\iota}'\boldsymbol{F} = 0$. The first stage penalised regression provides a matrix $\hat{\boldsymbol{F}} = \boldsymbol{X}\hat{\Pi}$ as a linear combination of the instruments. Using $\hat{\boldsymbol{F}}$ as instruments, and with $\boldsymbol{\iota}$ a T-vector of ones, the IV estimator for δ in (12) becomes

$$\hat{\delta} = (\hat{\boldsymbol{F}}'\boldsymbol{F})^{-1}\hat{\boldsymbol{F}}'\boldsymbol{\iota} \tag{13}$$

with an asymptotic covariance matrix consistently estimated as

$$\text{Var}(\hat{\delta}) = s_u^2(\hat{\boldsymbol{F}}'\boldsymbol{F})^{-1}\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}}(\boldsymbol{F}'\hat{\boldsymbol{F}})^{-1} \tag{14}$$

with $s_u^2 = \frac{1}{T}\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}$ (for $\hat{\boldsymbol{u}} = \boldsymbol{\iota} - \boldsymbol{F}\hat{\delta}$) the sample second moment of the residuals in the SDF model (9). The scaled numerator term in the estimator (13), $\frac{1}{T}\hat{\boldsymbol{F}}'\boldsymbol{\iota}$, is nothing but the time series mean of the tracking portfolios. This is a standard estimator for the risk premium of a factor using factor mimicking portfolios.

The denominator differs. The usual transformation from risk premiums to risk prices is by the inverse of the factor covariance matrix. In (13) we have $\frac{1}{T}\hat{\boldsymbol{F}}'\boldsymbol{F}$, which for large T converges to the second moment matrix of the mimicking portfolios $\text{E}[\hat{f}\hat{f}']$. Since mimicking portfolios for priced factors do not have a zero mean, this is not the covariance matrix. The difference between second moments and covariances will be small when returns are measured with reasonably high frequency.[4]

The first stage projection becomes redundant for factors that are traded and also among the test assets. For a traded factor we observe the excess return $\tilde{f}$ (note that $f = \tilde{f} - \text{E}[\tilde{f}]$ are demeaned factors), which is assumed to be perfectly priced by the moment condition $\text{E}[m\tilde{f}] = 0$. Its price is

$$\delta_T = \text{E}[\tilde{f}f']^{-1}\,\text{E}[\tilde{f}1], \tag{15}$$

where the subscript 'T' indicates that we consider $\tilde{f}$ as a 'T'raded factor. The structure is identical to the general IV estimator (12), but now with $\tilde{f}$ instead of $\hat{f}$ as the obvious instruments. Since this does not involve the N-dimensional matrix inversion $\text{E}[xx']^{-1}$, estimation of the risk prices does not pose a high dimensional challenge.

When the traded factors are not among the test assets, or if the moments $\text{E}[m\tilde{f}]$ are allowed to deviate from zero, we still need the first stage regression to find the δ that

[4] We obtain the second moment matrix due to the second moment weighting matrix in the HJ distance. Replacing the second moment matrix in the HJ distance by a covariance matrix, as in Giglio and Xiu (2021), leads to adding a constant term in the tracking portfolio regression (10) and to a covariance matrix for the mimicking portfolio $\hat{f}$ in the transformation from risk premium to risk prices. We keep the second moments to remain close to the original Hansen-Jagannathan distance.

minimises the HJ distance. The difference in treatment of the traded factors is similar to the choice between estimating risk prices from a cross-sectional regression of (expected) returns on beta's (covariances) or estimating them directly from the traded factors. The evaluation of the fit on the test assets x remains, however, a large N problem, with or without traded factors. Using the estimate $\hat{\delta}$ in (3) we form the SDF and assess the HJ distance using the regularised regression (7).

2.3 Efficient Frontier

For the interpretation of the HJ distance it is useful to split the projection of m in two parts,

$$\hat{m} = \text{Proj}(m|x) = \text{Proj}(1|x) - \delta' \text{Proj}(f|x). \tag{16}$$

When the tracking portfolios span the efficient frontier, the residual $\hat{m}$ will be zero. The projection of a constant on the excess returns, $\hat{1} \equiv \text{Proj}(1|x)$, defines a mean-variance (MV) efficient portfolio (Cochrane, 2005; Britten-Jones, 1999), which can be constructed by the regression model

$$1 = \pi_1' x + v_1. \tag{17}$$

The HJ-distance is measured by the Mean-Squared-Error (MSE) of the difference between an unrestricted MV portfolio and the optimal portfolio implied by the factor mimicking returns.

The decomposition suggests that the projection of the SDF on the space of excess returns can also be computed using $\hat{m} = \hat{1} - \delta' \hat{f}$. In population this is an identity, but due to sparsity conditions it is not an identity at estimation stage with regularised regressions. Sparsity constraints on the tracking portfolio weights Π in (10) and the mispricing weights ξ in (7) do not imply sparsity in π_1 for the mean-variance optimal portfolio, and vice versa. We impose sparsity on the tracking portfolio weights Π and the anomaly portfolio ξ, not on the overall MV efficient portfolio π_1. Constructing a mean-variance efficient portfolio, without referring to factors, is of independent interest. We include appendix C as an example.

3 L_2Boosting regressions

We need regularised linear regressions for several purposes. First, to evaluate the magnitude of the pricing errors through (7); second, to construct tracking portfolios for non-traded factors as in (10); and third, for additional insight, to estimate the mean-variance efficient portfolio in (17). All three can be analysed in the regression model

$$\boldsymbol{y} = \sum_{j=1}^{N} \theta_j \boldsymbol{x}_j + \boldsymbol{v}, \tag{18}$$

where $\boldsymbol{y}$ is the T-vector with observations of the dependent variable y_t, $\boldsymbol{v}$ is the T-vector of errors, and $\boldsymbol{x}_j$ are T-vectors of observations on the excess returns x_{jt}. All excess returns together are stored in the $(T \times N)$ matrix $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$. The dependent variable y is either a factor f, the constant 1, or an SDF m. Fitted values are always excess portfolio returns: a mean-variance efficient portfolio, a factor mimicking portfolio, or an anomaly.

Boosting algorithms were developed as a technique for producing a projection by aggregating weak predictors. Our version closely follows Bühlmann (2006). Although it is a learning algorithm that penalizes the L_2-norm of the parameters, boosting differs from ridge regression. In ridge regression all coefficients are non-zero, but shrunk towards a target. With L_2Boosting, the final result will have many exact zeros, very much like Lasso and Elastic Net estimators (Hastie, Tibshirani, and Friedman, 2009).

Algorithm 1 L_2Boosting

1: **Initialize**
    Step size parameter $\nu \in (0, 1]$
    Maximum number of iterations L
    Projection $\hat{\boldsymbol{y}} = 0$
    Coefficients $\hat{\theta}_j = 0$ $(j = 1, \ldots, N)$
2: **for** $\ell = 1$ to L **do**
3:     Compute residuals $\hat{\boldsymbol{v}} = \boldsymbol{y} - \hat{\boldsymbol{y}}$
4:     Find univariate regression coefficients $p_j = (\boldsymbol{x}_j' \boldsymbol{x}_j)^{-1} \boldsymbol{x}_j' \hat{\boldsymbol{v}}$
5:     Find best predictor $j^* = \text{argmin}_j(\hat{\boldsymbol{v}}_j' \hat{\boldsymbol{v}}_j)$ with $\hat{\boldsymbol{v}}_j = \hat{\boldsymbol{v}} - p_j \boldsymbol{x}_j$
6:     Update the projection: $\hat{\boldsymbol{y}} \leftarrow \hat{\boldsymbol{y}} + \nu p_{j^*} \boldsymbol{x}_{j^*}$
7:     Update regression coefficients: $\hat{\theta}_{j^*} \leftarrow \hat{\theta}_{j^*} + \nu p_{j^*}$

The detailed steps of the algorithm are listed in algorithm 1. At each iteration the algorithm searches for the univariate predictor that improves the fit the most. Implementation requires two tuning parameters: the stepsize ν and the number of iterations L. The exact value of the stepsize parameter does not have much of an effect on the results, as long as it is sufficiently small. It should be large enough for the algorithm to make progress, yet small enough to enable shrinkage and deal with multicollinearity. The value $\nu = 0.1$ is often recommended. The number of iterations is more critical, since eventually the solution will converge to the least squares estimator and therefore be prone to overfitting.

Each iteration increases model complexity. Since boosting is a linear method, the fitted values can be expressed as $\hat{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{y}$, where $\boldsymbol{H}$ is the $(T \times T)$ projection (*hat*) matrix (see appendix A for the updating formula). Model complexity is defined as $q = \text{tr}(\boldsymbol{H})$. In standard regression models, doing OLS on a full data matrix $\boldsymbol{X}$, $\text{tr}(\boldsymbol{H})$ equals the number of explanatory variables, *i.e.* the number of columns in $\boldsymbol{X}$. Due to the tuning parameters ν and L, model complexity q is less than the number of included variables. Bühlmann (2006) suggests to use the model complexity as an input in the corrected Akaike Information Criterion

$$AIC(\hat{\boldsymbol{v}}, q) = \ln\left(1 - R^2\right) + \frac{T+q}{T-q-2}, \tag{19}$$

where $(1 - R^2) = \hat{\boldsymbol{v}}'\hat{\boldsymbol{v}}/\boldsymbol{y}'\boldsymbol{y}$. The penalty is added to the log of the sum of squared residuals $\hat{\boldsymbol{v}}'\hat{\boldsymbol{v}}$. Minimising AIC leads to an optimal stopping time L.

The L_2Boosting algorithm is consistent for the conditional expectation of the response, if the regression coefficients in (18) satisfy the sparsity condition

$$\sum_{j=1}^{N} |\theta_j| = o(\sqrt{T/\ln(N)}) \tag{20}$$

for $T \to \infty$, maintaining that $\ln(N)/T \to 0$ (assuming existence of sufficient moments for x and y; see Bühlmann and Van de Geer (2011, section 12.6.2.2)). This sparsity condition is attractive for our purpose of constructing a tracking portfolio. With more data the model may become more complex and the number of assets in the tracking portfolio may grow very quickly, but the weight of each individual asset should shrink. The condition is likely to be satisfied for standard asset pricing models. For example, when excess returns follow a factor model, appendix B.2 shows that the portfolio weights for the tracking portfolio satisfy the condition. The same holds for the mean-variance portfolio with '1' as the dependent variable.

In our experience, and well-known in the literature, the AIC tends to select rather complex models. Complex models increase the risk of overfitting, which creates a bias for the IV estimator in the direction of the OLS estimator. In our case the OLS estimator $\hat{\delta}_{\text{ols}} = (\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{\iota} = 0$ because the factors are demeaned. Overfitting in the first stage regression will thus result in risk prices that may be biased towards zero. We therefore perform our model selection using cross-validation. With K-fold cross-validation the sample is split in K equal-sized subsamples (folds). For each fold k, the boosting algorithm performs a sequence of L iterations for the parameter vector θ on the complement of all data not in the k^{th} subsample. These parameters are used to generate fitted values and residuals for the k^{th} subsample. Doing this for all k gives a complete vector $\hat{\boldsymbol{v}}$ of T validation sample residuals. The value of L that minimises the residual sum of squares in the validation samples determines the optimal number of iterations. In practice we use 5-fold cross validation with random subsamples. We repeat the cross validation $\mathcal{T}$ times to minimise the sampling variation induced by the random subsample assignment.[5] Details are listed in algorithm 2.

Algorithm 2 Repeated cross-validation

1: **Initialize**
   Number of CV folds K
   Number of repeated cross-validations $\mathcal{T}$
2: **for** $\tau = 1$ to $\mathcal{T}$ **do**
3: Randomly order all time indices in the data to construct data matrices $\boldsymbol{X}$ and $\boldsymbol{y}$ with permuted rows.
4: **for** $k = 1$ to K **do**
5: Partition $\boldsymbol{X}$ in blocks $\boldsymbol{X}_k$ with the data for fold k and its complement $\boldsymbol{X}_{-k}$. Do the same for $\boldsymbol{y}$.
6: Run L iterations of L_2Boosting projecting $\boldsymbol{y}_{-k}$ on $\boldsymbol{X}_{-k}$, and save parameters $\hat{\theta}_{-k}^{(\ell)}$ at all boosting iterations $\ell \leq L$.
7: Compute fitted values $\hat{\boldsymbol{y}}_k^{(\ell)} = \boldsymbol{X}_k \hat{\theta}_{-k}^{(\ell)}$ and residuals $\hat{\boldsymbol{v}}_k^{(\ell)} = \boldsymbol{y}_k - \hat{\boldsymbol{y}}_k^{(\ell)}$
8: Find optimal stopping $L^* = \text{argmin}_\ell \sum_k \boldsymbol{v}_k^{(\ell)\prime} \boldsymbol{v}_k^{(\ell)}$
9: Save fitted values $\hat{\boldsymbol{y}}^\tau = (\hat{\boldsymbol{y}}_1', \ldots, \hat{\boldsymbol{y}}_K')'$ at the optimal stopping time L^*
10: Compute the average $\hat{\boldsymbol{y}} = \frac{1}{\mathcal{T}} \sum_\tau \hat{\boldsymbol{y}}^\tau$

We apply the L_2Boosting separately to each factor. For each factor j data are in the column vector $\boldsymbol{f}_j$ and the tracking portfolio returns in $\hat{\boldsymbol{f}}_j$. Joining all factors together we have the $(T \times M)$ data matrix $\hat{\boldsymbol{F}}$ for the instruments.

In algorithm 2 the fitted values at the optimal number of boosting iterations are out-of-sample estimates, with each fold k using portfolio weights that are estimated with the data not in fold k. The motivation for the out-of-sample fitted values comes from the split-sample IV estimator in Angrist and Krueger (1995). Its purpose is to reduce bias in cases with many potentially weak instruments. See appendix B.3 for details.

Because we use regularised regressions for the mimicking portfolios, the IV estimator (13) differs from a number of alternatives that would normally be equivalent. For example, with the complete set of N instruments, the projection matrix is $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and we obtain the tracking portfolios $\hat{\boldsymbol{F}} = \boldsymbol{HF}$ and the MV-efficient portfolio $\hat{\boldsymbol{\iota}} = \boldsymbol{H\iota}$. Risk

[5] Since the cross-validation randomly assigns observations to folds, some sampling variation will remain in the estimates. To minimise this effect we use the overly large $\mathcal{T} = 1000$ in the empirical work.

prices could then be estimated from the second stage OLS regression,

$$\hat{\boldsymbol{\iota}} = \hat{\boldsymbol{F}}\delta + \boldsymbol{u}, \tag{21}$$

which has the returns on an optimal mean-variance portfolio on the left-hand side, and the factor portfolio returns on the right-hand side. With unregularised projections this two-stage least squares estimator would be identical to the IV estimator (13). The identity holds because in this unrestricted case $\boldsymbol{H}^2 = \boldsymbol{H}$, implying $\hat{\boldsymbol{F}}'\boldsymbol{F} = \hat{\boldsymbol{F}}'\hat{\boldsymbol{F}}$. With boosting (and other statistical learning algorithms), $\boldsymbol{H}$ is a pseudo projection matrix and is not idempotent. Furthermore, when different macro variables use different optimal instruments we have $\hat{\boldsymbol{f}}_j = \boldsymbol{H}_j \boldsymbol{f}_j$ with individual projection matrices $\boldsymbol{H}_j$, which is a second reason the equivalence with 2SLS breaks down. In our empirical work we follow Belloni et al. (2012) and use the IV estimator (13).

