Empirical asset pricing with many assets and short time series

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

June 15, 2018

Abstract: We develop a new method for estimating the prices of macroeconomic risk factors in a stochastic discount factor model using a large cross-section of asset returns. The first step of our approach consists of constructing tracking portfolios for the macroeconomic factors. For this we use a boosting algorithm to project a time series of macro news on a large number of return time series. The method is designed for a setting where the cross-sectional dimension is large compared to the length of the return time series. In a second stage we use the tracking portfolio returns as instruments in a time series instrumental variables regression to estimate risk prices. The same learning algorithm also provides the weights of a mean-variance efficient portfolio. With this additional input we compute the Hansen-Jagannathan distance to compare how alternative models fit the cross-section of returns.

We apply the method to monthly data for 900 portfolio return series in the Kenneth French data library. We find that both consumption as well as inflation risk are priced.

Keywords: Boosting; Asset Pricing Tests; Hansen-Jagannathan Distance

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

[†] Department of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands.
Email: `s.lonn@maastrichtuniversity.nl`

[‡] Department of Finance, Maastricht University and NETSPAR, P.O. Box 616, 6200 MD Maastricht, The Netherlands.
Email: `p.schotman@maastrichtuniversity.nl`

[§] We thank conference participants at SOFIE 2017 (New York), ESEM 2017 (Lisbon) and the Conference on the Econometrics of Financial Markets (Stockholm), for helpful comments and suggestions on an early draft.

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 recurring problem in empirical asset pricing is the need to invert a large-dimensional covariance (or second moment) matrix. For example, in a standard time series based asset pricing test a small number of factors is proposed to explain the differences in expected returns across a large number of assets. The null hypothesis is that Jensen's alpha should be zero for all N test assets. But doing the test when N is large relative to the length of the series T, or as in our application even bigger than T, is problematic. The N-dimensional covariance matrix is also at the heart of mean-variance portfolio construction.

Since naive estimation does not work, various methods have been developed to work around the problem by carefully selecting portfolios or building factors. The practice of building portfolio has by now produced so many portfolios (and factors) that the number of interesting portfolios by itself has become a big-N problem. In this context Cochrane (2011) coined the term 'factor zoo' referring to the literally hundreds of interesting portfolio sorts that may be informative on which factors are priced.

We will formulate and interpret standard asset pricing tests in a framework that allows application of statistical learning techniques to deal with the big-N problem. Many statistical learning techniques are concerned with predicting an output variable when there are many potential explanatory variables. Since they allow the number of variables N to grow very large, explicit inversion of an N-dimensional moment matrix to compute a standard least squares regression will be infeasible. A useful solution is to assume that many of the regression coefficients are zero (without a priori knowing which) or that coefficients are mostly small. In our approach we write the asset pricing model explicitly as a regression problem in order to apply machine learning techniques.

Our approach builds on an insight of Hansen and Jagannathan (1997). They define the distance between a valid stochastic discount factor (SDF) m^* — that correctly prices all available assets — and an asset pricing model m as the second moment of the residuals from a regression of the tracking portfolio for a correct SDF on a tracking portfolio for the asset pricing model. Constructing a tracking portfolio is a regression problem. Hence, if we can construct the tracking portfolios for both m and m^*, we don't need to estimate N different regressions of returns on factors. Starting from the Hansen and Jagannathan (1997) analysis, we split the computation of their HJ-distance measure in two parts. In a first step we estimate the parameters of an SDF model using an instrumental variables regression with a constant term as dependent variable and the macro factors as endogenous regressors, for which we use their tracking portfolios as instruments. This part does not require a mean-variance efficient portfolio. The latter is needed, however, as a benchmark to compute the HJ-distance.

Our main contribution is to use statistical learning to construct such tracking portfolios. This is a big-N problem, but formulated as a regression problem. A large literature has developed for constructing sparse regression fits with many potential explanatory variables and a relatively small number of observations. We can therefore rely on standard techniques from the statistical learning literature to construct the tracking portfolios. In our empirical work we find tracking portfolios for standard macroeconomic variables such as consumption, inflation, interest rates and unemployment. Candidate assets for the tracking portfolios are the approximately 900 US equity portfolios that are available on Kenneth French's online data library. Since a portfolio return is a linear combination of asset returns, we use an additive learning algorithm for the construction of a sparse tracking portfolio. One of the most promising algorithms in this area is L_2-Boosting for regression models developed by Friedman (2001) and Bühlmann (2006).

Tracking portfolio regressions were developed by Lamont (2001). But whereas Lamont (2001) regresses macro variables on 13 different assets, we include about 900. Of course overfitting is a major concern in this case. To avoid overfitting Boosting and other statistical learning techniques often make use of an information criterium that penalizes model complexity. A secondary safeguard is randomized cross-validation, which estimates on a random subset of the data and evaluates on the remaining data.

The tracking portfolio for the true stochastic discount factor is an unrestricted mean-variance (MV) efficient portfolio. Technically, Jobson and Korkie (1983) show that the problem can be represented as a regression of a vector of ones on the collection of N excess returns. For this regression problem we employ the same L_2Boosting algorithm.

In the remainder of the paper we start by reviewing estimation of a stochastic discount factor model and the Hansen-Jagannathan distance to assess its cross-sectional fit. We then provide the details of the machine learning algorithm that we use to construct tracking portfolios. After explaining what we do, we discuss how our method relates to the literature, and espcially to some recent studies that apply machine learning tools to asset pricing problems. The second part of the paper contains the empirical application using a small number of macroeconomic factors and a large number of return portfolios. We conclude with some loose ends that require further research.

2 Stochastic discount factor projections

Let y be a vector of excess returns on N different assets or portfolios of assets. The stochastic discount factor methodology holds that the returns satisfy the N moment conditions

$$\mathrm{E}[my] = 0, \tag{1}$$

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

$$m = 1 - \delta' x \tag{2}$$

for a set of M macroeconomic factors x and parameters δ. Since we are modeling excess returns, the intercept of "1" is an arbitrary normalization parameter. Much of the interest in empirical asset pricing is in estimating the risk prices δ. It provides information on which macroeconomic factors are priced in the cross-section.

A separate question is how much of the cross-section is explained by the factors. Ideally the factors explain the entire cross-section of expected returns, in which case all the moment conditions (1) hold exactly. In practice, anomalies exist either due to mispricing or due to omitted factors. If the moment conditions are only approximate, the deviations $\mathrm{E}[my]$ in (1) are pricing errors. Hansen and Jagannathan (1997) propose a distance measure to evaluate the fit of the stochastic discount factor model. Their distance measure is defined as

$$HJ = \min_{m^* \in \mathcal{M}} \mathrm{E}\left[(m - m^*)^2\right], \tag{3}$$

where $\mathcal{M}$ is the set of all valid discount factors that satisfy the pricing condition $\mathrm{E}[m^* y] = 0$ for the set of N test assets. The criterion finds a valid discount factor m^* that is closest to the model m in (2). The solution is

$$HJ = \mathrm{E}[my]' \, \mathrm{E}[yy']^{-1} \, \mathrm{E}[my] \tag{4}$$

This is a quadratic form in the pricing errors $\mathrm{E}[my]$ with weighting matrix $\mathrm{E}[yy']^{-1}$. Hansen and Jagannathan (1997) discuss the difference between the distance measure HJ and a GMM weighting matrix. Most importantly the HJ weighting matrix, $\mathrm{E}[yy']^{-1}$, is independent of the model for m.

When the parameters δ are unknown the distance is further minimized with respect to δ. The resulting estimator, minimizing (4), is

$$\hat{\delta} = \left(\mathrm{E}[xy'] \, \mathrm{E}[yy']^{-1} \, \mathrm{E}[yx']\right)^{-1} \mathrm{E}[xy'] \, \mathrm{E}[yy']^{-1} \, \mathrm{E}[y] \tag{5}$$

In applications the population moments in (4)–(5) are replaced by sample moments assuming that we have a sample of T observations for both x and y. The big challenge is when N is large relative to T. In that case the weighting matrix contains $O(N^2)$ elements to be estimated, and finding any quantity depending on $\mathrm{E}[yy']^{-1}$ involves a huge matrix inversion that can be very sensitive to estimation error. In the really big N case, when $N > T$, the sample second moment matrix of excess returns will even be singular.