For the HJ distance we need the projection of the SDF $\boldsymbol{m} = \boldsymbol{\iota} - \boldsymbol{F}\delta$ on the excess returns $\boldsymbol{X}$. We therefore apply the L_2Boosting algorithm with dependent variable $\boldsymbol{\iota} - \boldsymbol{F}\hat{\delta}$, *i.e.* the residuals $\hat{\boldsymbol{u}}$ of the IV regression (13). Denoting the fitted values by $\hat{\boldsymbol{m}}$ the Hansen-Jagannathan distance is computed as

$$\text{HJ} = \frac{1}{T}\hat{\boldsymbol{m}}'\hat{\boldsymbol{m}}. \tag{22}$$

As for the IV estimator (13), there are alternative ways to compute the HJ distance, which would be equivalent without regularisation, but are different when using shrinkage and model selection. As mentioned before in section 2.3 an alternative option would be to use the returns of the mean-variance efficient portfolio $\hat{\boldsymbol{\iota}}$ and construct $\hat{\boldsymbol{m}} = \hat{\boldsymbol{\iota}} - \hat{\boldsymbol{F}}\hat{\delta}$. Due to the different amounts of shrinkage applied in construction $\hat{\boldsymbol{\iota}}$ and $\hat{\boldsymbol{F}}$ the two sets of fitted values have a different scaling, and hence the quantity $\hat{\boldsymbol{m}} = \hat{\boldsymbol{\iota}} - \hat{\boldsymbol{F}}\hat{\delta}$ is ill-behaved. The problem does not occur with a regularised regression of $\boldsymbol{\iota} - \boldsymbol{F}\delta$ on $\boldsymbol{X}$, since the shrinkage is applied to the entire left-hand side, and not independently (with different amounts of shrinkage) to separate components $\boldsymbol{\iota}$ and $\boldsymbol{F}$.

4 Monte Carlo Evidence

We conduct a Monte Carlo study to evaluate the properties of the IV estimator of the risk parameters and HJ distance using the L_2Boosting algorithm to select instruments. Asymptotic theory in Belloni et al. (2012) indicates that the IV estimator is consistent and asymptotically normal with standard errors as if we would have the optimal instruments. We choose a setting that may be challenging for two reasons. First, we consider cases where T is small relative to N, such that standard GMM with an optimal weighting matrix is not feasible. Second, we simulate data that resemble noisy equity returns that have a very low correlation with a non-traded factor. Both may lead to poor small sample performance of the IV-boosting estimator. In addition we allow for noisy observations of the factor and mispricing.

For the simulations excess returns are generated by the 2-factor model,

$$x = \beta_1(f_1 + \lambda_1) + \beta_2(f_2 + \lambda_2) + e, \tag{23}$$

where β_j are N-vectors of factor loadings, and e is an N-vector of idiosyncratic risk. Factors f_1 and f_2 are mutually independent with variances ω_1^2 and ω_2^2, and also independent of the idiosyncratic risk e. In estimation we only include the first factor. The second factor controls the cross-sectional error structure and mispricing. To lighten notation we will mostly drop the subscript on the first factor, and write the stochastic discount factor model $m = 1 - \delta f$, with $f = f_1$, $\delta = \delta_1$, $\lambda = \lambda_1$, and $\omega = \omega_1$.

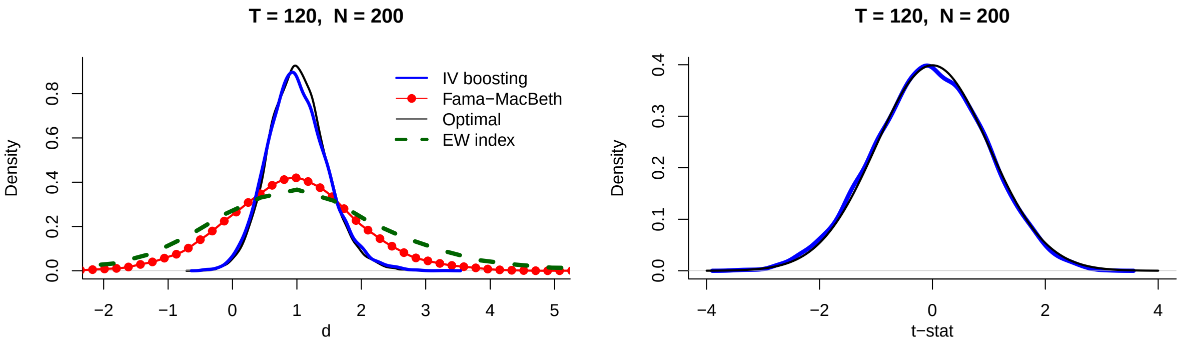
4.1 Correct specification

The model is correctly specified if $\lambda_2 = 0$. Misspecification will be considered in section 4.2. To satisfy the pricing condition $\mathrm{E}[mx] = 0$, we must have $\lambda = \delta\omega^2$. As a normalisation we set $\delta = 1$ as the true value. Since the SDF prices all assets, the HJ distance is zero.

The parameters in the DGP are calibrated to meet a number of design criteria. Appendix B contains a full specification of the calibration. Below we summarise the main ingredients. First, the factor variance ω^2, which is also the maximum Sharpe ratio for any trading strategy, is set equal to 0.8 on an annual basis. Second, for a low correlation between x and f we select the factor loadings β such that the squared correlation between f and an equally weighted portfolio is equal to 0.1. A third design criterion fixes the idiosyncratic noise variance as the difference between the average variance of individual assets and the variance of the equally weighted portfolio. As a fourth criterion, the cross-sectional variation in expected returns determines how much can be explained by a regression of average returns on beta's. This defines the cross sectional dispersion in β. Finally we specify the cross-correlations in the errors $\eta = \beta_2 f_2 + e$ with implied error covariance matrix $\Sigma = \mathrm{E}[\eta\eta']$. The larger and more dispersed the elements in β_2, the bigger the difference between the optimal weighting matrix and the identity weighting matrix for a cross-sectional regression of average returns on beta's. We assume that factor loadings β_2 are cross-sectionally independent of β. This resembles a setting where we wish to estimate the price of risk of a macro factor, knowing that a strong factor structure will remain in the test assets.

We discuss one of the Monte Carlo experiments in detail. The example has $N = 200$

(a) $\hat{\delta}$ densities (b) t-stat density



Panel (a) shows densities of the estimates of $\hat{\delta}$ in the SDF model $m = 1 - \delta f$ using simulated data with $N = 200$ test assets and $T = 120$ time series observations. In panel (b) the blue solid line is the Monte Carlo density of the t-statistic $t = (\hat{\delta} - 1)/s(\hat{\delta})$, where $s(\hat{\delta})$ is the asymptotic standard error of the IV-estimator. The thin black line is the standard normal density. See Table 1 for further notes.

Figure 1: Monte Carlo densities

test assets and $T = 120$ time series observations. Other combinations of N and T are discussed in appendix B. Since $N > T$ this is a setting where Fama-MacBeth would usually be the only option to estimate δ, since the large N precludes estimation of a GMM weighting matrix. The IV estimator with optimal instrument selection is designed for this setting. We compare it with the Fama-MacBeth estimator, which is implemented by running a cross-sectional regression of the N sample average excess returns on the N covariances between returns and the factor.[6]

Figure 1(a) shows densities of $\hat{\delta}$ for four alternative estimators; Table 1 provides summary statistics. The IV estimator appears almost unbiased. Moreover, it is also efficient: in figure 1(a) its density nearly overlaps with the infeasible optimal instrument that uses the population weights for the tracking portfolio. This would also be the GMM estimator with the optimal population weighting matrix. The boosting algorithm succeeds in constructing a tracking portfolio that performs nearly as well as an estimator with optimal weights. The efficiency loss of the IV-boosting estimator is a tiny difference in standard deviation: 0.47 versus 0.44. The difference between the boosting and the infeasible optimal estimator is only in the tails of the sampling distribution. These simulation results therefore confirm that the asymptotic theory in Belloni et al. (2012) is relevant is an asset pricing setting with a low signal-to-noise ratio and highly correlated regressors.

The most interesting comparison is with the Fama-MacBeth estimator. As expected the Fama-MacBeth estimator is slightly downward biased due to the well known errors-in-

[6] For a fair comparison between FM and IV we estimate the FM cross-sectional regression without a constant term.

Estimator	ave	std	quantiles 1%	25%	med	75%	99%
A: independent design							
IV boosting	1.01	0.47	0.01	0.69	0.98	1.30	2.27
Optimal	1.02	0.44	0.05	0.71	1.00	1.30	2.17
Fama-MacBeth	0.93	1.01	-1.47	0.27	0.93	1.57	3.43
Equal Weight	1.15	3.16	-2.42	0.13	1.01	1.93	5.87
B: orthogonal design							
IV boosting	1.02	0.44	0.06	0.72	1.00	1.30	1.76
Optimal	1.01	0.43	0.07	0.73	1.01	1.30	2.10
Fama-MacBeth	0.98	0.44	-0.02	0.68	0.96	1.26	2.08
C: noisy factor							
IV boosting	1.12	0.75	-0.16	0.65	1.01	1.45	3.47
Optimal	1.09	0.52	0.04	0.73	1.03	1.38	2.54
Fama-MacBeth	0.80	0.93	-1.43	0.23	0.77	1.35	3.17
D: mispricing							
IV boosting	0.99	0.46	0.03	0.68	0.96	1.27	2.18
Optimal	1.04	0.44	0.06	0.74	1.02	1.30	2.14
Fama-MacBeth	3.45	1.27	1.45	2.65	3.41	4.25	6.62

The table report averages, standard deviations, and quantiles for alternative estimators for δ in the discount factor model $m = 1 - \delta f$ with true value $\delta = 1$. Simulated data are generated for $N = 200$ test assets and $T = 120$ time series observations. *IV boosting* denotes the Instrumental Variables estimator using a tracking portfolio for f obtained by L_2Boosting; *Optimal* is the infeasible optimal IV estimator using the population portfolio weights for the tracking portfolio; *Equal Weight* is the IV estimator using the equally weighted portfolio of the test assets as the instrument; *Fama-MacBeth* is the cross-sectional regression of the sample average excess returns on the sample covariances of excess returns with the factor.
Panel A refers to a design with independent factor loadings for β and β_2 such that $\text{cov}[\beta, \beta_2] = 0$. Panel B has an orthogonal design with $\text{E}[\beta\beta_2] = 0$. Panel C has the same design as panel A apart from adding measurement noise to the factor. Panel D is similar to panel A, except for setting $\lambda_2 = 0.9$ to introduce pricing errors. Further details on the simulation design are provided in appendix B. Statistics are from 10,000 replications.

Table 1: Monte Carlo risk price estimates

variables problem of using sample covariances instead of the true population values. It is also much less efficient with more than double the standard deviation of the IV estimator.

Without pricing errors, any asset is a valid instrument, but not every instrument performs well. As an example we consider the equally weighted (EW) portfolio of all N assets as an instrument. With 200 test assets the EW portfolio is well-diversified. As an instrument it performs poorly. The variance of $\hat{\delta}$ is almost 16 times as large as for the optimal IV. Careful choice of instrument therefore matters for estimating the price of risk.

Important for inference are the estimated standard errors of the risk price estimates. For the boosting-IV estimator the asymptotic variance from the standard IV formula is on average almost identical to the Monte Carlo variance. Moreover, figure 1(b) shows that the density of the t-statistic is very close to normal.

The difference between IV-boosting and Fama-MacBeth depends on the calibration of the parameters in the error covariance matrix Σ. The Fama-MacBeth estimator uses an identity weighting matrix, whereas the cross sectional error distribution is far from diagonal due to the dispersion in the β_2 factor loadings. Fama-MacBeth would be optimal, and should perform at least as good as IV boosting, when we alter the design such that the cross-sectional covariance matrix Σ satisfies $\Sigma\beta = \sigma^2\beta$. In this case there is no benefit in doing a GLS or GMM cross-sectional regression. The condition will hold under an orthogonal design with $\beta'\beta_2 = 0$.[7] Panel B in table 1 shows that this orthogonal design does not affect the IV estimator, whereas it leads to a huge improvement for the Fama-MacBeth estimator. The two now perform equally well.

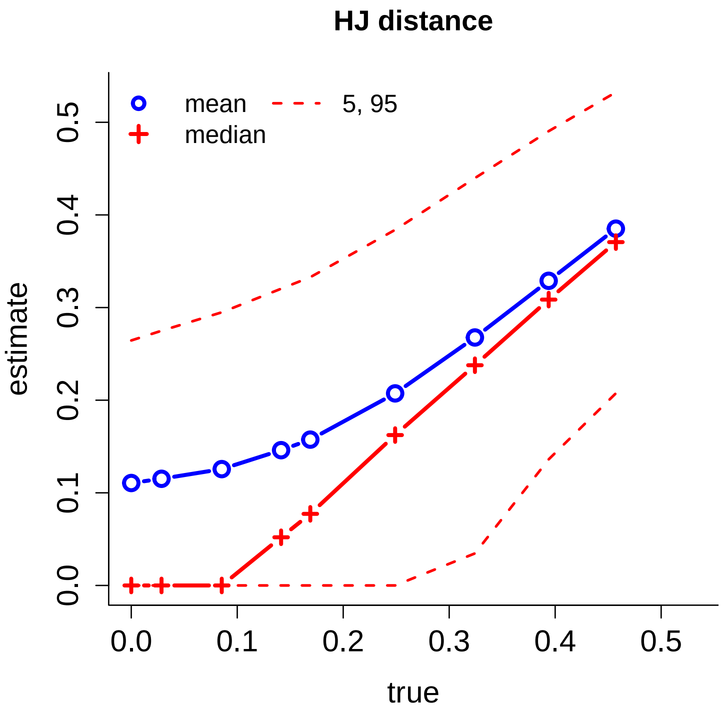
Since macroeconomic factors are noisy, we also consider the case that the observed factor has substantial measurement error. For this we assume that returns are as in the baseline model, but the factor itself is observed with noise: $f_n = f + n$, where n is independent zero mean noise. More specifically, we set $\text{var}\,[n] = 2\omega^2$, meaning that the noise variance is twice as big as the signal itself. This specification resembles a weak instrument setting. The noise does not affect the pricing condition, but considerably reduces the correlation between the factor and excess returns. Panel (c) in table 1 shows that the large amount of noise affects the IV and FM estimators in different ways. The IV-estimator is upward biased, whereas FM is downward biased. The biases are not very large compared to the much larger variance for both estimators. The IV estimator still performs better than FM.

Data are generated by a model without pricing errors. The population HJ distance is therefore equal to zero. For the baseline calibration the sampling distribution for HJ contains 58% exact zeros. An exact zero occurs when the boosting algorithm cannot find a single excess return that has out-of-sample predictive power in the cross validation. In other words, in 58% of cases the boosting algorithm cannot find an anomaly. Since HJ is non-negative by construction, the average HJ is positive.

4.2 Pricing errors

As a next step we add pricing errors. We introduce mispricing by adding a risk premium λ_2 in (23). Otherwise all parameters are the same as in the baseline calibration for panel A in table 1: same factors and factor loadings. The only difference is a positive value for λ_2. Cross-sectionally this type of mispricing is an omitted factor. If we would observe f_2, and construct a tracking portfolio, the HJ distance would still be zero. This is the same design

[7] Under the 'independent' design the covariance between the two vectors of factor loadings is equal to zero. With 'orthogonality' the inner product is zero, which implies that the cross-sectional covariance equals $-\,\text{E}[\beta]\,\text{E}[\beta_2]$, such that $\text{E}[\beta\beta_2] = \text{E}[\beta]\,\text{E}[\beta_2] + \text{cov}\,[\beta, \beta_2] = 0$.



The figure shows true and estimated HJ distances. Data are generated by a 2-factor model, but the model is estimated using a single factor. The true HJ distance is a function of the risk premium λ_2 for the omitted factor using (65) in appendix B.4 averaged over replications of the factor loadings. Except for λ_2, factors and factor loadings are identical to the design in panel A of table 1. The dashed lines are the 5% and 95% quantiles of the HJ distance estimates. For scaling the axes show the square root of HJ.

Figure 2: Pricing error estimates

as in Giglio and Xiu (2021) for studying mispricing.

In the Monte Carlo design this implies that the true discount factor changes to $m = 1 - \delta f - \delta_2 f_2$ and thus gets a higher variance. By necessity the maximum Sharpe ratio then also increases. The unrestricted mean-variance portfolio will thus have a larger Sharpe ratio than the value of 0.8 in the original design. In figure 2 the range of λ_2 corresponds to a maximum Sharpe ratio moving from 0.8 (annually) for $\lambda_2 = 0$ to $\max(\text{Sh}) = 2$ for $\lambda_2 = 0.15$ at the far end of the x-axis.

Figure 2 shows that estimates of HJ will be biased upwards for small pricing errors. This is because HJ, as a distance measure, is non-negative by construction. Since the boosting algorithm sets HJ to zero in the majority of cases, the median remains at zero for small mispricings. The larger the mispricing the closer mean and median move together. For large mispricing our estimator slightly underestimates the true value of HJ. The estimator thus has power to detect mispricing.

Panel D in table 1 shows the effects of pricing errors on the estimates of the risk price.

The results refer to the case that λ_2 is chosen such the maximum Sharpe ratio is 1.2, 50% higher than the baseline design in panel A. The IV estimates are hardly affected by the mispricing. Both the average as well as standard error in the misspecified model are very close to the correctly specified model in panel A. The same holds for the asymptotic standard errors, which remain close to the Monte Carlo standard error. Results for the Fama-MacBeth estimator are completely off. The reason is that the FM estimator depends on the factor loadings β and β_2 in the design, whereas for large N the IV estimator is independent of factor loadings. The IV estimator only requires the tracking portfolio instrument. Appendix B.4 provides analytical details.

5 Empirical results

We use our methodology to assess pricing kernels of both traded and non-traded factors. The non-traded factors are macroeconomic variables that are well established in the asset pricing literature. The traded factors serve as a benchmark, both for our methodology as well as for the macro factors. Since the methodology is for a fixed number of factors, we only include a few basic macro variables, without an exhaustive model selection on all macro variables reviewed in Cochrane (2017) or the library of macro factors discussed in McCracken and Ng (2015).

5.1 Data

All data are monthly for the period July 1963 – December 2017. Table 2 contains the sources for the macroeconomic variables. The first variable is *consumption*, being the fundamental macro factor in asset pricing. We take real expenditures of nondurables plus services.[8] Following much of the literature we consider news about the annual growth of monthly consumption as a factor.[9] Denoting the log consumption flow in month t by c_t, we define the annual growth in monthly consumption as $C_t^{12} = c_t - c_{t-12}$.

Other standard macroeconomic factors, at least since Chen, Roll, and Ross (1986), are *inflation*, the *credit spread*, and the *term spread*. For inflation we construct the factor as the

[8] We refrain from searching among the many proposed measures for consumption, such as, *e.g.* garbage (Savov, 2011), unfiltered data (Kroencke, 2017), or durables (Yogo, 2006).

[9] We consider annual growth rates, again without an elaborate search on the optimal horizon as in Parker and Julliard (2005) or Malloy, Moskowitz, and Vissing-Jørgensen (2009). A pragmatic reason to work with annual growth rates is the timing of information. With nowcasting and analyst expectations much of the news of current month macroeconomic data is already known before the end of the month. Also, since consumption growth seems to have a persistent component (Schorfheide, Song, and Yaron, 2018), an annual horizon may be preferable. See Jagannathan and Wang (2007, section III.B) for a detailed analysis of the trade-offs in selecting the horizon in an SDF model.

Consumption, services	U.S. Bureau of Economic Analysis, Real personal consumption expenditures: Services (chain-type quantity index), DSERRA3M086SBEA
Consumption, non-durable	U.S. Bureau of Economic Analysis, Real personal consumption expenditures: Nondurable goods (chain-type quantity index), DNDGRA3M086SBEA
Consumer Price Index	US Bureau of Labor Statistics, Consumer Price Index for All Urban Consumers: All Items, CPIAUCSL
Baa Corporate Bond Yield	Board of Governors of the Federal Reserve System, Moody's Seasoned Baa Corporate Bond Yield, BAA
10-Year Treasury Bond	Board of Governors of the Federal Reserve System, 10-year treasury constant maturity rate, DGS10
3-Month Treasury Rate	Board of Governors of the Federal Reserve System, 3-Month Treasury Bill: Secondary Market Rate, DTB3

Table 2: Macro data sources

innovation in the annual change in the log Consumer Price Index (p), $Inf_t^{12} = p_t - p_{t-12}$. The term spread ($TS$) is the difference between the 10-year Government Bond rate and the 3-month Treasury Bill rate. The credit spread (CS) is the difference between the BAA yield and the 10-year Government Bond rate.[10] Both spreads are business cycle indicators (Harvey, 1993; Gilchrist and Zakrajsek, 2012).

For the return data we use a large number of managed portfolios based on common industry sorts along with well established financial anomalies.[11] The industry portfolios enable us to relate our macroeconomic factors to typical portfolios that span the entire market and are likely to have different exposures to macro risk factors. Industry portfolios have been used for this since the seminal Breeden, Gibbons, and Litzenberger (1989) study. We take the data for the 49 industries from Kenneth French's data library. The industry sorts are expressed in excess of the one-month Treasury bill rate. We drop series that are not fully observed between July 1963 and December 2017. For the anomaly sorts we use the collection of characteristic sorted portfolios from Kozak, Nagel and Santosh (2020) available at Serhiy Kozak's website. We remove the value-weighted and equally-weighted market portfolios along with the size and value anomalies from the set of test assets. Traded factors are the five Fama and French factors (Fama and French, 2015). We thus obtain a set of $N = 79$ returns and $T = 654$ time series observations.

Many of the anomaly portfolios have a Sharpe ratio that is above the Sharpe ratio

[10] Since we do not observe the daily Baa Corporate Bond Yield as far back, we instead use the monthly average corporate bond yield for the missing data.

[11] We use managed portfolios instead of individual stocks. Literature is divided on this choice. In recent studies, *e.g.* Kelly, Pruitt, and Su (2019), individual stocks are used to simultaneously construct factors and portfolios. Other recent studies, *e.g.* Giglio and Xiu (2021) and Kozak, Nagel, and Santosh (2020), rely on portfolio sorts.

of the market portfolio (equal to 0.42). The maximum Sharpe ratio is for the *Industry Relative Reversals* anomaly, which on its own has an annualised Sharpe ratio of 1.14. The portfolios also provide a large cross-sectional dispersion in average returns, ranging from -1.4% to $+1.0\%$ per month with a cross-sectional standard deviation 0.5%. Both the mean and median volatility of the anomalies and the industry sorts are higher than the market portfolio, although some portfolios appear to have very low risk. This set of portfolios is thus challenging for any asset pricing model.

For the tracking portfolios we expect that they load primarily on the Industry portfolios. For the pricing errors we expect stronger weights for the anomaly sorts. It is up to the learning algorithm to check if this is true.

5.2 Tracking Portfolios

For the construction of the tracking portfolios we mostly follow Lamont (2001). As in Lamont (2001) the target for the tracking portfolios for consumption and inflation are the annual growth rates observed at a monthly frequency. For both variables we regress the macro variable on the excess returns plus a small number of conditioning variables Q_t,

$$F_{t+12} = \pi' x_{t+1} + \phi' Q_t + v_{t+12}, \tag{24}$$

where F_{t+12} is either consumption growth (C^{12}_{t+12}) or inflation (Inf^{12}_{t+12}). The tracking portfolio returns are the fitted values $\hat{f}_t = \hat{\pi}' x_t$. They represent the information embedded in financial returns about the macro growth for the coming year. As conditioning variables we use a constant and the past annual, quarterly and monthly growth rates.

The main difference with Lamont (2001) is the L_2Boosting algorithm for estimating the portfolio weights. This also affects the treatment of the lagged predictor variables Q_t. Instead of the multiple regression with both x_{t+1} and Q_t, we first regress F_{t+12} on the controls Q_t, and then run the L_2Boosting algorithm on the residuals. Since we expect very little correlation between returns and lagged macro control variables Q_t, this should not affect the tracking portfolio, while substantially simplifying the construction.[12]

Since the two financial spreads are already forward looking variables, their tracking portfolios are obtained using the spread that is concurrent with the excess returns, *i.e.* the regression specification is

$$F_{t+1} = \pi' x_{t+1} + \phi' Q_t + v_{t+1}, \tag{25}$$

[12] The alternative is to apply the model selection after partialling out Q_t from all excess returns and the macro data. However, regressing all elements of x_{t+1} on Q_t and working with the residuals introduces a lot of noise.

		Consumption (C)	Inflation (Inf)	Term Spread (TS)	Credit Spread (CS)
Target (F)		C^{12}_{t+12}	Inf^{12}_{t+12}	TS_{t+1}	CS_{t+1}
Controls (Q)		C^{12}_t, C^3_t, C^1_t	$Inf^{12}_t, Inf^3_t, Inf^1_t$	TS_t	CS_t
CV	R^2	0.06	0.06	0.06	0.12
	Sharpe	0.38	0.07	0.16	0.06
	trH*	7.92	6.37	5.79	7.99
AIC	R^2	0.14	0.16	0.20	0.25
	Sharpe	0.30	0.01	0.01	0.02
	trH	14.09	18.40	29.78	19.60
ρ(CV, AIC)		0.96	0.92	0.82	0.95

The table shows summary statistics for the tracking portfolios of non-traded factors. The target for the tracking portfolio is the variable F defined in the first line. The residuals from regressing F on the controls Q in the second line form the dependent variable for the L_2Boosting regression. Stopping time is either AIC or 5-fold cross-validation (CV). The R^2 fit refers to the partial R^2 after projecting the target on the controls; trH is the trace of the boosting-projection matrix; for the cross-validated results trH* is computed at the average optimal stopping time $L^*_{\mathcal{T}}$. Sharpe is the annualised Sharpe ratio of the tracking portfolio returns. ρ is the correlation between the returns of the AIC and CV tracking portfolio returns. Control variables for consumption and inflation are defined as $C^j_{t+k} \equiv c_{t+k} - c_{t+k-j}$ and $Inf^j_{t+k} = p_{t+k} - p_{t+k-j}$.

Table 3: Tracking portfolios for macro factors

with F_t either TS_t or CS_t. Analogously to the annual macro variables we first filter the spreads. The filter for both is an AR(1) correction.

Table 3 presents summary statistics of the tracking portfolios. For consumption the fit based on the out-of-sample fitted values in the cross-validation algorithm is $R^2 = 6\%$. That seems low, but is of the same order of magnitude as the partial $R^2 = 4\%$ reported in Lamont (2001). Direct comparison is, however, difficult due to various differences in sample and design. We use shrinkage in a regression with many assets, and estimate over a different sample period. Our fitted values are out-of-sample, which will typically reduce the reported fit. Indeed the R^2 increases to 14% with the in-sample AIC stopping. Contrary to Lamont (2001) our excess returns coincide with the first month of the annual consumption growth, which makes our fit a little better.[13]

Since the tracking portfolios are for excess returns, without restrictions on the weights, the mean and standard deviation are subject to scaling and also affected by the shrinkage of the L_2Boosting regressions. Important for their properties as an instrument is the Sharpe ratio, which scales the mean by the standard deviation. For consumption the annualised

[13] Comparison with other literature is also difficult, since many studies relate the growth in annual consumption to cumulative annual returns over the same period. Empirically the strongest correlation is between the contemporaneous excess returns and (unexpected) consumption growth. Therefore regressing annual (unexpected) growth on annual returns over the same interval will provide a better fit than regressing annual consumption just on one month returns.

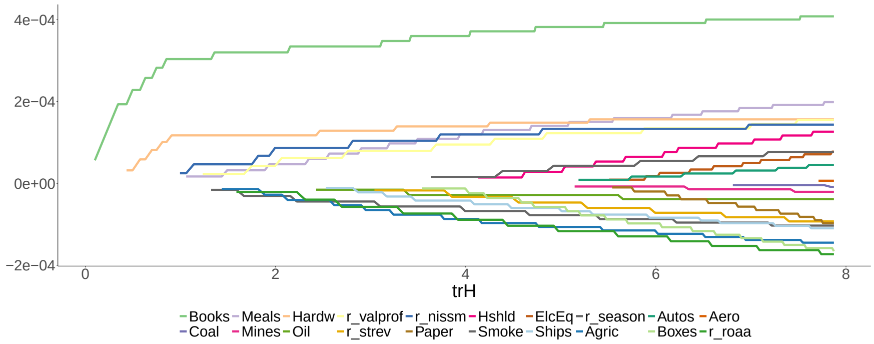
Sharpe ratio of 0.38 for the out-of-sample returns is the highest among all the estimated tracking portfolios, and close to the Sharpe ratio 0.42 for the market portfolio.

Figure 3(a) presents the consumption tracking portfolio weights and how they evolve up to the optimal stopping time. The allocation assigns the most weight to industry sorted portfolios. By far the largest weight is for the Printing and Publishing industry (`Books`), which has around twice the weight of the second to largest position. The majority of the weights, including the larger ones, are positive. The tracking portfolio appears fairly well-diversified with 20 assets included at the optimal stopping time. The weights are still subject to considerable shrinkage with a model complexity trH = 7.92, which is less than half of the full OLS weights for these 20 assets. Continuing the algorithm to the AIC stopping time the model complexity almost doubles, but the Sharpe ratio nevertheless goes down. The in-sample AIC portfolio returns are still highly correlated with the out-of-sample CV returns ($\rho = 0.96$).

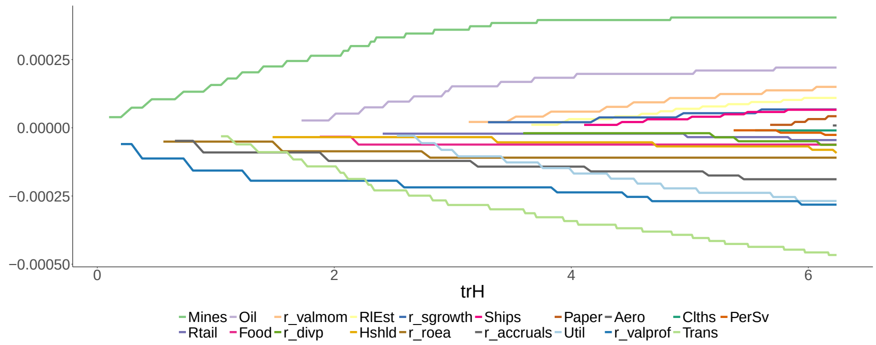
For inflation the tracking portfolio has a similar fit as for consumption, but a much lower Sharpe ratio. With the CV stopping criterion, the tracking portfolio has a Sharpe ratio close to zero. The algorithm could choose from many candidate portfolios, many of which have large Sharpe ratios that are above that of the market. Diversifying over these portfolios would easily enable a much higher Sharpe ratio, but the algorithm does not select them due to lack of correlation with inflation news. Running the algorithm until the AIC stopping time improves the fit (or overfits) but reduces the Sharpe ratio to virtually zero. An equity portfolio that hedges against inflation thus earns a zero expected excess return.

For both the term spread and the credit spread our cross-validation selects tracking allocations using a large number of assets. The CV stopping criterion identifies portfolios with an out-of-sample R^2 of 6 and 12 percent, respectively. The most prominent assets in the term spread tracking portfolio are anomaly sorts: the *Value-Profitability* and *Price* anomaly.[14] The credit spread tracking portfolio assigns a large long position to the *Utilities* industry. Similar to the inflation tracking allocation, these portfolios are composed of a mix of anomaly sorts and industries. As with most tracking portfolios the correlation between the alternative stopping rules, either CV or AIC, is large. An important difference between the two is the Sharpe ratio. Adding more complexity to the tracking portfolio increases its volatility, but not the average return. The Sharpe ratio for AIC terminated portfolios is generally lower. The effect illustrates that overfitting will reduce the mean of the tracking portfolio until ultimately it has zero mean, just as the demeaned factor itself. We see the effect for all four macro variables. In the further empirical analysis we will use the

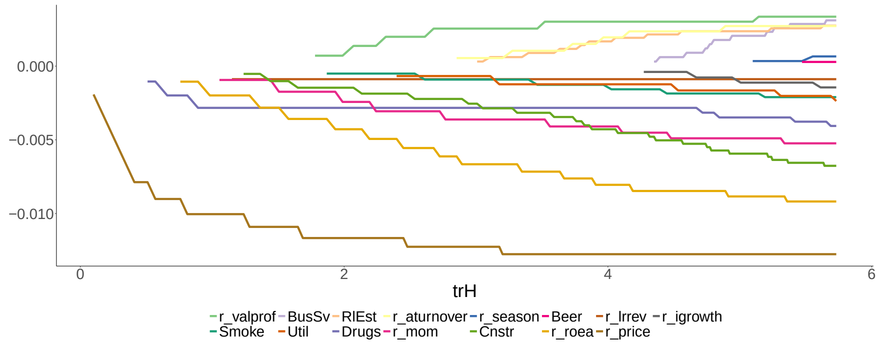
[14] See the internet appendix of Kozak, Nagel, and Santosh (2020) for definitions.