Our approach circumvents the need to explicitly estimate the large $(N \times N)$ weighting matrix $\mathrm{E}[yy']^{-1}$. For this we interpret the estimator (5) as the instrumental variables (IV) estimator of δ in the regression

$$1 = \delta' x + u \tag{6}$$

using y as instruments. To implement the IV estimator we perform a first stage regression to construct the tracking portfolios

$$\hat{x} = \mathrm{Proj}(x|y) = \mathrm{E}[xy'] \, \mathrm{E}[yy']^{-1} y \tag{7}$$

The second stage IV estimator then becomes

$$\hat{\delta} = \mathrm{E}[\hat{x}x']^{-1} \, \mathrm{E}[\hat{x}1], \tag{8}$$

which is identical to the original expression (5). For our purpose the main difference between (5) and (8) is the dimension of the matrix inversion. In (5) we need the large dimensional

$\mathrm{E}[yy']^{-1}$, whereas in (8) we only have the low dimensional $\mathrm{E}[\hat{x}x']^{-1}$. To avoid the explicit need for estimating $\mathrm{E}[yy']^{-1}$ we construct the tracking portfolios $\hat{x}$ applying tools from machine learning. The machine learning tools are designed to fit a regression with a large number of potential explanatory variables. Details are discussed in the next section.

The analogy with the IV formulation puts our model directly within the framework of Belloni, Chen, Chernozhukov and Hansen (2012). We have the regression model (6) with endogenous regressors x, for which we have many instruments y. All y_j are valid instruments, but including too many is not efficient and will lead to biases. Given instruments $\hat{x}$, the IV estimator for δ has covariance matrix

$$\mathrm{Var}(\sqrt{T}\hat{\delta}) = \omega^2 \, \mathrm{E}[\hat{x}x']^{-1} \, \mathrm{E}[\hat{x}\hat{x}'] \, \mathrm{E}[x\hat{x}']^{-1} \tag{9}$$

with $\omega^2 = \mathrm{E}[u^2]$ the variance of the stochastic discount factor. This reduces to

$$\mathrm{Var}(\sqrt{T}\hat{\delta}) = \omega^2 \, \mathrm{E}[\hat{x}\hat{x}']^{-1} \tag{10}$$

when instruments are chosen optimally. Under sparsity conditions for the projections, Belloni *et al* (2012) prove that various machine learning methods lead to optimal instruments. One of the machine learning methods is the L_2-Boosting algorithm we will discuss in the next section.

Tracking portfolios for the macro variables are all we need for estimating the risk prices δ. To evaluate the fit of the discount factor model using the HJ distance we need one more projection. Let $\hat{m} = \mathrm{E}[my]' \, \mathrm{E}[yy']^{-1} y$ be the projection of m onto the excess returns y. The HJ distance can be rewritten as

$$\begin{aligned} HJ &= \mathrm{E}\left[\left(\mathrm{E}[my]' \, \mathrm{E}[yy']^{-1}\right) yy' \left(\mathrm{E}[yy']^{-1} \, \mathrm{E}[my]\right)\right] \\ &= \mathrm{E}\left[\hat{m}^2\right], \end{aligned} \tag{11}$$

This expresses the distance as the squared projection of the discount factor on the excess returns. From the definition of the linear discount factor model, $m = 1 - \delta' x$, we can split the projection in two parts,

$$\hat{m} = \hat{1} - \delta' \hat{x} \tag{12}$$

The new element in (12) and (13) is the projection $\hat{1}$, which is defined as the fitted values from regressing the unit vector on the space of excess returns. The constant tracking portfolio $\hat{1}$ is nothing but the excess return of a mean-variance efficient portfolio. Once we have the projections, the HJ distance can now be estimated as the second moment $\mathrm{E}\left[\hat{m}^2\right]$.

Another way to look at the HJ distance through the lens of tracking portfolios is as the expected squared residual $\mathrm{E}[u^2]$ from the time series regression

$$\hat{1} = \delta' \hat{x} + u \tag{13}$$

The left-hand of this regression is the (excess) return on a mean-variance efficient portfolio, while the right-hand side variables represent a collection of portfolio returns. With return data, evaluation of the stochastic discount factor model only involves a time series regression of $\hat{1}$ on a set of macroeconomic tracking portfolios $\hat{x}$. The regression assesses how well a mean-variance efficient portfolio can be explained by the excess returns of macroeconomic tracking portfolios. If the tracking portfolios span the efficient frontier, the residual will be zero. The HJ distance is measured by the Mean-Squared-Error (MSE) of the regression. Adding an intercept to (13) we can further decompose the MSE into a bias and variance component as in standard spanning tests (Huberman and Kandel, 1987).

With the complete set of N instruments, the IV estimator (8) using the individual tracking portfolios as instruments, is numerically identical to the least squares estimator for δ in (13). In terms of data matrices, if $\boldsymbol{Y}$ be the $(T \times N)$ matrix of sample excess returns, and $\boldsymbol{X}$ the $(T \times M)$ matrix of macroeconomic news data, then the IV estimator

$$\hat{\delta} = (\hat{\boldsymbol{X}}' \boldsymbol{X})^{-1} \hat{\boldsymbol{X}}' \iota \tag{14}$$

equals the least squares estimator

$$\tilde{\delta} = (\hat{\boldsymbol{X}}' \hat{\boldsymbol{X}})^{-1} \hat{\boldsymbol{X}}' \hat{\iota} \tag{15}$$

when we use the unrestricted $(T \times T)$ projection matrix $\boldsymbol{H} = \boldsymbol{Y}(\boldsymbol{Y}'\boldsymbol{Y})^{-1}\boldsymbol{Y}'$ to generate the projections $\hat{\boldsymbol{X}} = \boldsymbol{H}\boldsymbol{X}$ and $\hat{\iota} = \boldsymbol{H}\iota$. The identity holds because the projection is idempotent, *i.e.* $\boldsymbol{H}^2 = \boldsymbol{H}$. When different macro variables (and the mean-variance portfolio) use different sets of instruments with different levels of shrinkage, we have $\hat{\boldsymbol{x}}_j = \boldsymbol{H}_j \boldsymbol{x}_j$ with individual projection matrices $\boldsymbol{H}_j$ ($\boldsymbol{x}_j$ is the j^{th} column of $\boldsymbol{X}$). Since in general $\boldsymbol{H}_i \boldsymbol{H}_j \neq \boldsymbol{H}_j \neq \boldsymbol{H}_i$, the two estimators are no longer identical. For estimating the prices of risk we will use the IV estimator (14). For the more informal assessment whether the tracking portfolios span the efficient frontier we use the least squares estimator (15).

3 Boosting Tracking Portfolios

Our approach for dealing with a large number of asset returns relies on the construction of tracking portfolios. Tracking portfolios have been explored in detail by Lamont (2001). His specification starts from the time series regression

$$X_t = \beta' y_t + \phi' Q_{t-1} + \epsilon_t, \tag{16}$$

where X_t is the observation at time t for some macroeconomic state variable, y_t is the N-vector of excess returns at time t and Q_{t-1} a q-vector of predetermined variables such as a constant term and lagged values of X. The tracking portfolio is defined as the linear combination $\hat{x}_t = \beta' y_t$. As in Lamont (2001) we work with excess returns to avoid imposing the portfolio restriction that elements of β sum to one. As emphasized by Lamont (2001) a large number of variables in y_t may lead to overfitting and spurious results.

When N is large relative to T, sparsity conditions are needed to make this a feasible regression. This can be done in many ways, for example by Bayesian shrinkage priors or by imposing zero restrictions as in Lasso. The boosting algorithm that we employ works under the assumption that $\sum_{j=1}^{N} |\beta_j|$ is bounded as N increases. Adding more returns should lead to a more balanced tracking portfolio with smaller individual coefficients $|\beta_j|$. In contrast to the standard tracking portfolio where the assets are specified a priori, we are constructing a portfolio sequentially by adding assets that complement each other in terms of predictive properties.

Boosting algorithms were developed as a technique for producing a projection by aggregating weak predictors. The statistical foundation of the algorithm is derived from the functional gradient descent representation established by Breiman (1998). The version we use here closely follows Bühlmann (2006). It is applicable to continuous response variables and known as L_2-Boosting with component-wise linear least squares.