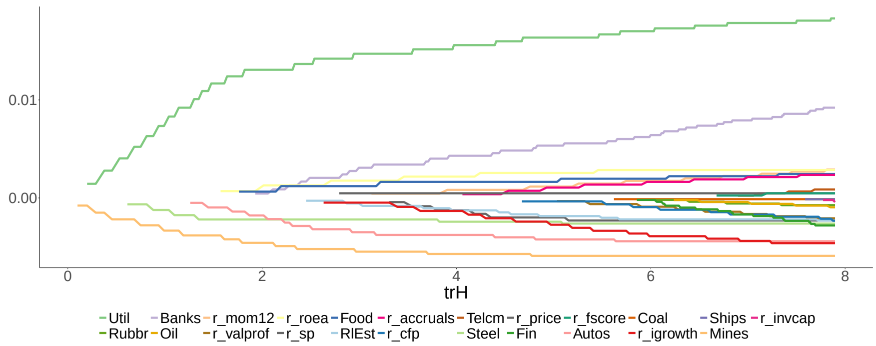
(a) Consumption



(b) Inflation



(c) Term spread



(d) Credit spread

The figure shows the evolution of the factor tracking portfolio weights as a function of the model complexity (trH). Final point on the horizontal axis is the average model complexity at the optimal stopping time implied by cross-validation. The projection is performed on the full sample. The legend lists the assets included in the tracking portfolio sorted in descending order on the final weight.

Figure 3: Tracking portfolio composition

	C		Inf		TS		CS	
Intercept	0.31	(1.4)	−0.31	(1.0)	0.22	(1.8)	0.13	(0.3)
MKT	1.03	(13.2)	0.01	(0.1)	0.08	(1.8)	0.22	(1.6)
SMB	0.52	(4.6)	0.07	(0.6)	−0.18	(2.4)	−1.45	(7.1)
HML	0.34	(2.3)	0.34	(2.1)	−0.43	(4.2)	−0.35	(1.2)
RMW	−0.17	(1.3)	1.40	(9.0)	0.28	(2.0)	0.88	(3.3)
CMA	−0.71	(3.7)	−0.11	(0.5)	−0.03	(0.2)	0.73	(1.7)
R^2	0.52		0.16		0.30		0.19	

The table shows results from regressing the macro tracking portfolio returns on the five Fama-French factors. Tracking portfolio returns have been scaled by dividing by the sum of the portfolio weights, such that they represent the excess return on a long-short portfolio. Robust t-statistics in parentheses.

Table 4: Tracking portfolios and Fama-French factors

out-of-sample tracking portfolios based on the CV stopping times as instruments.

The tracking portfolios contain all pricing information about the macro variables they are tracking. They will, however, only have independent meaning when the returns differ from what is available from common traded factors such as the five Fama-French factors. To characterise the tracking portfolios we regress their returns on the five Fama-French factors. Table 4 shows that the loadings for the tracking portfolios on the five Fama-French factors have a very distinct pattern. Consumption is mostly associated with the market and the investment factor. Inflation is primarily related to profitability — without any loading on the market—, the credit spread loads on size, while the term spread is a mix of book-to-market and size.

The tracking portfolios are correlated with the Fama-French factors, but far from perfect. The highest R^2 is 0.52 for the consumption tracking portfolio, while R^2 for the other three tracking portfolios is much lower. The low correlations can be due to genuinely different pricing information in the industry and anomaly portfolios related to the macro variables, but it could also be just noise related to the construction of the tracking portfolios. A GRS test on the intercepts rejects ($F = 2.93$, $p = 0.02$) that the Fama-French factors price the four tracking portfolios, which could be because the tracking portfolios genuinely contain relevant macro pricing information, but could also just be because some of them load on the anomaly portfolios. For now we just conclude that the tracking portfolios differ from the Fama-French factors.

For comparison table 5 reports the summary statistics for tracking portfolio for traded factors using the same set of test assets. The traded factors are the five Fama-French factors. The factors are not among the test assets, so may not be perfectly replicable (and we also do not include test portfolios sorted on the characteristics of the Fama-French factors).

		MKT	SMB	HML	RMW	CMA
	Sharpe	0.42	0.28	0.43	0.42	0.50
CV	R^2	0.97	0.83	0.85	0.77	0.77
	Sharpe	0.31	0.08	0.33	0.17	0.35
	trH*	42.42	45.80	33.68	36.87	24.58
AIC	R^2	0.98	0.87	0.89	0.82	0.82
	Sharpe	0.29	0.08	0.32	0.14	0.32
	trH	30.37	34.50	31.92	33.28	30.19

The first line reports the annualised Sharpe ratio of the factor based on the sample mean and standard deviation. Tracking portfolios are formed by projecting the demeaned Fama-French factors onto the set of excess returns using L_2Boosting. The subsequent Sharpe ratios are for the tracking portfolios based on either CV or AIC stopping.

Table 5: Traded factors tracking portfolios

Nevertheless the fit for the value-weighted market index is close to perfect with many included assets and a large model complexity. Fit for the Fama-French sorted portfolios is not that perfect, but still very strong. Most remarkable are the Sharpe ratios of the tracking portfolios. These are generally lower than those of the factors themselves.

5.3 Factor news

With monthly data the annual growth rate does not represent a news variable that can be used as an input in the SDF model. To eliminate autocorrelation, news about annual future growth for consumption and inflation is defined as the revision in expectations $(\mathrm{E}_{t+1} - \mathrm{E}_t)[F_{t+12}]$. As in Xiao et al. (2013) (and others) we construct the news factor using the two low dimensional projections

$$F_{t+12} = h_1' Z_t + \zeta_{1,t+12}, \tag{26}$$

$$F_{t+12} = h_0' Z_{t+1} + \zeta_{0,t+12}, \tag{27}$$

where $Z_t = (Q_t' \ \ \hat{f}_t)'$ is the vector of controls used for the tracking portfolio augmented by the returns on the tracking portfolio. The tracking portfolios returns are added, since they evidently have predictive power. From (26)-(27) we obtain the news factor

$$f_{t+1} = (\mathrm{E}_{t+1} - \mathrm{E}_t)\,[F_{t+12}] = h_0' Z_{t+1} - h_1' Z_t. \tag{28}$$

The coefficients h_0 and h_1 are estimated by standard OLS regression.[15] For the credit spread and term spread we already have monthly data with monthly innovations and therefore do

[15] The alternative would be a vector autoregression (VAR) for the macro variables (and possibly some returns) as for example in Petkova (2006). In a VAR, revisions in 12-month ahead expectations can be constructed by iterating on the VAR prediction equations analogously to Campbell and Vuolteenaho (2004) and others. That would link the coefficients h_0 and h_1, but at the cost of many additional parameters in a full-fledged VAR. Our specification is kept parsimonious to save on degrees of freedom.

		Tracking portfolios ($\hat{f}$)				Factors (f)			
		C	Inf	TS	CS	C	Inf	TS	CS
$\hat{f}$	Consumption	1							
	Inflation	−0.16	1						
	Term spread	0.09	−0.04	1					
	Credit spread	−0.15	−0.48	−0.37	1				
f	Consumption	0.45	−0.14	0.05	−0.05	1			
	Inflation	−0.08	0.53	0.02	−0.21	−0.23	1		
	Term spread	0.04	0.02	0.24	−0.17	0.03	0.03	1	
	Credit spread	−0.03	−0.25	−0.17	0.35	0.00	−0.21	−0.15	1

The table shows correlations constructed from the sample second moment matrix of factor news (f) and their tracking portfolio excess returns ($\hat{f}$) constructed using cross-validation. The lower left panel contains the correlations between the factor (row) and the tracking portfolio returns (column).

Table 6: Factor and instrument correlation matrix

not need the additional regressions.

Crucial for the instrumental variables estimator are the correlations between the tracking portfolio and the factors news. For consumption and inflation the correlations, reported in table 6, are 0.45 and 0.53. These numbers are substantially larger than implied by the partial R^2 in table 3. The reason is the construction of news as the revision in expectations. The factors f_t have much lower volatility than the annual growth targets for the tracking portfolios. From the correlation matrix in table 6 we also learn that correlations among the tracking portfolios generally are larger than those among the factor news variables.

5.4 Risk prices

With the estimates of economic news and their respective tracking portfolios as instruments we estimate the risk prices. Results are presented in table 7. The IV estimator produces a statistically significant estimate for the price of consumption risk. In the basic CCAPM model with constant relative risk aversion, a linear approximation of the pricing kernel is $m = 1 - \gamma C$, and thus the coefficient on consumption news equals the risk aversion parameter. The estimate implies very high risk aversion, consistent with the empirical literature.[16] Estimates for the consumption risk are stable across different specifications that add other macro variables.

To further interpret the estimate for the consumption risk price, consider the implied pricing kernel $m_t = 1 - f_t\hat{\delta}$. The pricing kernel is only identified up to a scalar multiple

[16] For example, our estimate is close the γ reported in Kroencke (2017, table III) for year on year consumption growth.

Factor	Sparse IV					Fama-MacBeth				
	C	$+Inf$	$+TS$	$+CS$	All	C	$+Inf$	$+TS$	$+CS$	All
Consumption	36.66	36.47	41.44	40.32	39.81	40.81	39.98	39.97	43.78	45.12
	(2.7)	(2.6)	(2.7)	(2.7)	(2.6)	(2.5)	(2.4)	(2.4)	(2.5)	(2.6)
Inflation		−0.51	1.62	5.66	−2.15		−19.13	−19.15	−14.16	−12.34
		(0.0)	(0.1)	(0.4)	(0.1)		(1.6)	(1.6)	(1.0)	(0.9)
Term spread			−0.61		−0.73			−0.01		0.15
			(1.3)		(1.0)			(0.0)		(0.3)
Credit spread				0.36	−0.24				0.36	0.47
				(0.7)	(0.3)				(0.9)	(1.0)

The table reports estimates of the risk prices δ in the SDF model $m = 1-\delta' f$ with t-statistics in parentheses. The left panel refers to the IV boosting estimates. The IV-boosting estimates are based on out-of-sample tracking portfolios with optimal stopping calibrated through cross-validation. The right panel reports Fama-MacBeth cross-sectional regressions of average excess returns on the sample covariances between factors and excess returns. Standard errors are computed under the Shanken correction.

Table 7: Risk price estimates for non-traded factors

(since we work with excess returns), but its volatility is still the maximum Sharpe ratio for any portfolio. As consumption news has a standard deviation of 0.66% per month, the implied annualised maximum Sharpe ratio is $\sqrt{12} \times 36.66 \times 0.0066 = 0.84$. However, that portfolio must be perfectly correlated with the factor news. By construction, the maximally correlated portfolio is the tracking portfolio, which only has a correlation of 0.45 (see table 3), and therefore implies a maximum Sharpe ratio equal to $0.45 \times 0.84 = 0.378$, equal to the Sharpe ratio for the tracking portfolio in table 3 and close to the sample Sharpe ratio for the market portfolio in table 5.

Risk prices of the other macro variables are insignificant and do not contribute to explain the cross section. The main reason is that their tracking portfolios have a much lower Sharpe ratio than the consumption tracking portfolio.

The table also contains results from a Fama-MacBeth (FM) regression. The FM estimates are based on covariances of the same candidate assets with the factors. The macroeconomic news data are identical to the time series used for the IV approach. For consumption the IV and FM estimates are similar. The only difference are the somewhat smaller standard errors for the IV estimator. The standard errors are smaller, but the difference is not as big as in the Monte Carlo simulations. The results of the IV estimates deviate from the Fama-MacBeth estimates for the other macro factors. Estimates for the inflation risk price are very different. For IV they are never significant due to the close to zero mean of the inflation tracking portfolio. The FM results are a bit more erratic, with inflation being close to significant in some specifications.

	MKT	SMB	HML	RMW	CMA
Sample	0.0539	0.0398	0.0011	0.1010	0.1260
	(0.0109)	(0.0150)	(0.0206)	(0.0209)	(0.0309)
IV-boosting	0.0438	0.0108	−0.0077	0.0549	0.1133
	(0.0111)	(0.0163)	(0.0266)	(0.0236)	(0.0420)
Fama-MacBeth	0.0606	0.0031	−0.0811	0.0889	0.2182
	(0.0110)	(0.0184)	(0.0326)	(0.0251)	(0.0522)

Risk prices in the first row are estimated from (15) treating the factors as traded assets. The second row shows IV estimates using the sparse tracking portfolios from table 5. The third row are the Fama-MacBeth estimates based on sample covariances between factor news and excess returns of test assets. In parenthesis we report standard errors.

Table 8: Risk price estimates for 5 Fama-French factors

Estimates for the Fama-French 5-factor model on the same test assets provide some further insights in the properties of the sparse IV estimator. As the Fama-French factors are traded, we can estimate the risk prices directly, using the factors themselves as instruments as in estimator (15), without requiring any test assets. By construction, we should not be able to get better estimates. Indeed, in table 8 these estimates have the lowest standard errors. As already noted by Fama and French (2015), the HML factor seems redundant. The other risk price estimates in the table are based on the test assets, without using the average returns of the five Fama-French factors. Both the IV as well as FM estimator recover the market premium, which can be readily identified from the industry portfolios. Standard errors are only marginally above those for the direct estimates in the top row of the table. For the other factors the standard errors have the predicted theoretical ranking. Direct estimates are most precise, followed by the sparse IV, while the FM estimates have largest standard errors. The point estimates are quite different. When we need to learn about the factor risk prices through the lens of the test assets, the size factor is insignificant using either IV or FM. Furthermore, the FM cross-sectional regression finds a significant role (with the wrong sign) for HML, contrary to the other estimates. The test data allow identification of the profitability and investment factors. The IV estimator tends to generate lower risk prices than FM, sometimes correctly so (as for CMA) and sometimes not (as for RMW).

5.5 Pricing errors

Projecting the fitted SDF, $1 - \hat{\delta}' f_t$ onto the test assets, the boosting algorithm stepwise builds a portfolio of mispriced assets. The fitted values are $\hat{m}_t$, from which we estimate the HJ distance.

	C	$+Inf$	$+TS$	$+CS$	All	CAPM	FF5
HJ	0.33	0.33	0.33	0.34	0.32	0.35	0.31
	(0.027)	(0.027)	(0.027)	(0.026)	(0.027)	(0.027)	(0.025)

The Hansen-Jagannathan distance (HJ) is the time series average of $\hat{m}_t^2$, where $\hat{m}_t$ are the fitted values from projecting the SDF $1 - \hat{\delta}' f_t$ on the excess returns with $\hat{\delta}$ estimated using the sparse IV estimator. The FF5 model uses the tracking portfolios of the Fama-French factors. CAPM refers to the single factor model with only the market portfolio. The column 'All' denotes the model that includes all macro factors. Standard errors are computed from a regression of $\hat{m}_t^2$ on a constant using Newey-West correction for autocorrelation.

Table 9: SDF mispricing

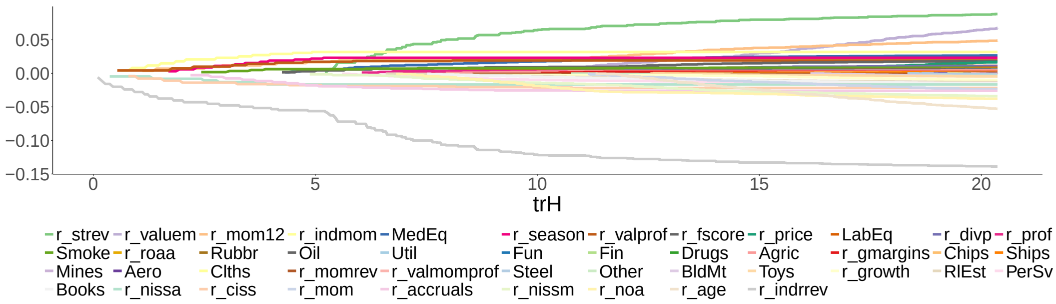
Table 9 shows that the HJ distance for all models is of a similar magnitude. In the FF5 model the risk parameters δ are estimated from the IV estimator (see table 8), and thus not restricted to the traded factor sample means. A noticeable result is that the Consumption CAPM performs better the standard CAPM with the market portfolio as the single factor. As expected, models that add inflation, credit spread and/or term spread do not make a difference relative to the consumption model. The Fama-French 5-factor model (FF5) performs slightly better than most of the macro factors on these test assets. Still, the model with all macro factors achieves the same HJ distance. The standard errors in the table probably underestimate the true uncertainty, as they do not adjust for the uncertainty in the L_2Boosting projection.

Figure 4 shows which assets are selected to explain the pricing errors by the L_2Boosting algorithm up to the optimal stopping time. The largest pricing errors are associated with anomaly characteristics. For both the CCAPM and the FF5 model, the *Industry Relative Reversal* anomaly obtains by far the largest (negative) weight. Indeed, the top 3 positive and top 5 negative weights in the mispricing portfolios are identical in both models and are all anomaly portfolios. These anomalies are consistently identified as the most difficult to price.

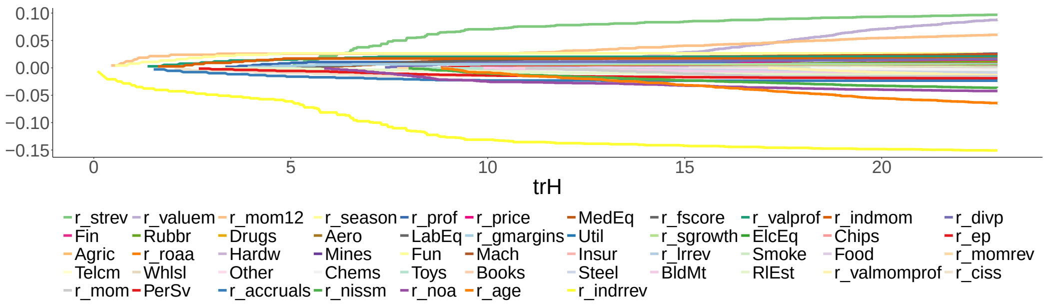
From the HJ distances it seems that the models are fairly similar in their ability to explain average returns. For a closer look at the dissimilarities we project the difference in pricing errors for two models (A and B), $m_t^A - m_t^B$, onto the set of excess returns. The boosting projection will select assets for which the mispricing greatly differs between the models. We limit ourselves to two pairwise comparisons, CCAPM - Fama-French and CCAPM - CAPM. The coefficients of the projections are presented in figure 5.

The pricing differences between the FF5 model and the CCAPM are mainly driven by how they price the *asset growth* anomaly; all other coefficients are small. For the pricing differences between the CAPM and CCAPM, the algorithm mostly selects various industry

(a) CCAPM



(b) Fama-French 5-factors



The legend is sorted on the final weight of the assets in descending order. The stopping time for the L_2Boosting algorithm is determined by cross validation. The figure shows the full sample estimates up to the average optimal stopping time.

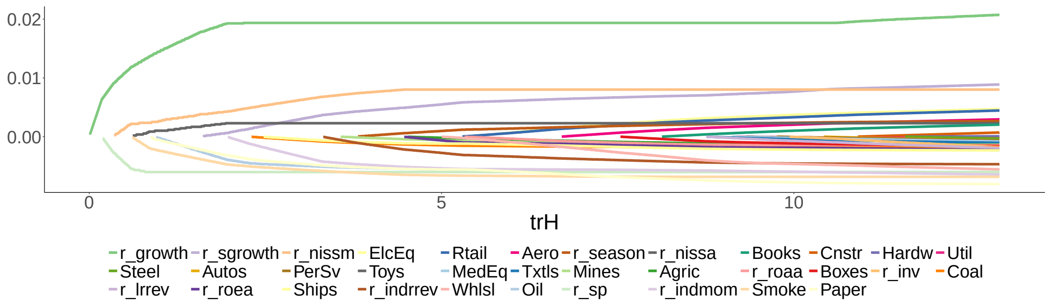
Figure 4: Pricing error portfolios

portfolios. Neither of the two sets of coefficients contains the *Industry reversal* anomaly, despite being the most mispriced asset in figure 4. Clearly, none of the models is able to price this anomaly.

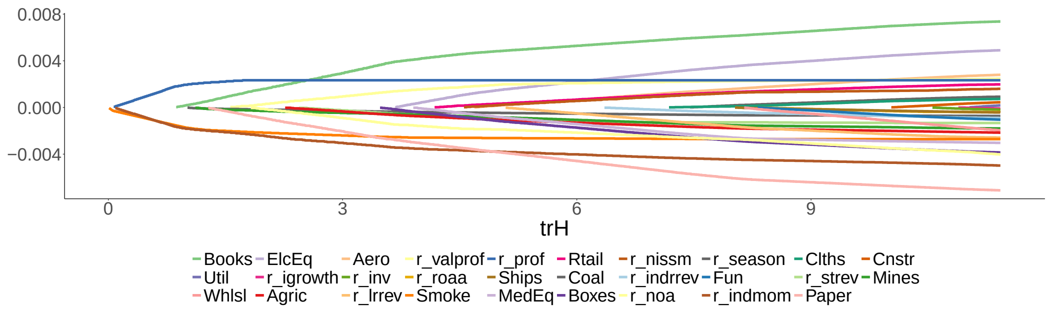
6 Conclusion

We have reformulated the problem of estimating risk prices in a stochastic discount factor model as an instrumental variables regression. For an asset pricing model that contains non-traded factors and is tested on many assets the main benefit of writing the problem in this form is that regularised regression techniques for optimal instrument selection provide a large efficiency gain relative to the two-pass Fama-MacBeth estimator. In a simulation study the IV estimator is close to the infeasible GMM estimator for short time series for factors and a large collection of test assets. The estimator approximates the GMM weighting matrix without the need to explicitly estimate a high-dimensional covariance matrix. In an empirical application the IV estimator shows that consumption is a priced factor for the cross-section of excess equity returns.

(a) CCAPM - Fama-French 5-factor



(b) CCAPM - CAPM



The figure shows the evolution of portfolio weights that explain the difference between two pricing kernels. Optimal stopping of the L_2Boosting determined by cross validation. Full sample estimates are shown up to the average optimal stopping time.

Figure 5: Pricing differences between SDF models

A similar regularised regression is used to evaluate the pricing error of the asset pricing model when there are many test assets. The stochastic discount factor is projected on all available test assets to construct a maximally mispriced portfolio. The average squared returns of the projection are an estimate of the Hansen-Jagannathan distance.

For the empirical results we implemented L_2Boosting for high-dimensional instrument selection problem for data that are strongly collinear, such as returns data. Using a large asset space of characteristics and industry sorts, the algorithm finds a consumption tracking portfolio that is strongly correlated with consumption news. Although it has a market beta close to one, it consists of a limited number of mostly industry portfolios and differs markedly from an aggregate market index. This tracking portfolio performs better than the standard CAPM in explaining the cross-section of the included tests assets. The HJ distance indicates that its pricing errors are similar to the Fama-French five factor model. Other macro-economic factor, inflation, term spread and credit spread, contribute little to performance of the SDF model for our set of test assets, but also do not affect the pricing implications of the consumption factor.

The proposed methodology appears promising for estimating and testing asset pricing models. Still, given the huge empirical literature on macro-finance asset pricing, many extensions remain unexplored. First, is the need to extend the factor space, since the few standard macro factors we considered do not explain the cross-section. Our analysis has taken the number of factors as given, while concentrating on dealing with a large number of test assets. When combined with the large literature on bringing order in the 'factor zoo' (see, *e.g* Feng, Giglio, and Xiu (2020)), we face the double regularisation problem to simultaneously deal with many assets and many factors.

Second, we have restricted the tracking portfolio weights to be constant over time. For the industry portfolios in our empirical work Fama and French (1997) already document time varying risk exposures. Extending the tracking portfolios to allow for time-varying weights, whether by rolling windows or through conditioning information (Ferson, Siegel, and Xu, 2007), will enrich the empirical contents, but also further challenge the large N regularisation. With rolling windows there will be fewer time series observations, while conditional information increases the asset space.

A third extension is moving to individual stocks. Individual stocks introduces further technical issues, for example an unbalanced panel with missing data. In its current the instrument selection algorithm cannot tackle this. Individual stocks will also greatly increase the amount of noise. With the managed portfolio sorts we could focus on assets that are informative, either for constructing a tracking portfolio or discovering mispricing.

References

Angrist, J., and A. Krueger. 1995. Split-Sample Instrumental Variables Estimates of the Return to Schooling. *Journal of Business and Economics Statistics* 13:225–35.

Aretz, K., S. Bartram, and P. Pope. 2010. Macroeconomic Characteristics and Characteristic-Based Factor Models. *Journal of Banking and Finance* 34:1383–99.

Bai, J., and S. Ng. 2009. Selecting Instrumental Variables in a Data Rich Environment. *Journal of Time Series Econometrics* 1:105–10.

Balduzzi, P., and C. Robotti. 2008. Mimicking Portfolios, Economic Risk Premia, and Tests of Multi-Beta Models. *Journal of Business and Economics Statistics* 26:354–68.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen. 2012. Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain. *Econometrica* 80:2369–429.

Bollerslev, T., N. Meddahi, and S. Nyawa. 2019. High-Dimensional Multivariate Realized Volatility Estimation. *Journal of Econometrics* 212:116–36.

Breeden, D., M. Gibbons, and R. Litzenberger. 1989. Empirical Tests of the Consumption-oriented CAPM. *Journal of Finance* 44:231–62.

Britten-Jones, M. 1999. The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights. *Journal of Finance* 54:655–71.

Bryzgalova, S. 2016. Spurious Factors in Linear Asset Pricing Models. Working paper.

Bühlmann, P. 2006. Boosting for High-Dimensional Linear Models. *Annals of Statistics* 34:559–83.

Bühlmann, P., and S. Van de Geer. 2011. *Statistics for High-Dimensional Data.* Springer.

Campbell, J., and T. Vuolteenaho. 2004. Bad Beta, Good Beta. *American Economic Review* 94:1249–75.

Chen, N., R. Roll, and S. Ross. 1986. Economic Forces and the Stock Market. *Journal of Business* 59:383–403.

Cochrane, J. H. 2005. *Asset Pricing.* Princeton University Press.

———. 2017. Macro-Finance. *Review of Finance* 21:1–42.

Fama, E., and J. MacBeth. 1973. Risk, Return and Equilibrium: Empirical Tests. *Journal of Political Economy* 71:607–36.

Fama, E. F., and K. R. French. 1997. Industry Costs of Equity. *Journal of Financial Economics* 43:153–93.

———. 2015. A Five-Factor Asset-Pricing Model. *Journal of Financial Economics* 116:1–22.

Feng, G., S. Giglio, and D. Xiu. 2020. Taming the Factor Zoo. *Journal of Finance* 75:1327–70.

Ferson, W., A. Siegel, and P. Xu. 2007. Mimicking Portfolios with Conditioning Information. *Journal of Financial and Quantitative Analysis* 41:607–35.

Giglio, S., and D. Xiu. 2021. Asset Pricing with Omitted Factors. *Journal of Political Economy* 129:1947–90.

Giglio, S., D. Xiu, and D. Zhang. 2022. Test assets and weak factors. Working paper 29002.

Gilchrist, S., and E. Zakrajsek. 2012. Credit Spreads and Business Cycle Fluctuations. *American Economic Review* 102:1692–720.

Gospodinov, N., R. Kan, and C. Robotti. 2014. Misspecification-Robust Inference in Linear Asset-Pricing Models with Irrelevant Risk Factors. *Review of Financial Studies* 27:2139–70.

Hansen, L., and R. Jagannathan. 1997. Assessing Specification Errors in Stochastic Discount Factor Models. *Journal of Finance* 52:557–90.

Harvey, C. 1993. Term Structure Forecasts Economic Growth. *Financial Analysts Journal* 49:6–8.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning.* Springer.

Jagannathan, R., and Y. Wang. 2007. Lazy Investors, Discretionary Consumption, and the Cross-Section of Stock Returns. *Journal of Finance* 52:1623–61.

Kan, R., and C. Zhang. 1999. Two-Pass Tests of Asset Pricing Models with Useless Factors. *Journal of Finance* 54:203–35.

Kelly, B., S. Pruitt, and S. Su. 2019. Characteristics are Covariances: A Unified Model of Risk and Return. *Journal of Financial Economics* 134:501–24.

Kleibergen, F. 2009. Tests of Risk Premia in Linear Factor Models. *Journal of Econometrics* 149:149–73.

Kleibergen, F., and Z. Zhan. 2020. Robust Inference for Consumption-Based Asset Pricing. *Journal of Finance* 75:507–50.

Kozak, S., S. Nagel, and S. Santosh. 2020. Shrinking the Cross Section. *Journal of Financial Economics* 135:271–92.

Kroencke, T. 2017. Asset Pricing without Garbage. *Journal of Finance* 72:47–98.

———. 2021. Robust Inference for Consumption-based Asset Pricing With Power. SSRN working paper 3562169.

Lamont, O. 2001. Economic Tracking Portfolios. *Journal of Econometrics* 105:161–84.

Ledoit, O., and M. Wolf. 2003. Improved Estimation of the Covariane Matrix of Stock Returns with an Application to Portfolio Selection. *Journal of Empirical Finance* 10:603–21.

———. 2004. Honey, I Shrunk the Sample Covariance Matrix. *Journal of Portfolio Management* 30:110–9.

———. 2017. Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks. *Review of Financial Studies* 30:4349–88.

Malloy, C., T. Moskowitz, and A. Vissing-Jørgensen. 2009. Long-Run Stockholder Consumption Risk and Asset Returns. *Journal of Finance* 64:2427–79.

McCracken, M., and S. Ng. 2015. FRED-MD: A Monthly Database for Macroeconomic Research. St. Louis FED working paper 2015-012B.

Murphy, K. 2012. *Machine Learning: A Probabilistic Perspective.* MIT Press.

Parker, J., and C. Julliard. 2005. Consumption Risk and the Cross Section of Expected Returns. *Journal of Political Economy* 113:185–222.

Petkova, R. 2006. Do the Fama-French Factors Proxy for Innovations in Predictive Variables? *Journal of Finance* 51:581–612.

Savov, A. 2011. Asset Pricing with Garbage. *Journal of Finance* 66:177–201.

Schorfheide, F., D. Song, and A. Yaron. 2018. Identifying Long-Run Risks: A Bayesian Mixed-Frequency Approach. *Econometrica* 86:617–54.

Staiger, D., and J. Stock. 1997. Instrumental Variables Regression with Weak Instruments. *Econometrica* 65:557–86.

Vassalou, M. 2003. News Related to Future GDP Growth as a Risk Factor in Equity

Returns. *Journal of Financial Economics* 68:43–74.

Xiao, Y., R. Faff, P. Gharghori, and B.-K. Min. 2013. Pricing Innovations in Consumption Growth: A Re-evaluation of the Recursive Utility Model. *Journal of Banking and Finance* 3:4464–75.

Yogo, M. 2006. A Consumption-Based Explanation of Expected Stock Returns. *Journal of Finance* 51:539–80.

Zou, H., and T. Hastie. 2005. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society B* 67:301–20.

A IV Boosting and Fama-MacBeth

The boosting algorithm produces a time series of tracking portfolio returns $\hat{f}_t$. With a single factor, the time series of the factor and the tracking portfolio are collected in the T-vectors $\boldsymbol{f}$ and $\hat{\boldsymbol{f}}$, respectively. The relation between $\boldsymbol{f}$ and $\hat{\boldsymbol{f}}$ can be written as

$$\hat{\boldsymbol{f}} = \boldsymbol{H}\boldsymbol{f}, \tag{29}$$

where $\boldsymbol{H}$ is the estimated hat matrix. The second stage IV estimator for δ is defined as

$$\hat{\delta}_{\text{IV}} = \frac{\hat{\boldsymbol{f}}'\boldsymbol{\iota}}{\hat{\boldsymbol{f}}'\boldsymbol{f}} = \frac{\boldsymbol{f}'\boldsymbol{H}'\boldsymbol{\iota}}{\boldsymbol{f}'\boldsymbol{H}'\boldsymbol{f}} \tag{30}$$

The hat matrix is defined recursively in the projection update step of the boosting algorithm as

$$\boldsymbol{H}_{\ell+1} = \boldsymbol{H}_\ell + \nu \boldsymbol{P}_{k_{\ell+1}}(I - \boldsymbol{H}_\ell) \tag{31}$$

where $\boldsymbol{P}_i$ is the univariate projection matrix $\boldsymbol{x}_i\boldsymbol{x}_i'/d_i^2$, $d_i^2 = \boldsymbol{x}_i'\boldsymbol{x}_i$, k_ℓ indexes the predictor selected at step ℓ, and $\boldsymbol{H}_\ell$ is the hat matrix after ℓ steps. Note that $\boldsymbol{H}$ is not symmetric, even though all $\boldsymbol{P}_i$ are. Using an induction argument (proof omitted) it follows that

$$\boldsymbol{H}_\ell = \sum_{i,j} a_{\ell,ij}\boldsymbol{P}_{ij}, \tag{32}$$

with $(N \times N)$ coefficient matrices $A_\ell = \{a_{\ell,ij}\}$ and data matrices $\boldsymbol{P}_{ij} = \boldsymbol{x}_i\boldsymbol{x}_j'/(d_i d_j)$. Obviously, for $\ell = 0$, we have $\boldsymbol{H}_0 = 0$, which satisfies (32) with $A_0 = 0$.