Our empirical implementation deviates slightly from the specification in (16). Instead of the multiple regression with both y_t and Q_{t-1} we first regress X_t on the fixed regressors Q_{t-1} before starting the boosting algorithm. The residuals $x_t = X_t - \text{Proj}(X_t|Q_{t-1})$ have mean zero, and therefore preserve the scaling of the discount factor specification. We define these residuals as the macro news factors in our model.

The model selection procedure is summarised in algorithm 1. It is based on a model

$$\boldsymbol{x} = \sum_{j=1}^{N} \beta_j \boldsymbol{y}_j + \boldsymbol{e} \tag{17}$$

where $\boldsymbol{x}$ is the T-vector of dependent variables and $\boldsymbol{y}_j$ are T-vectors of observations on the explanatory variables.

Algorithm 1 L_2 Boosting with component-wise least squares

Initialize:
- Define a stepsize parameter $\nu \in (0, 1]$
- Set initial projection $\hat{\boldsymbol{x}} = 0$ and coefficients $\beta_j = 0$ $(j = 1, \ldots, N)$
- Set iteration counter $\ell = 0$

for $\ell = 1$ to L **do**
- Compute residuals $\boldsymbol{e} = \boldsymbol{x} - \hat{\boldsymbol{x}}$
- Find univariate regression coefficients $b_j = (\boldsymbol{y}_j' \boldsymbol{y}_j)^{-1} \boldsymbol{y}_j' \boldsymbol{e}$ and residuals $\boldsymbol{u}_j = \boldsymbol{e} - \boldsymbol{y}_j b_j$
- Find $i = \text{argmin}_j (\boldsymbol{u}_j' \boldsymbol{u}_j)$
- Update the projection: $\hat{\boldsymbol{x}} \leftarrow \hat{\boldsymbol{x}} + \nu \boldsymbol{y}_k b_i$
- Update regression coefficients: $\beta_i \leftarrow \beta_i + \nu b_i$

end for

At each iteration the algorithm searches for the univariate predictor that improves the fit the most. For this the algorithm uses univariate regressions of the current residuals on the explanatory variables. It chooses the variable that best fits the current residuals. The fitted values are updated by the newly selected regressors with a shrinkage parameter ν.

Boosting has two important tuning parameters: the step size ν and the number of iterations L. The literature concludes that the exact value of the tuning parameter does not have much of an effect on the results, as long as it is sufficiently small. The value $\nu = 0.1$ is often recommended. The stepsize should be large enough for the algorithm to make progress, and small enough to enable shrinkage and deal with multicollinearity. The number of iterations is more critical, since eventually the solution will converge to the least squares solution and therefore be prone to severe overfitting.

To avoid overfitting the actual number of iterations should be set based on an information criterion or through cross-validation. In our empirical work we consider both options. For the information based stopping rule we use the Bayesian Information Criterion (BIC)

defined as

$$BIC = \ln(\boldsymbol{e}'\boldsymbol{e}) + k\frac{\ln T}{T}, \tag{18}$$

where k is the model complexity. For the definition of the model complexity parameter k we follow Bühlmann (2006). The fitted values after L iterations can be expressed as

$$\hat{\boldsymbol{x}} = \boldsymbol{H}^L \boldsymbol{x} \tag{19}$$

with $\boldsymbol{H}^L$ the $(T \times T)$ pseudo projection (*hat*) matrix

$$\begin{aligned} \boldsymbol{H}^L &= \boldsymbol{H}^{L-1} + \nu \boldsymbol{P}_{i_L}(\boldsymbol{I} - \boldsymbol{H}^{L-1}) \\ &= \boldsymbol{I} - \prod_{\ell=1}^{L}(\boldsymbol{I} - \nu \boldsymbol{P}_{i_\ell}) \end{aligned} \tag{20}$$

The *hat* matrix aggregates the iterations using the univariate projection matrices $\boldsymbol{P}_i = \boldsymbol{y}_i(\boldsymbol{y}_i'\boldsymbol{y}_i)^{-1}\boldsymbol{y}_i'$ starting from $\boldsymbol{H}_0 = 0$ and using the selected regressors i_ℓ in iteration ℓ. The matrix $\boldsymbol{H}_\ell$ is a pseudo projection matrix since it is not idempotent. The complexity parameter after L iterations is then defined as

$$k = \mathrm{tr}(\boldsymbol{H}^L) \tag{21}$$

In standard regression models this definition of k would be equal to the number of explanatory variables. Due to the shrinkage parameter ν in the boosting algorithm k is less than the number of included variables, and also much smaller than the number of iterations L. The stopping criterion uses the value of L^* that minimizes the BIC criterion.

The alternative stopping rule is based on cross validation. With J-fold cross-validation we split the sample in J equal-sized subsamples. For each fold j we run the boosting algorithm for L steps on the complement of all data not in the j^{th} subsample. We then construct predicted values for the j^{th} subsample using the estimated parameters $\beta^{L,j}$. Doing this for all j gives a sample of T test residuals. The value of L that minimizes the residual sum of squares of the test residuals determines the optimal number of iterations. With this L^* we then run the boosting algorithm one more time on the full sample to construct the tracking portfolio.

In practice we use 5-fold cross validation with random subsamples. We randomly draw (without replacement) the time series observation numbers to be part of subsample j. To minimize the sampling variation induced by the random subsample assignment we repeat this 10 times and compute the cumulative sum of squares over all 10 cross-validation samples.

The tracking portfolio may be written as

$$\hat{x}_t = (\beta^{L^*})' y_t \tag{22}$$

We construct a separate tracking portfolio for each of the M macroeconomic variables in the SDF model. The algorithm as presented above does not impose that some assets must always be part of the tracking portfolio. It is however straightforward to force variables which a priori known to be relevant into the projection. The only necessary adjustment is in the initial condition.

The boosting algorithm is also used to construct the mean-variance optimized portfolio $\hat{\iota}$. The only difference with the macroeconomic tracking portfolios is the initial condition. The ι-vector of ones is not prefiltered in any way before being used as the dependent variable.

Bühlmann (2006) shows that the L_2Boosting algorithm is consistent for the conditional expectation of the response if the regression coefficients in (17) satisfy a sparsity condition. More precisely, when the number regressors N is allowed to grow as $\ln N/T \to 0$ as T grows, then the sparsity condition $\sum_{j=1}^{N} |\beta_j| = o(\sqrt{T/\log(N)})$ as $T \to \infty$ leads to the consistent projection

$$\parallel \boldsymbol{Y} \left(\hat{\beta} - \beta\right) \parallel /T = o_p(1) \tag{23}$$

(Bühlmann and Van de Geer, 2011) subject to further regularity conditions on the existence of moments for x and y. The estimator is consistent for the parameters β_T if the number of iterations $L_T \to \infty$ such that $L_T = o(\sqrt{T/\ln N}$ as T tends to infinity. The theoretical foundation of the algorithm is attractive for our purposes because of the sparsity conditions. 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.

Boosting for instrument selection was suggested in Bai and Ng (2009). Belloni, Chen, Chernozhukov and Hansen (2012) propose it as an alternative to LASSO for instrument selection in an IV regression. The two procedures have a number of differences and similarities. The similarity is in the simultaneous model selection and shrinkage performed by both methods. The difference is in the form of the model complexity penalty. LASSO penalizes according to the L_1 norm, while boosting is an L_2 procedure. The difference shows up in the type of sparsity that they support and produce. LASSO assumes that just a few instruments have non-zero coefficients, whereas boosting allows for many small coefficients with a bound on their sum of absolute values. The latter seems more suitable for a portfolio interpretation. With many assets, diversification would usually imply that all assets have small weights. The distinction is not completely clear, however. One could consider the case that some of the assets being considered are themselves well diversified portfolios like the market portfolio or the Fama-French factors. Including such well-diversified assets blurs the distinction between many small' and few large' coefficients.

Although being an L_2 learning algorithm, boosting differs from ridge regression. In ridge regression all coefficients are shrunk. With boosting, the final result will still have many exact zeros. Hastie, Tibshirani and Friedman (2009) evaluate the boosting algorithm as one of best off-the-shelf' choices for data mining.

Summarizing, booting provides a sensible methodological foundation for the purpose of tracking portfolio and pricing kernel estimation. The linear construction accommodates the linearity in the assets implied by the pricing conditions, the least squares fitting procedure minimizes the second norm which under (Hansen and Jagannathan, 1997) provides a sensible measure of the misspecification in a pricing model, and the component-wise feature accommodates the large cross-sectional dimension.