With this notation, omitting the ℓ-subscript, we can rewrite the numerator and denominator in (30) as

$$\frac{1}{T^2}\boldsymbol{f}'\boldsymbol{H}'\boldsymbol{\iota} = \sum_{i,j} a_{ij}\frac{\boldsymbol{f}'\boldsymbol{x}_j}{T}\frac{\boldsymbol{x}_i'\boldsymbol{\iota}}{T} = \sum_{i,j} a_{ij}C_j\bar{x}_i = \bar{x}'AC \tag{33}$$

$$\frac{1}{T^2}\boldsymbol{f}'\boldsymbol{H}'\boldsymbol{f} = \sum_{i,j} a_{ij}\frac{\boldsymbol{f}'\boldsymbol{x}_j}{T}\frac{\boldsymbol{x}_i'\boldsymbol{f}}{T} = \sum_{i,j} a_{ij}C_jC_i = C'AC \tag{34}$$

The vector $\bar{x}$ holds the average excess returns of the N assets, while C is the vector of N covariances of returns with the factor. The C_i are sample covariances, and not second moments, because factors have been demeaned. The IV estimator (30) is therefore equivalent to a cross-sectional regression of average returns on covariances with a weighting matrix A. This provides a direct relation with the Fama-MacBeth estimator which estimates δ from the same cross-sectional regression, but with the identity weighting matrix $A = I$, *i.e.*

$$\hat{\delta}_{FM} = (C'C)^{-1}C'\bar{x} \tag{35}$$

For the analogy with Fama-MacBeth the FM cross-sectional regression does not have a constant term. The analogy also suggests that the boosting IV estimator could be subject to a similar errors-in-variables problem as Fama-MacBeth. The covariances C_i are estimated from a finite time series and differ from their population counterparts. As is known from Jagannathan and Wang (1998) this measurement error bias will disappear for large T. The difference is that boosting selects the a_{ij} that maximise the correlation of the asset portfolio with the factor. These will typically be the largest elements C_i. Since large elements C_i will on average correspond to relatively large true covariances, the measurement error bias is less severe for the IV estimator. Also, by concentrating on the assets that have the most information on the factor, we expect to gain precision.

Even $N = 1$ is sufficient to obtain a consistent estimate, as $T \to \infty$, for δ (in this single factor model). Larger N has two opposing effects. More data generally increases the efficiency of the estimator for given matrix A, but the quadratically expanding number of elements in the weighting matrix A may lead to less efficient estimates. It is up to the boosting algorithm to find a balance in the structure for A.

When boosting overfits, the IV estimator will be biased towards zero. Overfitting means that $\hat{\boldsymbol{f}}$ becomes too close to $\boldsymbol{f}$ itself. In that case the IV estimator approaches the OLS estimator, $\hat{\delta}_{\text{OLS}} = \frac{\boldsymbol{f}'\boldsymbol{\iota}}{\boldsymbol{f}'\boldsymbol{f}} = 0$. It is identically equal to zero, since the factor has mean zero by construction.

B Monte Carlo evidence

B.1 Calibration

The text refers to five design criteria. Below we list them in detail.

1. With the normalisation $\delta = 1$ we must have $\lambda = \omega^2$. In that case ω is the maximum Sharpe ratio for any trading strategy. We take the stylised fact that the Sharpe ratio of the market portfolio is about 0.4 for annual data. In the simulations we allow for a mean-variance optimal portfolio with an annualised Sharpe ratio of 0.8. Since we simulate monthly data, we set $\omega_1 = \omega = 0.8/\sqrt{12}$.

2. Consider the equally weighted index of all stocks, such that all idiosyncratic noise has been diversified away,

$$X_I = \lim_{N\to\infty} \frac{1}{N} \sum x_i = \beta_{I,1}(f_1 + \lambda_1) + \beta_{I,2}(f_2 + \lambda_2), \tag{36}$$

   where $\beta_{I,1}$ and $\beta_{I,2}$ are the average loadings of $\beta_{1,i}$ and $\beta_{i,2}$, respectively. The variance of X_I can be decomposed as

$$\text{var}\,[X_I] = R_I^2 \beta_{I,1}^2 \omega_1^2 + (1 - R_I^2)\beta_{I,2}^2 \omega_2^2 \tag{37}$$

   where R_I^2 is the proportion of the index variance attributed to the first (priced) factor. A low value for R_I^2 generates data with a low correlation between returns and the included factor. We set $R_I^2 = 0.1$. For the index variance we use the empirical estimate $\text{var}[X_I] = (0.2)^2/12$, *i.e.* an annual volatility of 20% for an equally weighted index. The remaining common variance $(\beta_{I,2}\omega_2)^2$ is due to the omitted factor. Since only the product $\beta_{I,2}\omega_2$ matters, scaling the variance of f_2 to $\omega_2^2 = 1/12$ for monthly data is just a normalisation.

3. For the calibration of the idiosyncratic noise we look at the average variance of individual assets relative to the variance of the index. The large sample cross-sectional average variance of returns is

$$\bar{V}^2 \equiv \text{plim}\,\frac{1}{N}\sum_i \text{var}[x_i] = \text{E}[\beta_{i,1}^2]\omega_1^2 + \text{E}[\beta_{i,2}^2]\omega_2^2 + \sigma^2, \tag{38}$$

   where $\text{E}[\beta_{i,j}^2]$ is the second moment of the cross-sectional distribution of the $\beta_{i,j}$'s, *i.e.* $\text{E}[\beta_{i,j}^2] = \beta_{I,j}^2 + \Xi_j^2$, with Ξ_j the cross-sectional standard deviation of the $\beta_{i,j}$'s. Subtracting the variance of the index gives

$$\bar{V}^2 - V_I^2 = \Xi_1^2\omega_1^2 + \Xi_2^2\omega_2^2 + \sigma^2 \tag{39}$$

   Setting the average variance of individual stocks to $\bar{V}^2 = 0.3^2/12$, a 30% annual volatility, we still need two assumptions to fix Ξ_1 and Ξ_2 before we can use (39) to solve for σ^2.

4. One additional moment for calibration is the cross-sectional dispersion in expected returns,

$$s_{\bar{x}}^2 \equiv \text{plim}\,\frac{1}{N}\sum(\text{E}[x_i] - \overline{\text{E}[x_i]})^2 = \Xi_1^2\lambda^2, \tag{40}$$

   assuming $\lambda_2 = 0$. We set $s_{\bar{x}}$ is 2% annually; this determines Ξ_1.

5. Finally, to fix the cross-sectional dispersion Ξ_2 in the loadings of the omitted factor we split the remainder in (39) equally between σ^2 and $\Xi_2^2\omega_2^2$.

Table 10 summarises the Monte Carlo design parameters. The large difference between Ξ_1 and Ξ_2 implies that only very little of the cross-sectional covariance structure is due to the factor that is included in the SDF model.

Moments (annual)

Sharpe	R_I^2	sd(Y_I)	$\bar{V}$	$s_{\bar{x}}$
0.80	0.10	0.20	0.30	0.02

Derived parameters (monthly)

$\omega \equiv \omega_1$	ω_2	σ	$\lambda \equiv \lambda_1$	$\beta_{I,1}$	$\beta_{I,2}$	Ξ_1	Ξ_2
0.2309	0.2887	0.0454	ω^2	0.0791	0.1897	0.0313	0.1571

Table 10: Monte Carlo design parameters

For the baseline model we assume that the factor loadings $\beta_{i,1}$ and $\beta_{i,2}$ are independent. In each iteration of the Monte Carlo experiments both factor loadings as well as factors are random. We first draw factor loadings $\beta_{i,1}$ and $\beta_{i,2}$ given the parameters in table 10. We subsequently draw the time series of factors and noises. The factors are generated from a normal distribution. Finally we combine loadings, factors and idiosyncratic noise to generate excess returns and the SDF. Since the pricing kernel must also be positive, strictly speaking a normal distribution for f is inadmissible. The volatility ω is, however, small enough that the probability for m to become negative is negligible with monthly data.

To introduce pricing errors we set $\lambda_2 > 0$, such that the true pricing kernel is given by

$$m = 1 - \delta f - \delta_2 f_2 \tag{41}$$

with $\delta_2 = \lambda_2/\omega_2^2$. Except for λ_2 all other parameters remain as in the bottom row of table 10. With mispricing there will be a wedge between the maximum Sharpe ratio from the factor mimicking portfolio and an unrestricted mean-variance portfolio. In the baseline design without pricing errors the two Sharpe ratios are equal. Since the term $\delta_2 f_2$ increases the variance of the true pricing kernel, the resulting excess returns exhibit more cross-sectional dispersion in expected returns and to allow for a larger maximum Sharpe ratio. Values for λ_2 used in figure 2 range from zero to 0.15 in monthly units. Results in panel D in table 1 in the text and in figure 7 below are for $\lambda_2 = 0.075$.

B.2 Population properties

Written as a single factor model, and assuming correct specification with $\lambda_2 = 0$, (23) becomes

$$x = \beta(f + \lambda) + \eta, \tag{42}$$

with error covariance matrix $\mathrm{E}[\eta\eta'] \equiv \Sigma = \beta_2\beta_2'\omega_2^2 + \sigma^2 I$. Given the design (with $\lambda_2 = 0$) we can compute the projections $\hat{f} \equiv \mathrm{Proj}(f|x)$ and $\hat{1} \equiv \mathrm{Proj}(1|x)$ using the population parameters. Due to the normalisation $\delta = 1$, we have $\hat{f} = \hat{1}$. In words: the factor mimicking portfolio equals the mean-variance efficient portfolio. From the pricing condition we have

$$\mathrm{E}[fx] = \mathrm{E}[1x] = \beta\omega^2, \tag{43}$$

Next, using the matrix inversion lemma, the second moment matrix

$$\mathrm{E}[xx'] = (\omega^2 + \lambda^2)\beta\beta' + \Sigma, \tag{44}$$

has the inverse

$$\mathrm{E}[xx']^{-1} = \Sigma^{-1} - \frac{\omega^2 + \lambda^2}{1 + (\omega^2 + \lambda^2)S}\Sigma^{-1}\beta\beta'\Sigma^{-1} \tag{45}$$

with $S = \beta'\Sigma^{-1}\beta$. Using these intermediate results gives the projection coefficients for both the factor mimicking portfolio and the mean-variance efficient portfolio as

$$\pi = \mathrm{E}[xx']^{-1}\,\mathrm{E}[fx] = \frac{\omega^2}{1+(\omega^2+\lambda^2)S}\Sigma^{-1}\beta \tag{46}$$

Mean and standard deviation of the portfolio excess returns follow as

$$\mathrm{E}[\hat{f}] = \frac{\omega^2\lambda S}{1+(\omega^2+\lambda^2)S} \tag{47}$$

$$\mathrm{stdev}(\hat{f}) = \frac{\omega^2\sqrt{\omega^2 S^2 + S}}{1+(\omega^2+\lambda^2)S} \tag{48}$$

(and of course the same for $\mathrm{E}[\hat{1}]$ and $\mathrm{stdev}(\hat{1})$), from which we find the Sharpe ratio

$$\mathsf{Sh} = \omega \times \left(\frac{S}{S+1/\omega^2}\right)^{1/2}, \tag{49}$$

given that $\lambda = \omega^2$. Under our model design the quadratic form $S = \beta'\Sigma^{-1}\beta$ increases with N.[17] The limiting Sharpe ratio, as $S \to \infty$, is therefore equal to ω, while for all finite N the Sharpe ratio is strictly less than ω.

Although the two projections are the same, their fit is very different. For the MV-efficient portfolio the second moment of the dependent variable is obviously $\mathrm{E}[1^2] = 1$, while for the factor it is $\mathrm{E}[f^2] = \omega^2$. The projections have second moment

$$\begin{aligned}\mathrm{E}\left[(\pi' f)^2\right] &= \mathrm{E}[\pi' f]^2 + \mathrm{Var}[\pi' f] \\ &= \frac{\omega^4 S}{1+(\omega^2+\lambda^2)S}\end{aligned} \tag{50}$$

For large N (large S) we therefore have the measures of fit

$$R_f^2 \equiv \lim_{S\to\infty}\frac{\mathrm{E}[\hat{f}^2]}{\mathrm{E}[f^2]} = \frac{1}{1+\omega^2} \tag{51}$$

$$R_1^2 \equiv \lim_{S\to\infty}\frac{\mathrm{E}[\hat{1}^2]}{\mathrm{E}[1^2]} = \frac{\omega^2}{1+\omega^2} \tag{52}$$

In this simulation design these R^2's of the tracking portfolio regressions solely depend on the volatility of the pricing kernel, which is also the risk premium associated with the factor. By construction the tracking portfolio will never fully identify the factor, nor will it produce a fully risk free portfolio (this would violate the no-arbitrage condition). The more volatile the factor, the better the tracking performance of the mimicking portfolio. For our design parameters the fit for the tracking portfolio will converge to $R_f^2 = 0.81$. The fit for the mean-variance portfolio can therefore not be larger than $R_1^2 = 0.19$. Given the difference

[17] The two exceptions are when many assets have zero beta's and do not contain any information on the factor, and when some assets have zero idiosyncratic risk. In the first case S need not go to infinity with N, while in the latter case S will be infinite without N going to infinity.

in the signal/noise ratio a statistical learning algorithm will typically perform much better for the tracking portfolio problem than in constructing a mean-variance efficient portfolio.

For consistency of the boosting algorithm it is important that the portfolios are sufficiently sparse. The sparsity condition requires $\sum_{i=1}^{N} |\pi_i|$ to be bounded as sample size T goes to infinity while N grows exponentially in T (see Bühlmann (2006)). In the current setting with two factors we need to work out the details of S and $\Sigma^{-1}\beta$ as N becomes large. For Σ we have the inverse

$$\Sigma^{-1} = \frac{1}{\sigma^2}\left(I - \frac{\omega_2^2}{\sigma^2 + \omega_2^2 \beta_2' \beta_2} \beta_2 \beta_2'\right) \tag{53}$$

As a result, by simply substituting and simplifying,

$$\bar{S} = \text{plim}_{N\to\infty} \frac{1}{N} S = \frac{1}{\sigma^2}\left(\mathrm{E}[\beta_i^2] - \frac{\mathrm{E}[\beta_i \beta_{2,i}]^2}{\mathrm{E}[\beta_{2,i}^2]}\right) \tag{54}$$

When β_i and $\beta_{2,i}$ are independent, we have that $\mathrm{E}[\beta_i\beta_{2,i}] = \beta_I \beta_{I,2}$. Therefore, $\bar{S} > 0$ as long as $\mathrm{E}[\beta_i^2] > 0$. Moreover, all elements

$$\left(\Sigma^{-1}\beta\right)_i = \frac{1}{\sigma^2}\left(\beta_i - \frac{\omega_2^2 \beta' \beta_2}{\sigma^2 + \omega_2^2 \beta_2' \beta_2} \beta_{2,i}\right)$$

have finite expected absolute value. Since the denominator S in (46) is of order N, the sum over $|\pi_i|$ will be bounded in N, and hence satisfies the Bühlmann condition.

As a benchmark for the efficiency of the IV estimator we consider the case where the optimal instrument is known, *i.e.* we have the instruments

$$\hat{f}_{\text{opt}} = \pi' x, \tag{55}$$

with π the population tracking portfolio weights defined in (46). The resulting estimator $\hat{\delta}_{\text{opt}} = \frac{\boldsymbol{\iota}' \boldsymbol{X} \pi}{\boldsymbol{f}' \boldsymbol{X} \pi}$ is infeasible in practice, but it provides an upper bound on the precision of the actual IV estimator.

B.3 In-sample versus Out-of-Sample

The first use of the simulation design is to evaluate the difference between in-sample and out-of-sample fits of the tracking portfolios. In section 3 we propose to implement the IV estimator using the cross-validation predictions $\hat{\boldsymbol{F}}_k = \boldsymbol{X}_k \hat{\Pi}_{-k}$ for each subsample k with $\hat{\Pi}_{-k}$ the estimator of the tracking portfolio weights using the training data $(\boldsymbol{F}_{-k}, \boldsymbol{X}_{-k})$. The alternative is to use cross-validation to determine the optimal stopping, and then run the boosting algorithm one more time on the full sample $(\boldsymbol{F}, \boldsymbol{X})$ with the optimal stopping to obtain the final estimates $\hat{\Pi}$ and tracking portfolio returns $\hat{\boldsymbol{F}}$. We use Monte Carlo to illustrate that the latter results in a strong downward bias in the estimated δ when T and N are of similar magnitude. Using the design in table 10 we simulate both estimators to

N	T	Estimator	ave	std	quantiles 1%	25%	med	75%	99%
200	120	Out-of-sample	1.01	0.47	0.01	0.69	0.98	1.30	2.27
		In-sample	0.52	0.28	0.01	0.32	0.49	0.70	1.26
200	600	Out-of-sample	1.00	0.21	0.58	0.86	1.00	1.14	1.51
		In-sample	0.82	0.17	0.47	0.71	0.82	0.93	1.25
200	1200	Out-of-sample	1.00	0.13	0.73	0.91	1.00	1.09	1.30
		In-sample	0.89	0.11	0.64	0.82	0.90	0.98	1.13
20	120	Out-of-sample	1.03	0.58	-0.19	0.63	1.00	1.39	2.54
		In-sample	0.84	0.46	-0.17	0.54	0.82	1.13	1.98

The table reports averages, standard deviations, and quantiles for alternative estimators for δ in the discount factor model $m = 1 - \delta f$ with true value $\delta = 1$. Simulated data are generated for N test assets and T time series observations. *Out-of-sample* uses the cross-validated predictions and is identical to the design in panel A of table 1; *In-sample* denotes the Instrumental Variables estimator using a tracking portfolio for f obtained by boosting on the full sample of T observations with the same number of boosting iterations. Statistics are from 10,000 replications.

Table 11: In-sample versus out-of-sample tracking portfolio returns

illustrate the difference. As benchmark we set $(N, T) = (200, 120)$. We then vary N and T to show the dependence on the sample size. The tracking portfolio is estimated by boosting using stepsize $\nu = 0.1$ and 5-fold cross-validation to determine the optimal stopping.[18]

In table 11 the in-sample estimator for the benchmark has a mean of only 0.52 instead of the true value equal to one. The out-of-sample estimator is nearly unbiased with the same sample sizes. The bias slowly diminishes if either T grows bigger, or N decreases.

The intuition for the bias reduction is similar to what motivated the split sample IV estimator in Angrist and Krueger (1995) and its large N counterpart in Belloni et al. (2012). In both papers the sample is split in two parts, whereas we exploit the sample splits we already have for computing the cross-validated fit. Below we illustrate the intuition for the bias, and its correction, in the simple case of unpenalised regressions in a setting with $N < T$. The tracking portfolio from unrestricted least squares in the first stage regressions

$$\boldsymbol{F} = \boldsymbol{X}\Pi' + \boldsymbol{V}, \tag{56}$$

yields (without regularisation),

$$\hat{\boldsymbol{F}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{F} = \boldsymbol{X}\Pi' + \boldsymbol{H}\boldsymbol{V} \tag{57}$$

The crucial element in estimating δ is the cross-product $\hat{\boldsymbol{F}}'\boldsymbol{F}$,

$$\hat{\boldsymbol{F}}'\boldsymbol{F} = \Pi\boldsymbol{X}'\boldsymbol{X}\Pi' + \Pi\boldsymbol{X}'\boldsymbol{V} + \boldsymbol{V}'\boldsymbol{X}\Pi' + \boldsymbol{H}\boldsymbol{V} \tag{58}$$

[18] All computations have been done in R. For boosting we use the function glmboost from package mboost where the options (`offset=0, center=F`) suppress the constant from the regression function.

Denoting the sample second moment matrix by $\boldsymbol{M} = \boldsymbol{X}'\boldsymbol{X}/T$, and taking expectations conditional on the instruments $\boldsymbol{X}$, assuming the orthogonality $\mathrm{E}[v_t x_t'] = 0$, we have

$$\mathrm{E}\left[\frac{1}{T}\hat{\boldsymbol{F}}'\boldsymbol{F}|\boldsymbol{X}\right] = \Pi\boldsymbol{M}\Pi' + \frac{N}{T}\Psi, \tag{59}$$

where $\Psi = \mathrm{E}[v_t v_t']$ is the second moment matrix of the regression errors. The second term creates a bias. Of course, for fixed N and growing T the bias becomes negligible relative to the first term, but for T and N of similar magnitudes the bias can be substantial.

No such bias exists for the cross-product $\hat{\boldsymbol{F}}'\boldsymbol{\iota}$, since

$$\mathrm{E}\left[\frac{1}{T}\hat{\boldsymbol{F}}'\boldsymbol{\iota}|\boldsymbol{X}\right] = \mathrm{E}\left[\left(\frac{1}{T}\boldsymbol{F}'\boldsymbol{X}\right)\boldsymbol{M}^{-1}\left(\frac{1}{T}\boldsymbol{X}'\boldsymbol{\iota}\right)|\boldsymbol{X}\right] = \Pi\bar{x} \tag{60}$$

for $\bar{x} = \frac{\boldsymbol{X}'\boldsymbol{\iota}}{T}$. The magnitude of the bias in the IV estimator (13) depends on further distributional assumptions, because the bias term appears in the denominator and $\mathrm{E}[(\hat{\boldsymbol{F}}'\boldsymbol{F})^{-1}] \neq \left(\mathrm{E}[\hat{\boldsymbol{F}}'\boldsymbol{F}]\right)^{-1}$. The bias will be towards zero, since the additional term $\frac{N}{T}\Psi$ is positive definite.

The bias can be greatly reduced by using the cross-validated tracking portfolio returns. These out-of-sample fitted values are

$$\hat{\boldsymbol{F}}_k = \boldsymbol{X}_k(\boldsymbol{X}_{-k}'\boldsymbol{X}_{-k})^{-1}\boldsymbol{X}_{-k}'\boldsymbol{F}_{-k} = \boldsymbol{X}_k\hat{\Pi}_{-k} \tag{61}$$

(or other linear estimators for $\hat{\Pi}_{-k}$ that so not depend on $\boldsymbol{F}_k$). For each fold k we then have

$$\hat{\boldsymbol{F}}_k'\boldsymbol{F}_k = \boldsymbol{F}_{-k}'\boldsymbol{X}_{-k}(\boldsymbol{X}_{-k}'\boldsymbol{X}_{-k})^{-1}\boldsymbol{X}_k'\boldsymbol{F}_k, \tag{62}$$

which is unbiased conditional on $\boldsymbol{X}$ due to the time series independence of the blocks $\boldsymbol{F}_k$ and $\boldsymbol{F}_{-k}$ and the demeaned factors.

The out-of-sample rows in table 11 illustrate that the cross-validated tracking portfolios lead to almost unbiased estimates of δ. For this reason, and in light of the analytical intuition, all our empirical estimates use the cross-validated tracking portfolio returns.

B.4 Mispricing

With mispricing the HJ distance will be positive. Its value as a function of λ_2 is found by simply calculating $\mathrm{E}[mx]$ and $\mathrm{E}[xx']^{-1}$ that enter the HJ distance. The true expected excess returns are $\mathrm{E}[x] = \beta\lambda + \beta_2\lambda_2$, which implies that pricing errors under the misspecified model $m = 1 - \delta f$ (omitting the second priced factor) are

$$\mathrm{E}[mx] = \mathrm{E}[x] - \delta\,\mathrm{E}[xf] = \beta(\lambda - \delta\omega^2) + \beta_2\lambda_2 \tag{63}$$

Differentiating HJ with respect to δ, the minimum HJ obtains when δ satisfies

$$\lambda - \delta\omega^2 = -\lambda_2\frac{\beta'\,\mathrm{E}[xx']^{-1}\beta_2}{\beta'\,\mathrm{E}[xx']^{-1}\beta} \tag{64}$$

The expression shows that with misspecification the implied price of risk is no longer equal to $\delta = \lambda/\omega^2$ unless $\beta' \,\mathrm{E}[xx']^{-1}\beta_2 = 0$. The HJ-distance is equal to

$$\mathrm{HJ} = \lambda_2^2 \left(\beta_2' \,\mathrm{E}[xx']^{-1}\beta_2 - \frac{(\beta' \,\mathrm{E}[xx']^{-1}\beta_2)^2}{\beta' \,\mathrm{E}[xx']^{-1}\beta} \right) \tag{65}$$

It depends on λ_2 in a complicated way. Apart from the leading λ_2^2 term, there is a further nonlinear dependence through $\mathrm{E}[xx']$. That dependence would disappear when we would use covariances instead of second moments in the definition of the HJ-distance.[19] For a closer inspection we explicitly compute

$$\mathrm{E}[xx'] = BJ^{-1}B' + \sigma^2 I, \tag{66}$$

where we introduce the $(N \times 2)$ matrix $B = (\beta\ \beta_2)$ and the (2×2) matrix

$$J^{-1} = \begin{pmatrix} \omega^2 + \lambda^2 & \lambda\lambda_2 \\ \lambda\lambda_2 & \omega_2^2 + \lambda_2^2 \end{pmatrix} \tag{67}$$

By the matrix inversion lemma,

$$\mathrm{E}[xx']^{-1} = \frac{1}{\sigma^2} \left(I - \frac{1}{\sigma^2} B(G + J)^{-1})B' \right), \tag{68}$$

with $G = B'B/\sigma^2$ and thus

$$B' \,\mathrm{E}[xx']^{-1}B = G - G(G + J)^{-1}G = \left(G^{-1} + J^{-1}\right)^{-1}, \tag{69}$$

In general, neither G nor J is diagonal, so the off-diagonal element $\beta' \,\mathrm{E}[xx']^{-1}\beta_2$ will be non-zero, and therefore δ will be affected by the mispricing. Even in the special case of an orthogonal design, G is diagonal, but J is not. Elements of the matrix G will generally be of order N, whereas those in J do not depend on N. For large N we have the approximation

$$\lim_{N\to\infty} B' \,\mathrm{E}[xx']^{-1}B = J, \tag{70}$$

which leads to simple limiting expressions for the HJ distance and δ. In the limit, as $N \to \infty$, equations (64) and (65) become

$$\mathrm{HJ} = \lambda_2^2 \left(J_{22} - \frac{J_{21}^2}{J_{11}} \right) = \frac{\lambda_2^2}{\lambda_2^2 + \omega_2^2} \tag{71}$$

$$\delta = \frac{1}{\omega^2} \left(\lambda + \lambda_2 \frac{J_{21}}{J_{11}} \right) = \frac{\lambda}{\omega^2} \times \frac{\omega_2^2}{\omega_2^2 + \lambda_2^2} \tag{72}$$

[19] See our earlier discussion in footnote 4. Giglio and Xiu (2021, sect II.A) analyse exactly the same two-factor model. They conclude that the omitted variables bias – from leaving out the second factor – does not affect the estimated risk premium if all assets are included. That is the case we consider for our population properties with $N \to \infty$. The infinite N assumption is essentially the same as their assumption that idiosyncratic risk vanishes in the limit. We still get a small bias in δ because we estimate the factor mimicking regression without a constant term. The difference between covariances and second moments is negligible when returns are observed at a high enough frequency.

The final equalities in (71)-(72) follow by explicitly inverting (67). Because of (70) neither depends on the properties of the factor loadings.[20] Both the limiting HJ distance as well as the risk price that minimises the HJ distance depend on the mispricing λ_2. As explained before in footnote 19, the only reason for the dependence on λ_2 in (72) is the use of second moments instead of covariance in the HJ distance. With $\mathrm{Var}(x)$ instead of $\mathrm{E}[xx']$ the expressions would further simplify to $\mathrm{HJ} = \lambda_2^2/\omega_2^2$ and $\delta = \lambda/\omega^2 = 1$.

Properties of the FM estimator are very different. For the cross-sectional regression of expected average returns on covariances we have

$$\delta_{FM} = \frac{\mathrm{E}[xf]'\,\mathrm{E}[x]}{\mathrm{E}[xf]'\,\mathrm{E}[xf]}, \tag{73}$$

which can be written, analogously to expression (64) for the IV estimator, as

$$\lambda - \delta_{FM}\omega^2 = -\lambda_2 \frac{\beta'\beta_2}{\beta'\beta} \tag{74}$$

Unlike the IV estimator, the FM estimator for δ will depend on properties of the factor loadings, even for $N \to \infty$. Unless the loadings of the second factor are orthogonal to the included factor, the estimate will be asymptotically biased. This effect shows up in table 1 for the misspecified model with independent factor loadings. Factor loadings of a completely independent, but priced, factor matter. For the Monte Carlo design we can eliminate the bias by adding a constant term in the cross-sectional regression as is often done in empirical studies. But eliminating the bias will increase the variance of the estimator due to multicollinearity between the constant term and the factor loadings.

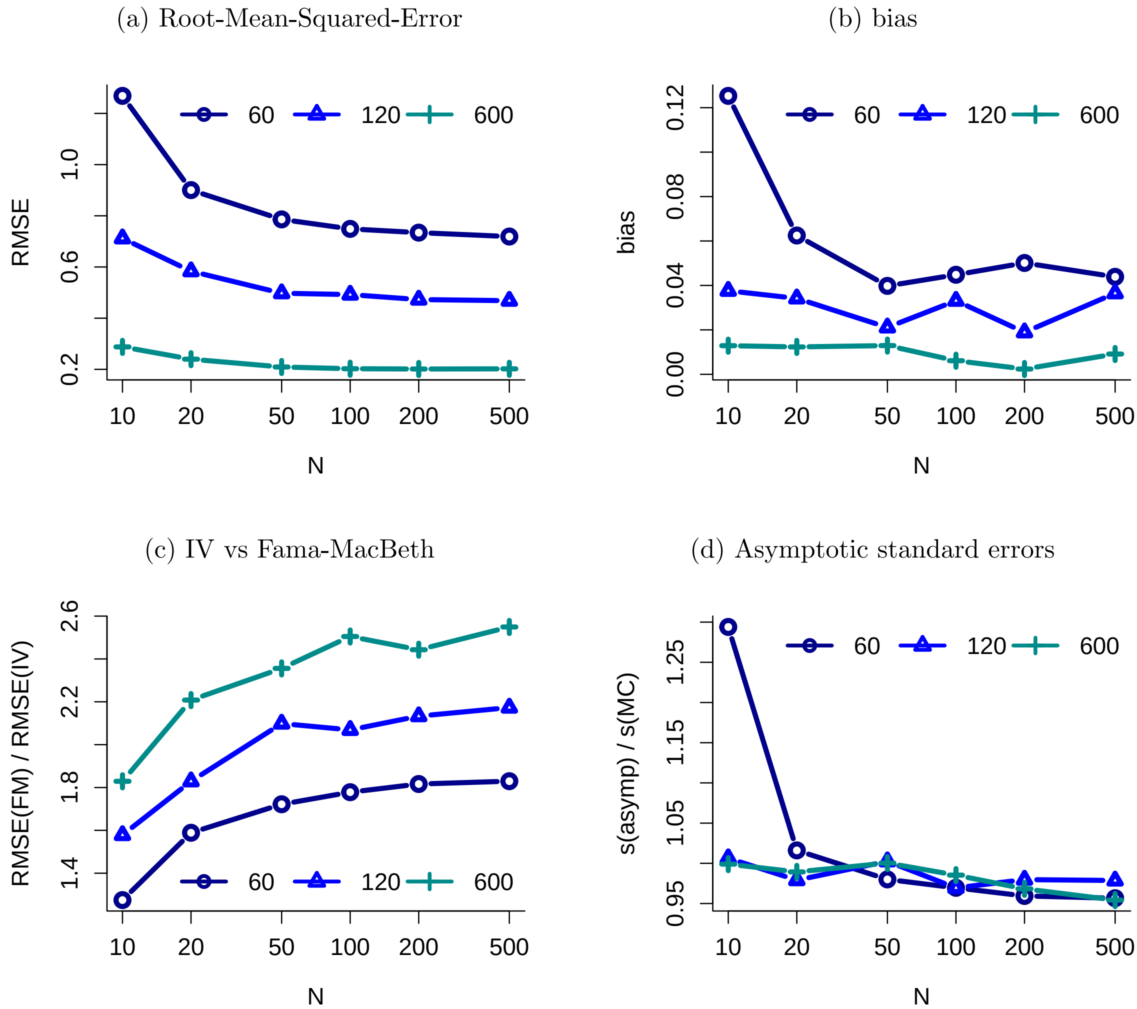
B.5 Results for different N and T

In the text we discuss the case $(N, T) = (200, 120)$. Here we consider different combinations of N and T. Using the design settings for the correctly specified model from panel (A) in table 10, figure 6(a) shows the Root-Mean-Squared-Error (RMSE) for $\hat{\delta}$ for given T as a function of N. Obviously, with larger T the error decreases for every N. Similarly, for given sample size T, enlarging the cross section improves the estimate, but reaches a saturation point for large N.[21] The bias, shown in figure 6(b) remains small for all (N, T) combinations except when both are very small.