4 Literature

Most distinguishing in our approach is the focus on macroeconomic risk factors as opposed to characteristic sorted portfolios. Like many other papers our interest is in estimating risk prices for factors in a stochastic discount factor. Since the macroeconomic factors are not traded, the main challenge is the selection of optimal instruments in the form of tracking portfolios. Tracking portfolios are common in asset pricing tests. Our innovation relates to the use of machine learning tools in constructing tracking portfolios from a large set of candidate asset portfolios.

Estimating the risk prices in a SDF model has a long history in empirical asset pricing.

Our specification is directly related to Hansen and Jagannathan (1997), who also note the relation between their distance measure and the use of factor-mimicking portfolios. The most common way to estimate risk prices from the cross section follows the Fama-MacBeth two-stage methodology. The principal difference with our instrumental variables estimator is the weighting matrix. Instead of the full $\mathrm{E}[yy']^{-1}$ it uses a N-dimensional diagonal matrix. Our tracking portfolio method is designed to put more emphasis on those assets that have a stronger correlation with the macro factors.

One of our tracking portfolios is the projection of excess returns on the unit vector, which identifies the weights of a mean-variance efficient portfolio. This part is related to Kozak, Nagel and Santosh and (2017). We estimate the weights from the time series regression

$$1 = w'y_t + u_t \tag{24}$$

Kozak, Nagel and Santosh and (2017) multiply both sides of (24) by y_t' and take the time series average. The left hand side then becomes the N-vector of mean excess returns, while the right hand side gives the $(N \times N)$ matrix of second moments. This is equivalent to a cross-sectional regression of means on covariances with N observations and N explanatory variables. Kozak, Nagel and Santosh and (2017) apply machine learning techniques to this derived regression model assuming that the return covariance matrix is free of measurement error and can be used as fixed regressors.

Our big N approach differs from a number of other recent studies that employ machine learning techniques to estimate the risk prices in a stochastic discount factor model. Feng, Giglio and Xiu (2017) specify a linear SDF model for m, and consider the problem that M, the dimension of x, can be very large. Doing so they address the problem of finding which factors are priced within a large set of potential factors. We assume that factors are given, and instead consider the evaluation of a pricing model using the selected factors. In our case the dimension M of potential macroeconomic factors is small, but the number of assets N is large. That puts our emphasis on determining which portfolios are most informative for estimating risk prices for a given set of factors.

We take our test assets y as a given collection of characteristic sorted portfolios, which we download from the dataverse of Kenneth French. These are not individual stocks, but managed portfolios. Other studies, such as Kozak, Nagel and Santosh (2017) construct portfolios from individual stocks. Kelly, Pruit and Su (2018) simultaneously construct factors and portfolios from individual returns. Freyberger, Neuhierl and Weber (2017) use machine learning techniques to find functions of characteristics that explain cross-sectional differences in expected returns.

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. Classic referenes are Jobson and Korkie (1982) and more recently DeMiguel, Garlappi and Uppal (2009). Many shrinkage and dimension reduction techniques have been suggested to obtain portfolios with reasonable out-of-sample performance. Our L_2-Boosting technique is, as far as we know, new in this respect. A direct horse-race between L_2-Boosting and state-of-the-art portfolio optimization techniques is outside the scope of the present paper.

Kleibergen and Zhan (2018) express doubts about the correlation between data and financial returns. Following earlier work by Kan and Zhang (1999) they suspect that macroeconomic factors may be useless, meaning that they lack any correlation with returns. Obviously, if the tracking portfolios are not sufficiently correlated with the macro news variables, the second stage instrumental variable regression can only create noise. An important diagnostic of the tracking portfolios is their correlation with the factor they are supposed to mimic.

5 Data

The financial data used for the construction of the tracking portfolios are all collected from the sorted equity portfolios maintained on the website of Kenneth French. We include the domestic US equity portfolios. Table 1 provides the list of the data that we consider for the tracking portfolios. In total the candidate set of instruments contains $N = 901$ excess returns, all observed on a monthly frequency from January 1965 to February 2016, providing $T = 638$ time series observations. Hence the final set of asset candidates is maintaining a cross sectional dimension well in excess of the time dimension of the data.

Except for the seven factors (five Fama-French plus Momentum and Long-Term Rever-

Table 1: Portfolios included from the Kenneth French data library

N	Portfolios formed on	File name
18	Size	Portfolios_Formed_on_ME
19	Book-to-Market	Portfolios_Formed_on_BE-ME
18	Investment	Portfolios_Formed_on_INV
18	Operating Profitability	Portfolios_Formed_on_OP
10	Momentum	10_Portfolios_Prior_12_2
10	Short-Term Reversal	10_Portfolios_Prior_1_0
10	Long-Term Reversal	10_Portfolios_Prior_60_13
19	Earnings/Price	Portfolios_Formed_on_E-P
19	Cashflow/Price	Portfolios_Formed_on_CF-P
19	Dividend Yield	Portfolios_Formed_on_D-P
100	Size and BtM	100_Portfolios_10x10
99	Size and Op. Prof.	100_Portfolios_ME_OP_10x10
100	Size and Inv.	100_Portfolios_ME_INV_10x10
25	BtM and Op. Prof.	25_Portfolios_BEME_OP_5x5
25	BtM and Inv.	25_Portfolios_BEME_INV_5x5
25	Op. Prof. and Inv.	25_Portfolios_OP_INV_5x5
25	Size and Momentum	25_Portfolios_ME_Prior_12_2
25	Size and STR	25_Portfolios_ME_Prior_1_0
25	Size and LTR	25_Portfolios_ME_Prior_60_13
25	Size and Accruals	25_Portfolios_ME_AC_5x5
25	Size and Market Beta	25_Portfolios_ME_BETA_5x5
25	Size and Variance	25_Portfolios_ME_VAR_5x5
25	Size and Residual Variance	25_Portfolios_ME_RESVAR_5x5
35	Size and Net Share Issues	25_Portfolios_ME_NI_5x5
32	Size, BtM, and Op. Prof.	32_Portfolios_ME_BEME_OP_2x4x4
32	Size, BtM, and Inv.	32_Portfolios_ME_BEME_INV_2x4x4
32	Size, Op. Prof., and Inv.	32_Portfolios_ME_OP_INV_2x4x4
48	Industry	49_Industry_Portfolios
5	Fama-French Factors	F-F_Research_Data_5_Factors_2x3
1	Long-term Reversal	F-F_LT_Reversal_Factor
1	Momentum	F-F_Momentum_Factor

All data is from `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. The table lists the file names as they are recorded on the website dataverse. From all files we use the value weighted portfolios. Abbreviations: Op. Prof. = Operating Profitability, Inv. = Investments, BtM = Book-to-Market, STR = Short-Term-Reversals, LTR = Long-Term-Reversals. Some of the sets contain fewer variables than in the original data library due to missing values. A time series of returns is excluded from our database if more than one year of returns is missing in our sample Jan 1965- Feb 2018. With fewer missing values we replace the missing data by the riskfree rate.

Table 2: Portfolio summary statistics

	Mean	St.dev	Min	Median	Max	market
Average	0.69	0.22	-0.20	0.69	1.43	0.52
Std dev	4.89	1.09	2.02	5.47	11.32	4.44
Sharpe	0.14	0.04	0.00	0.13	0.25	0.12

Average refers to the time series average of excess returns. The columns then provide the overall average (*Mean*), the cross-sectional standard deviation of the averages (*St.dev*), the minimum, median and maximum of the averages, and finally the average of the market excess return. The rows for *Std dev* provide the same cross-sectional information for the time series standard deviations and the individual Sharpe ratios, respectively. All returns are expressed in units of percent per month.

sal), which are already excess returns, all series are expressed in excess of the risk-free rate. All returns are discrete time monthly returns. Summary statistics of the excess return data are in Table 2. A number of facts stand out. First, both the median and the mean average excess return of all included portfolios is above the average excess return of the market portfolio. The cross-sectional standard deviation (and min and max of the averages) indicate that there is a large dispersion in average returns. Second, most portfolios have a larger standard deviation than the marker portfolio, although some portfolios appear to have very low risk. Third, and important for our further results, there exist many portfolios that on their own already have a Sharpe ratio that is above the Sharpe ratio of the market portfolio. The maximum Sharpe ratio is for the portfolio of small firms with low prior volatility (`ME_VAR_5x5_SMALL.LoVAR`).