Figure 6(c) compares the IV estimator to Fama-MacBeth. The larger N the larger the gain in efficiency from using the IV estimator. For small N and T the IV estimator achieves

[20] The first equality in (71) holds more generally in case of factor mispricing. With M priced factors of which M_2 are omitted in the pricing model, the general expression is $\lambda_2'(J_{22} - J_{21}J_{11}^{-1}J_{12})\lambda_2$ with J_{ij} blocks in the $(M \times M)$ matrix $J = (\Omega + \lambda\lambda')^{-1}$ and Ω block diagonal, *i.e* the omitted factors are independent of the included factors.

[21] For $T = 60$ it sometimes (less then 1% of replications) happens that the cross-validation cannot find any asset that has an out-of-sample correlation with the factor, such that the tracking portfolio is equal to zero. Evidently the IV estimator is not defined for such a clear rank-deficient portfolio. The sampling distributions are thus censored by excluding the cases for which L_2Boosting sets $\hat{f} = 0$.

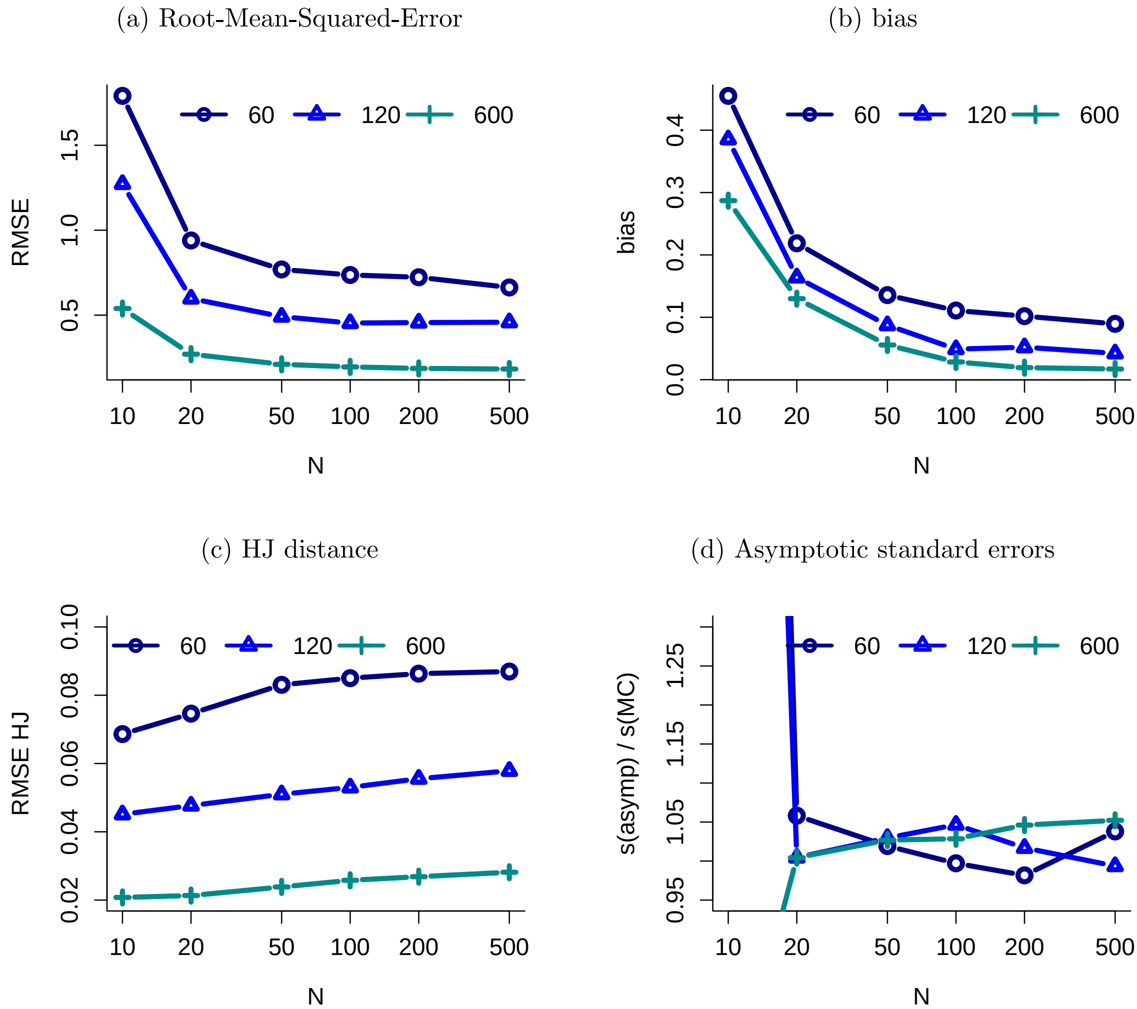


The figure shows Monte Carlo results for the instrumental variable estimator $\hat{\delta}$ where the first stage regression is performed by L_2Boosting. The upper left panel shows the bias relative to the true value $\delta = 1$; the upper right panel shows the Root-Mean-Squared-Error. The lower left panel compares the IV and FM estimators. The lower right panel compares the asymptotic and Monte Carlo standard errors for the IV estimator. The horizontal axis is the cross-sectional sample size N on a log scale. The different curves refer to different sample sizes $T = (60, 120, 600)$.

Figure 6: Monte Carlo results: IV estimator with correct specification

about 20% higher precision than the Fama-MacBeth estimator. The gain quickly increases to more than double the FM precision for larger (N, T). Finally, panel 6(d) shows that the asymptotic standard errors are reliable, except when N and T are both small.

Figure 7 shows similar plots related to the design in panel D in table 10, which analyses the effect of pricing errors. With small N the risk price estimator has substantial bias, which quickly diminishes as N becomes large. With mispricing we therefore require a reasonably large cross-section. The variance of the IV estimator is not much affected by the mispricing. The RMSE in figure 7(a) is of similar magnitude as it was in figure 6(a). An important result is the quality of the asymptotic standard errors under mispricing. They have been computed with the standard formula without any adjustment for the misspecification as in



Panels (a), (b) and (d) are similar to figure 6 but now for data generated with $\lambda_2 = 0.075$. The bias is with respect to the true value in (72) Panel (c) shows the RMSE of the HJ estimates. The y-axis in the lower right panel (d) has been truncated, since results for $N = 10$ are far off the normal scales.

Figure 7: Monte Carlo results: IV estimator with pricing errors

Gospodinov, Kan, and Robotti (2014), but even so they are still very close to the Monte Carlo standard errors. Only for small N they are completely off. The precision of the HJ estimates is not much affected by N, but improves steadily with increasing T.

C Sparse Mean-Variance Portfolio

As part of testing the performance of the boosting algorithm on noisy return data we apply the algorithm to the same example data set that Kozak, Nagel, and Santosh (2020, KNS) use to test their regularised regressions based on economically motivated priors. They estimate an SDF based on daily excess returns of the 25 Fama-French portfolios sorted on Size and Book-to-Market from July 1926 to December 2017. We represent that problem as the projection of excess returns on a vector of ones, as in (17), written in data notation as

$$\boldsymbol{\iota} = \boldsymbol{X}\pi_1 + \boldsymbol{v}_1, \tag{75}$$

where $\boldsymbol{X}$ contains the excess returns orthogonalised with respect to the 'value-weighted index return using β's estimated in the full sample' (KNS, p 280).

With '1' as the dependent variable the parameters in (75) are weights of a mean-variance efficient portfolio. Constructing a mean-variance efficient portfolio is notoriously difficult, especially when the cross-sectional dimension N is large relative to the time series sample size T. Dangers of overfitting when N is large and cross-correlations are substantial have been pointed out many times, see the references in KNS. Many shrinkage and dimension reduction techniques have been suggested to obtain portfolios with reasonable out-of-sample performance. L_2Boosting, as far as we know, new in this respect.

KNS explain that the OLS estimator $\hat{\pi}_1 = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\iota}$ has very poor out-of-sample performance for the mean-variance portfolio due to the noisy sample means $\frac{1}{T}\boldsymbol{X}'\boldsymbol{\iota}$, even with such long time series.[22] They suggest an Elastic Net estimator motivated by economic arguments. A maximum Sharpe ratio motivates L_2-norm shrinkage on π_1, while the strong factor structure in the data suggests an L_1-norm sparsity with many exact zeros.

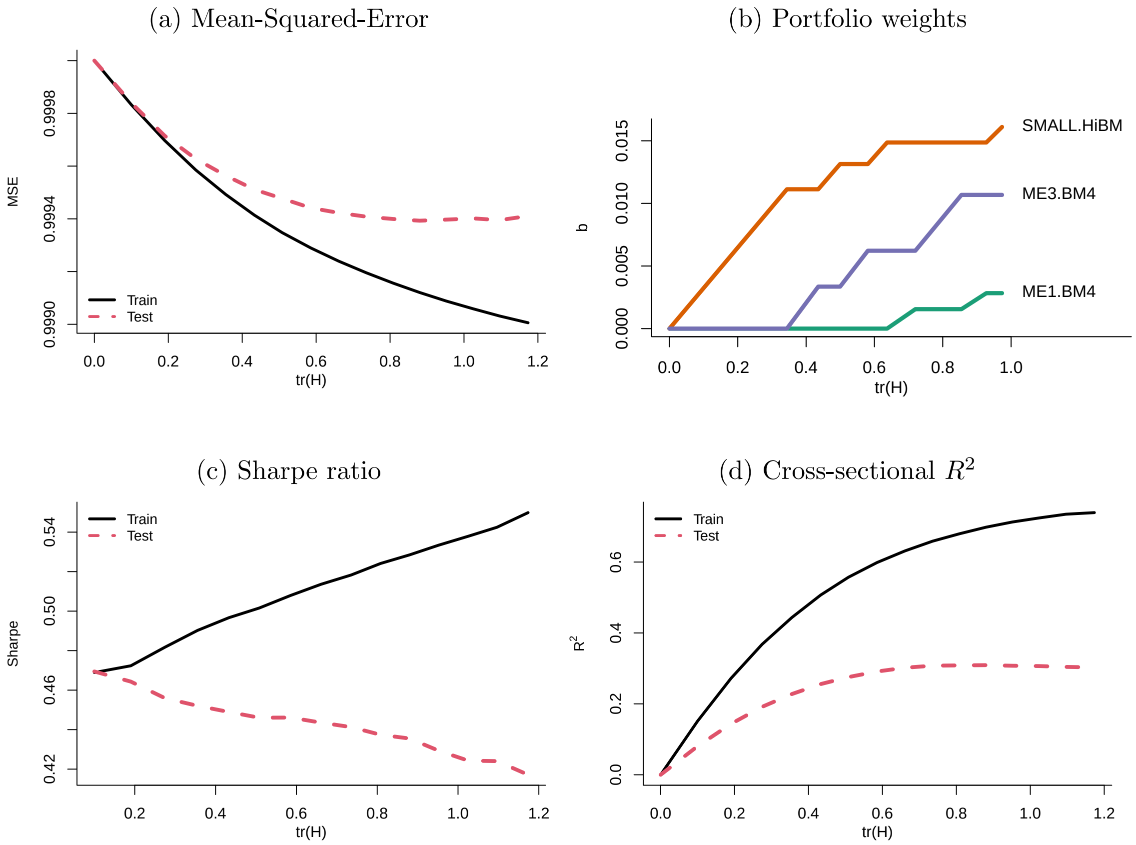
Applying L_2Boosting to (75) should have the same effect and lead to a very similar portfolio. We apply boosting on the same data set with standard tuning parameter $\nu = 0.1$ and 5-fold cross-validation. Figure 8 shows results for the first 15 steps of the algorithm.

Figure 8(a) shows that the cross-validated residual sum of squares reaches a saturation point at a model complexity of $\operatorname{tr}\boldsymbol{H} = 0.9$, after 10 iterations of the algorithm. In contrast the training fit improves steadily by construction. Figure 8(b) shows that the solution is sparse. At the optimal stopping time just three of the 25 portfolios obtain a non-zero weight, exactly as in KNS. Most prominent is the SMALL_HIBM portfolio, the most notorious outlier among the 25 portfolios.

One of the important lessons is the difference between in-sample and out-of-sample performance. In figure 8(c) the in-sample Sharpe ratio increases steeply while increasing the complexity of the portfolio. Out-of-sample, the Sharpe ratio already starts to decline after the first step. L_2Boosting minimises the out-of-sample residual sum of squares $\hat{\boldsymbol{v}}_1'\hat{\boldsymbol{v}}_1$ (shown in panel (a), which is not the same as maximising the out-of-sample Sharpe ratio in panel (c). Therefore the algorithm does not stop after the first step, when the Sharpe ratio starts to decrease.The out-of-sample Sharpe ratio in the test samples hovers around between 0.41 and 0.47, the same order of magnitude as the Sharpe of the market portfolio.

[22] The matrix $\frac{1}{T}\boldsymbol{X}'\boldsymbol{X}$ contains second moments. The choice between second moments or covariances does not matter for the mean-variance problem. Apart from an arbitrary scaling constant, the two will lead to the same mean-variance portfolio weights (Britten-Jones, 1999).

Figure 8: Mean-variance analysis of the 25 Fama-French Size/BtM portfolios



Horizontal axis in panels (a)-(d) is the model complexity with increasing number of boosting iterations in regression (75). Panel (a) displays the time series Mean-Squared-Error $\hat{\boldsymbol{v}}_1'\hat{\boldsymbol{v}}_1/T$ both in-sample ('training') as well as out-of-sample ('test'). Panel (b) shows the estimated portfolio weights. As the weights are for excess returns, scaling is indeterminate (multiplying by any positive constant will give the same Sharpe ratio). The Sharpe ratio for the portfolio $\hat{\pi}_1'x$ is in panel (c), again both in-sample and out-of-sample. Panel (d) shows the implied cross-sectional R^2 for the regression of sample means on the sample covariances using the estimated coefficients $\hat{\pi}_1$ from the time series regression. In panels (a), (c) and (d) the black solid lines are based on a random training sample containing 80% of the data used to estimate the parameters $\hat{\pi}_1$. The red dashed lines apply the estimated $\hat{\pi}_1$ to the complementary test samples. Coefficients in panel (b) are from running boosting on the full sample. The procedure is repeated 100 times. Figures shows averages over these 100 replications.

KNS evaluate their model based on its cross-sectional fit, derived from the regression

$$\boldsymbol{X}'\boldsymbol{\iota} = \boldsymbol{X}'\boldsymbol{X}\pi_1 + \eta, \tag{76}$$

with $N = 25$ observations on the average returns, and N regressors formed by the return covariances. Using the estimated $\hat{\pi}_1$ from (75), we evaluate the cross-sectional R^2 in figure 8(b). The results imply a very similar fit as reported by KNS in their figures 1 and 2: out-of-sample the cross-sectional R^2 is around 0.35. Consistent with the time series regression, the out-of-sample cross-sectional R^2 starts to slowly decline after 10 iterations.

The boosting regressions can replicate the results in Kozak, Nagel, and Santosh (2020) for these 25 FF portfolios with simple default settings for the L_2Boosting regressions.