The sorted returns contain several redundancies. For example, quintile portfolios are approximately the average of two decile portfolios. Similarly, double sorted portfolios on size and another characteristic can be aggregated to single sorts on the characteristics. For the boosting algorithm these redundancies are not a problem when constructing macroeconomic tracking portfolios. However, they do lead to some near arbitrage relations when we construct the unit tracking portfolio. If the dependencies would be exact, this would not cause any problems for the boosting algorithm. But with near exact linear dependence, the model selection for the constant mimicking portfolio will identify faux arbitrage portfolios by leveraging on the very small differences in the linear dependencies. For the Mean-Variance efficient portfolio we therefore only include the highest order quantiles in the candidate set, *i.e.* we take deciles instead of quintiles. We further drop the univariate

Table 3: Macro data sources

Personal Consumption Expenditures	US Bureau of Economic Analysis, Personal Consumption Expenditures: Chain-type Price Index [PCEPI], `https://fred.stlouisfed.org/series/PCEPI`
Consumer Price Index	US Bureau of Labor Statistics, Consumer Price Index for All Urban Consumers: All Items [CPIAUCSL], `https://fred.stlouisfed.org/series/CPIAUCSL`
Baa Corporate Bond Yield	Board of Governors of the Federal Reserve System, Moody's Seasoned Baa Corporate Bond Yield [BAA], `https://fred.stlouisfed.org/series/BAA`
10-Year Treasury-Rate	Board of Governors of the Federal Reserve System, 10-year treasury constant maturity rate [DGS10], `https://fred.stlouisfed.org/series/DGS10`
Unemployment	Federal Reserve Economic Data, Civilian Unemployment Rate, `https://fred.stlouisfed.org/series/UNRATE`

sorts when a bivariate sort with size is also available. With the bivariate and trivariate sorts the linear combinations that are necessary to form near exact dependencies are more complex. Here we maintain both of these sorts along with the trivariate portfolios.

Table 3 contains our list of macroeconomic variables. The first macro variable is consumption. We measure consumption growth as the monthly change of log personal consumption expenditures. Consumption news is consumption growth in deviation of its long-term mean. Consumption growth is the most basic variable in a stochastic discount factor model. When it is the only factor in the SDF its coefficient measures coefficient of relative risk aversion. The linear approximation may be poor, so an obvious extension is to have separate coefficients for positive and negative growth. Introducing nonlinearities by piecewise linear terms is a common powerful procedure in neural network models. We define the variable *Negative Growth* as the minimum of consumption growth and zero, which we again demean.

Other standard macroeconomic variables, at least since Chen et al. (1986), are inflation, credit spread, and term spread. We also include unemployment as a business cycle indicator. For inflation we use the consumer price index of all urban consumers. Unexpected inflation is constructed as the residual from an AR(1) model for the monthly change in the log of the index. Unemployment shocks are defined as the first difference of the unemployment rate. The change in the 10-year interest rate is used as a proxy for term structure news. For

Table 4: Summary statistics macro variables

Series	St. Dev.	Skewness	Kurtosis
Consumption (Δc)	0.502	0.13	6.15
Negative Growth ($\min(0, \Delta c)$)	0.553	-2.04	8.88
Inflation ($(1 - \phi \text{Ł})\Delta p$)	0.251	0.06	8.16
Unemployment (ΔU)	0.356	1.43	7.31
Negative Unemployment ($\min(0, \Delta U)$)	0.170	-1.20	5.21
Long-Term Interest rate (ΔR)	0.338	-0.28	6.15
Credit Spread ($(1 - \phi \text{Ł})C$)	0.223	0.61	9.35

Statistics refer to monthly data for January 1965 to February 2018. The news filter for the data is indicated in parentheses: Δ means first differencing, ϕŁ indicated an AR(1) adjustment. All series are demeaned. Filters are run on the complete data series, extending to the earliest available date in the source. Units are percent per month.

the credit spread we take the yield spread between the BAA rated yield and the 10/year government bond yield. Credit rate news are the residuals from an AR(1) filter. Summary statistics along with details of the construction of the news variables are reported in table 4.

Timing of information is not obvious for the macro series. With nowcasting and analyst expectations much of the news of current month data is already known before the end of the month. On the other hand, reported data and revisions imply that data for month t are often only known at a later date. To minimize the impact of timing uncertainty on the construction of the tracking portfolios, we use the sum of the news at times $t-1$, t and $t+1$ as the dependent variable x_t for the all macro variables except the Credit Spread and the Long-Term Interest Rate.

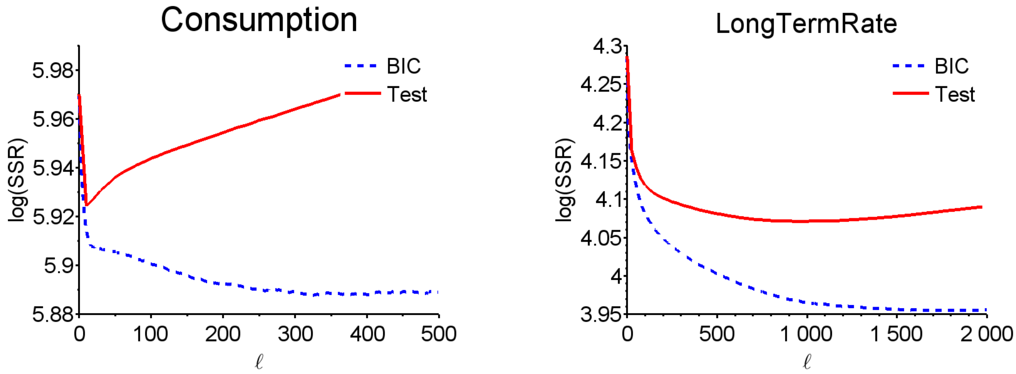
6 Results

6.1 Tracking portfolios

Crucial for estimating the risk prices are the first stage projections of the macroeconomic news variables on the space of excess returns. Table 5 provides summary statistics. For all series the BIC criterium selects a much larger model than cross-validation. When the number of boosting steps increases, the variance of the tracking portfolios also increases. Most of this is by construction, since the amount of shrinkage decreases with each additional iteration.

As we have just a small set of macro variables, we will discuss variables individually.

Figure 1: Tracking portfolio fit



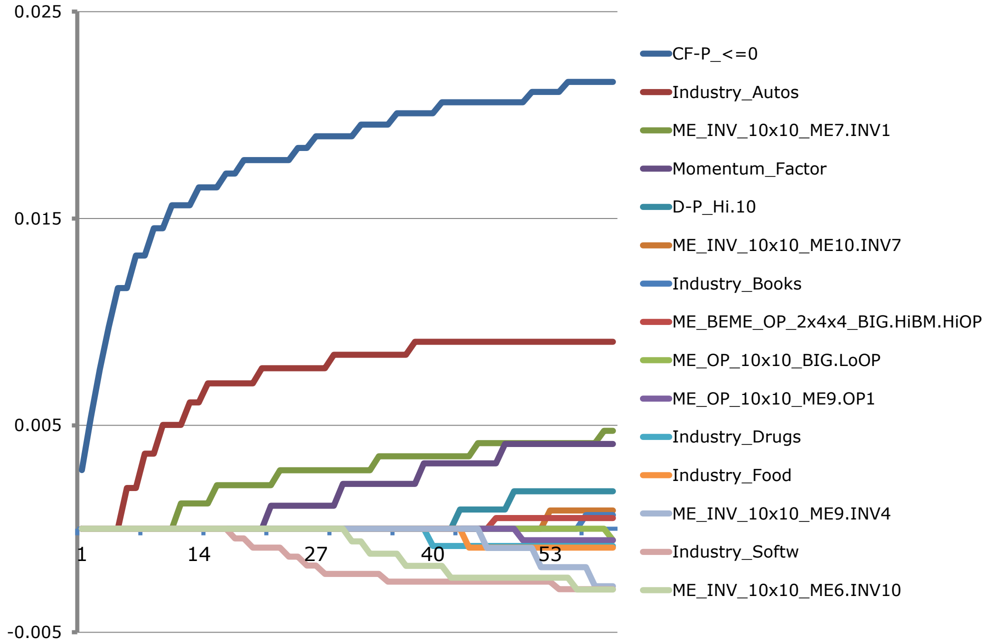
The figure shows the fit of the tracking portfolio for macroeconomic news as a function of the number of boosting steps (ℓ). The solid line *Test* refers to the sum of squared residuals (SSR) of the 5-fold cross-validation as is plotted as ln(SSR). The dashed line is the BIC criterion $\ln(\text{SSR}) + k \ln T/T$, where k is the model complexity and T the time series sample size. Without any regressor ($\ell = 0$) the two criteria are both equal to the log sum of squares $\sum_t x_t^2$ of the news series.

Starting with *Consumption*, the projection more or less delivers what we would expect. Consumption is related to returns, but the correlation is not that large, even when we allow searching among a large set of portfolios. Using cross validation the optimal tracking portfolio explains about 8% of the variation in the consumption shocks. Using the BIC criterium the algorithm continues much longer, and generates a tracking portfolio with an R^2 of 19%. As with most tracking portfolios the correlation between the two alternatives, termination by CV versus the BIC stopping rule, is large (see last column of table 5). Figure 1 shows how the fit of the model evolves with the number of boosting iterations. The fit of the model remains fairly stable as the iterations progress. Despite the large number of additional steps needed to minimize the BIC criterium, the algorithm does not make much progress in fitting the consumption growth. The correlation seems sufficiently large to provide a valid instrument; yet it remains low enough to suspect gross overfitting.

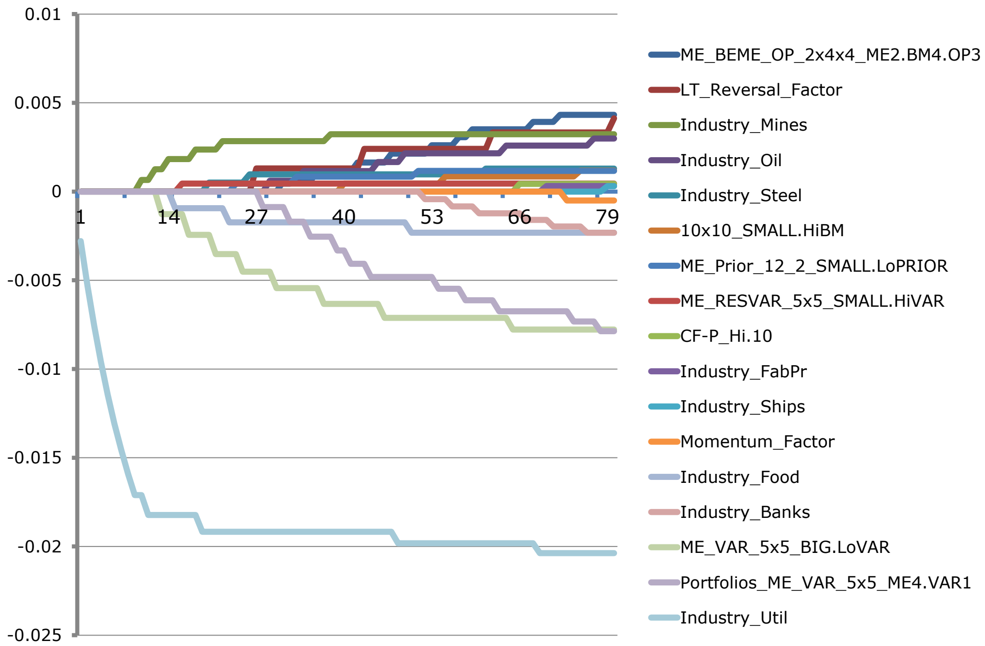
Figure 2 shows the evolution of the tracking portfolio as a function of the number of boosting iterations. The most highly correlated portfolio is the sort on firms with negative cashflow-to-price (`CF-P_<=0`). This is also the portfolio that in the end receives the largest weight in the consumption tracking portfolio. At later steps the algorithms starts to increase the coefficient on Automobile Industry (`49_Industry_Autos`) to make it the second most important component of the tracking portfolio based on the BIC stopping criterion.

Most remarkable is the low Sharpe ratio of the tracking portfolio. Whether we use CV or

Figure 2: Tracking portfolio composition



(a) Consumption



(b) Long-term interest rate

The figure shows the evolution of the tracking portfolio weights as a function of the number of iterations of the L_2-Boosting algorithm. At each step the algorithm adjusts the coefficient of the return portfolio that has the highest correlation with the current residuals.

Table 5: Tracking portfolio summary statistics

	CV				BIC				
	R^2	trH	SDev	Sh	R^2	trH	SDev	Sh	ρ
Consumption	0.076	1.0	0.125	0.068	0.186	9.0	0.213	0.070	0.88
Neg. growth	0.042	1.1	0.062	0.122	0.277	18.2	0.152	0.079	0.61
Inflation	0.218	8.2	0.132	0.044	0.430	29.4	0.199	0.129	0.89
Long-term rate	0.435	20.1	0.167	0.034	0.489	28.8	0.183	0.031	0.99
Credit spread	0.286	9.5	0.082	0.124	0.389	20.0	0.098	0.103	0.95
Unemployment	0	0	0	0	0.306	19.7	0.097	0.010	0
Neg. unemp.	0.069	3.3	0.018	0.137	0.178	11.7	0.036	0.231	0.80

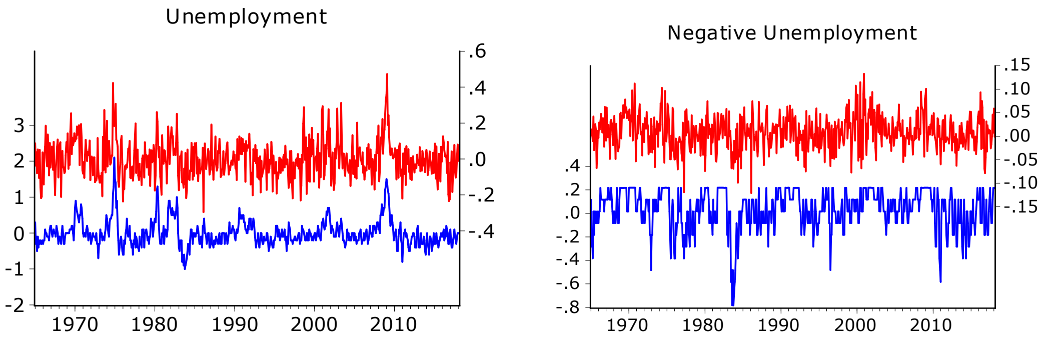
The table shows summary statistics for the tracking portfolios constructing using the L_2Boosting algorithm. R^2 is the squared correlation between the macro news and the tracking portfolio; trH is Bühlmann's degrees of freedom parameter k defined in (21); SDev is the standard deviation of the tracking portfolio excess returns; Sh is the Sharpe ratio of the tracking portfolio excess returns. For the left side of the table Boosting iterations are terminated by minimizing the test error sum of squares using 5-fold cross-validation (CV). For the right-side of the table the stopping criterium is based on the BIC. In the final column ρ is the correlation between the returns of the BIC and CV tracking portfolios. Abbreviations: Neg. growth = $\min(0, \Delta c)$, Neg. Unemp = $\min(0, \Delta U)$.

BIC as the stopping criterion, the tracking portfolio has a Sharpe ratio that is much below the Sharpe ratio of the market portfolio, and is in the lower part of the distribution over all the portfolio sorts in the database. Since the consumption tracking portfolio apparently combines assets in a way that leads to a low Sharpe ratio, it will never be able to explain the high Sharpe ratio of a mean-variance efficient portfolio. Partly this is known as the equity premium puzzle, but here it comes as more of a surprise, since the algorithm could choose from many candidate portfolios, most of which have Sharpe ratios that are above that of the market. Diversifying over these portfolios would easily enable much larger Sharpe ratios.

For *Negative Consumption Growth* the tracking portfolio results are qualitatively similar. Cross-validation identifies a very sparse model, this time consisting mostly of the Automobile industry. The BIC criterion again selects a more complex model. The most important additional variable selected by BIC is one of the low volatility portfolios (`ME_VAR_5x5_ME3.VAR1`). Since this portfolio is added at a later stage during the model selection, the correlation between the CV and BIC stopped tracking portfolio returns is lower than for most other variables. When Negative Growth is included in our later SDF pricing regressions we may thus expect some sensitivity with respect to which stopping criterium we use.

Among the macro variables *Inflation* news appears to be the easiest to fit. The CV stopping criterium already identified a large model with an R^2 of 22%. The most important weight is the Oil industry (`49_Industry_Oil`) followed by Coal (`49_Industry_Coal`).

Figure 3: Tracking portfolio returns



The figure shows time series of news (low, blue) and tracking portfolio excess returns (up, red) for the *Unemployment* and *Negative Unemployment* data. The tacking portfolio uses the regression fit identified by the BIC criterium.

The most puzzling results occur for *Unemployment*. Here cross-validation does not find any return variable that correlates with the macro news series. What this means is that if we randomly select 80% of the months in the sample and run the boosting algorithm on this sample to construct tracking portfolio weights, the out-of-sample fit on the remaining 20% of the data is worse than having no model at all. The BIC criterium, however, using the full sample both for fitting and evaluation, does find substantial explanatory power. The discrepancy between the two is most likely related to outliers that happen to coincide in both the news series and in some return portfolio. With full sample estimation these outliers are influential, whereas they will often be absent in randomly selected subsamples used by cross validation. Figure 3 shows that some of this can indeed explain the big difference between BIC and CV in this case. In the full sample the two biggest increases in unemployment (in 1975 and 2009) are fitted reasonably well by the BIC tracking portfolio.

Negative unemployment does not have the positive outliers and is easier to track with cross validation. By its definition a shock to negative unemployment is a decrease in unemployment, and thus positive economic news. If strong 'negative' news would be priced, the complement $U^+ = U - U^-$ will be priced in the SDF, leading to opposite coefficients for U and U^-. Anticipating the factor pricing regressions, it is important to note that the tracking portfolio for *Negative unemployment* attains the highest Sharpe ratio among all tracking portfolios, both according to CV as well as BIC.

The two financial variables, *Credit Spread* and *Long-term interest rate* are less subject to measurement error and timing of news. Cross validation already fits both with high R^2's.

Even though the BIC models are a bit less sparse, the tracking portfolios are very similar under BIC and CV. For both variables the BIC and CV returns are almost perfectly correlated. One of the more noticeable properties of the interest rate tracking portfolio is the large (negative) weight for the utilities industry (`49_Industry_Util`). The evolution of the portfolio weights is shown in figure 2. The tracking portfolio for the credit spread is dominated by various low volatility portfolios (`ME_VAR_5x5_ME3.VAR1`, `ME_VAR_5x5_ME2.VAR1`, `ME_VAR_5x5_ME4.VAR1`)

TO DO: TRACKING PORTFOLIOS FOR A BROADER RANGE OF MACRO AND FINANCIAL VARIABLES

6.2 Pricing kernel estimates

The tracking portfolios serve as instruments for the macro variables when estimating the SDF equation (6) using the IV estimator (8) and standard errors as in (10). Results are in table 6, for both the CV and BIC tracking portfolios.

Consistent with most of the empirical literature the basic Consumption CAPM with only consumption growth as a factor is not very successful. The price of consumption risk is barely significant.[1] More important are the time series properties of the SDF time series. The actual SDF, *i.e.* the regression residuals $\hat{m}_t = 1 - \hat{\delta}\Delta c_t$, is highly volatile, since consumption news has a standard deviation of 0.78 leading to pricing kernel volatility of $0.32 \times 0.78 = 0.25$. Even through the pricing kernel is only identified up to a scalar multiple (since we work with excess returns), the volatility of m is still the maximum Sharpe ratio for any portfolio. The consumption factor could thus account for a monthly Sharpe ratio of 0.25. However, that portfolio must be perfectly correlated with consumption. By construction the maximally correlated portfolio is the tracking portfolio, which only has a correlation of $0.076^2 = 0.28$ (see table 5). More relevant for pricing, therefore, is the projection on the space of asset returns. For this we construct $u_t = 1 - \hat{\delta}\hat{x}_t^C$, with $\hat{x}_t^C$ the tracking

[1] Due to the scaling of the variables, the coefficient itself (0.32) must be multiplied by 100 to interpret it as the coefficient of relative risk aversion. An estimate around 32 is not unusual, and may be even be on the low side. The consumption series used in the IV estimation as well as in the SDF construction is the sum of the time $t-1$, t, $t+1$ shocks, consistent with the dependent variable in the tracking portfolio construction. This adds roughly another factor $\sqrt{3}$ to the scaling and would bring the CRR estimate up to 55.

Table 6: Discount factor model estimates

	CV				BIC			
	CAPM	C+	macro	all	CAPM	C+	macro	all
Consumption	0.32	-0.51	0.54	-0.61	0.21	0.19	0.16	0.18
	(1.7)	(1.0)	(0.2)	(1.0)	(1.7)	(0.9)	(1.1)	(0.8)
Neg. growth		1.91		2.26		0.05		-0.12
		(1.9)		(1.6)		(0.2)		(0.4)
Inflation			0.30	0.56			0.57	0.39
			(1.2)	(1.6)			(3.8)	(2.3)
Long-term rate			0.12	-0.31			-0.13	-0.06
			(0.4)	(0.7)			(0.6)	(0.3)
Credit spread			-1.22	-0.12			-0.71	-0.82
			(2.0)	(2.40)			(2.0)	(2.0)
Unemployment							-0.12	-2.40
							(0.4)	(4.9)
Neg. unemp.				3.62				6.53
				(2.8)				(6.2)

The table reports estimates of the risk prices δ in the SDF model $m = 1 - \delta' x$ using instrumental variables $\hat{x}$. The instrument for macroeconomic variable x_j are the tracking portfolio returns $\hat{x}_j$. The left side of the table uses instruments based on cross-validation (CV); the right side of the table is based on the BIC for instrument selection. T-statistics are in parentheses.

portfolio excess return. The tracking portfolio has a standard deviation of 0.125, which gives a volatility of only 0.04 for the projected SDF, and therefore a very low maximum Sharpe ratio. Consistent with the equity premium literature consumption as a single factor cannot explain high Sharpe ratios. Using the BIC estimate for δ (0.21) and the standard deviation for the BIC tracking portfolio (0.213) leads to the same volatility for the projected SDF. The difference in the estimates for δ between CV and BIC is thus consistent with the different volatilities of the tracking portfolios.[2]

Adding negative consumption growth to the SDF specification leads to different results for CV and BIC instruments. For the CV tracking portfolio the nonlinearity in consumption growth is almost significant and greatly increases the volatility of the discount factor and its ability to generate high Sharpe ratios. The importance of the negative growth variable is related to the high Sharpe ratio of its tracking portfolio.[3] With the BIC instrument

[2] Technically, the IV estimator for δ in the single factor model equals $\hat{\delta} = \mathrm{E}[\hat{x}]/\operatorname{cov}[x, \hat{x}]$, which implies that the volatility of the SDF satisfies $\rho_{x,\hat{x}} s(m) = \mathrm{E}[\hat{x}]/s(\hat{x})$. The volatility of the SDF is directly related to the Sharpe ratio of the tracking portfolio, with a proportionality constant equal to the correlation between x and its tracking portfolio $\hat{x}$.

[3] In a multivariate IV regression the relation between the discount factor volatility and the tracking

Table 7: Mean-Variance tracking portfolio

	CV				BIC				
	RMSE	trH	SDev	Sh	RMSE	trH	SDev	Sh	ρ
MV	0.563	71.0	0.354	1.69	0.563	60.5	0.358	1.58	0.995

The table shows summary statistics for the mean-variance tracking portfolio constructed using the L_2Boosting algorithm. RMSE is the root-mean-squared-error of the regression residuals; trH is Bühlmann's degrees of freedom parameter k defined in (21); SDev is the standard deviation of the tracking portfolio excess returns; Sh is its Sharpe ratio. For the left side of the table Boosting iterations are terminated by minimizing the test error sum of squares using 5-fold cross-validation. For the right-side of the table the stopping criterium is based on the BIC. In the final column ρ is the correlation between the returns of the BIC and CV tracking portfolios.

selection the tracking portfolio for negative growth loses its high Sharpe ratio (table 5), and hence the estimate for its risk price becomes much smaller and becomes insignificant.

After the detailed discussion of the consumption risk prices, the remainder of table 6 is almost self explanatory. We find large t-statistics for risk prices for those variables that have a tracking portfolio with a high Sharpe ratio. For the CV based instruments the significant risk prices are associated with the *Credit spread* and *Negative unemployment*; for the BIC instruments *Inflation* and *Unemployment* are significant as well. The latter is the exception to the rule, since the Sharpe ratio for the unemployment tracking portfolio is very low. In the multiple regression it is significant due to its correlation with the tracking portfolio for negative unemployment. The two prices have opposite signs, indicating that it is mostly increases in unemployment that are priced.

6.3 Mean-Variance efficient portfolio

For further assessments of the ability of the multivariate SDF models for pricing the cross-section we need the Hansen-Jagannathan distance. For this we first need to construct a mean-variance efficient portfolio by projecting the unit vector on the space of excess returns. Figure 4 shows how the fit of the MV tracking portfolio evolves with the number of boosting steps. Both CV and BIC require many more steps than for any of the macro tracking portfolios. The two criteria produce almost identical portfolios with a correlation above 0.99. The tracking portfolio is very well diversified. It uses 192 out of the 809 available

portfolio Sharpe ratios is not as clear cut as in the univariate regression due to the correlations among tracking portfolios. Under the IV estimator the SDF variance is still a quadratic form in the average returns of the tracking portfolios: $s^2(m) = \mathrm{E}[\hat{x}]' \, \mathrm{E}[\hat{x}x']^{-1} \, \mathrm{E}[xx'] \, \mathrm{E}[x\hat{x}']^{-1} \, \mathrm{E}[\hat{x}]$ Since $\mathrm{E}[x] = 0$ by assumption, the second moments are covariances.

portfolios.[4] The largest absolute weight of a single constituent of the tracking portfolio is only 0.035, with an average absolute weight of 0.012.

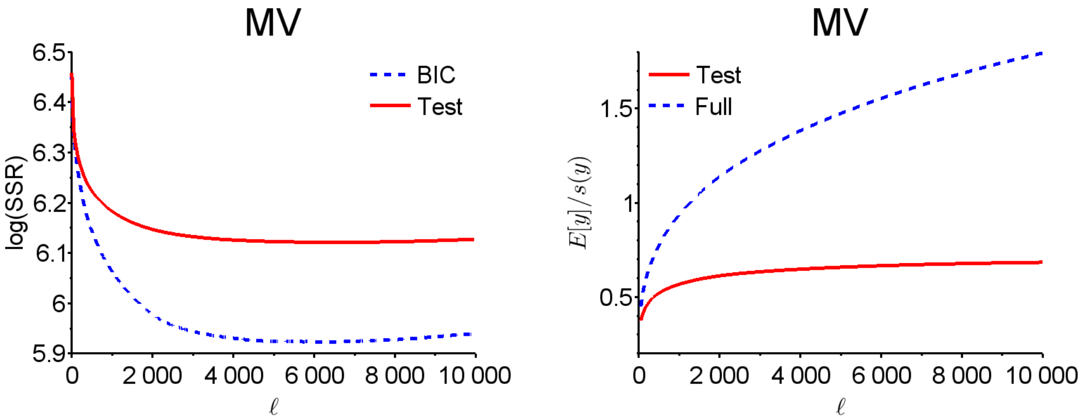
The big challenge to any asset pricing model is the huge Sharpe ratio of the tracking portfolio. Its value (in monthly units) is more than 1.5, more than 6 times as large as the largest single portfolio Sharpe ratio in the database (`ME_VAR_5x5_SMALL.LoVAR`). The tracking portfolio loads heavily on this low-volatility portfolio. Its counterpart, a portfolio with high residual volatility (`ME_RESVAR_5x5_SMALL.HiVAR`) obtains an equally negative weight. In addition there are large positive weights for some of the momentum and reversal portfolios (`ME_Prior_1_0_SMALL.LoPRIOR`, `ME_Prior_12_2_SMALL.HiPRIOR`) and large negative weights for Size/BM and reversal portfolios (`10x10_SMALL.LoBM`, `ME_Prior_1_0_SMALL.HiPRIOR`). Typically, many of the small stock portfolios are selected. One of the more remarkable results is that the seven factor portfolios (5 Fama-French plus Momentum and Reversal) do not obtain strong weights. In fact, the original three FF factors (Market, SMB, HML) are not selected at all and therefore have zero weight. Only the CMA and Momentum factor have a weight close to the average absolute weight.

The huge Sharpe ratio almost surely overestimates the true Sharpe ratio of the MV-portfolio. A clear indication for this is the Sharpe ratio obtained in test samples using cross validation. The graph on the right in figure 4 shows the Sharpe ratio ('Test') with increasing model complexity. For the test samples the portfolio weights are estimated on a random subsample with 80% of the observations, and the Sharpe ratio computed on the fitted portfolio returns in the remaining 20%. This out-of-sample Sharpe ratio is extremely stable around 0.55 irrespective of the model complexity. While still a large Sharpe ratio, it is not anywhere near the in-sample estimate when portfolio weights and performance measures are computed on the same sample.

For the purpose of optimal-portfolio construction the MV portfolio may not be optimal due to the unconstrained nature of its composition. Our interest here is to construct a benchmark for measuring the fit of asset pricing models. As such the MV portfolio generates a time series of excess returns on the mean-variance frontier.

[4] Recall that to avoid near arbitrage about 100 redundant portfolios are deleted when fitting the mean-variance portfolio, see section 5)

Figure 4: Efficient frontier estimates



The left graph shows the fit of the tracking portfolio for the unit vector as a function of the number of boosting steps. The solid line *Test* refers to log of the sum of squared residuals log(SSR) of the 5-fold cross-validation. The dashed line is the BIC criterion. The graph on the right compares the Sharpe ratio of the fitted values in test samples with results for the full sample. A complete test sample is defined as the T observations from all K segments of a single cross-validation. The Sharpe ratio is defined as the (absolute) average return divided by its standard deviation. Test sample results are obtained by averaging over 100 cross-validations. Returns are expressed in percent per month.

6.4 Asset pricing

Knowing the properties of the mean-variance efficient portfolio we can compute the Hansen-Jagannathan distance to evaluate how well the alternative specifications perform. Due to the huge Sharpe ratio of the MV-portfolio it is clear that none of the models can fully explain the cross-section. The absolute values of the HJ-distance may therefore not be very meaningful. Their relative magnitudes still provide an interesting comparison of the different models.

From the results in table 8 we learn that the BIC instruments always fit better than their CV counterparts. With the CV portfolios, adding all variables in the SDF model only leads to a modest 8% decrease in the HJ distance, indicating that the additional macro factors do not improve much on the consumption based model. The only substantial improvement occurs when all BIC instruments are included. Because of the large Sharpe ratio associated with the nonlinear unemployment effect, the HJ distance improves with nearly 20%.

To do: HJ distance on test samples instead of full sample

As a final test of the factor tracking portfolios, and the importance of the macroeconomic

Table 8: Hansen-Jagannathan distance

	CAPM	C+	macro	all
Cross Validation	0.468	0.458	0.454	0.430
BIC	0.439	0.436	0.415	0.360

For each of the SDF models in table 6 the Hansen-Jagannathan distance is computed as $HJ = \hat{u}'\hat{u}/T$ where $\hat{u} = \hat{\iota} - \hat{\boldsymbol{X}}\hat{\delta}$ are residuals of the MV-tracking portfolio returns with respect to the tracking portfolio returns.

factors, we consider their weight in approximating the mean-variance efficient portfolio. To do so we consider the spanning regression (13) with an additional intercept to measure the alpha of the MV portfolio relative to our factor portfolios. If the factors fully explain the cross-section the regression should fit perfectly and the intercept should be zero. What we observe in table 9, as expected, are large highly significant intercepts at the size of more than half a percent per month. Again, this relates to the low Sharpe ratios of the macro tracking portfolios. Combining these macro factors does not bring enough diversification benefits to come near the performance of the boosted mean-variance efficient portfolio.

Despite the high alpha's, the regressions offer some insights in the potential of macro factors. Consumption matters. Both on its own, as well as with a nonlinear effect added, the consumption tracking portfolios explain part of the return on the MV-portfolio. It is especially the nonlinearity, which emphasizes negative consumption growth, that contributes significantly to the return of the MV portfolio. The same nonlinearity also makes unemployment an important contributor. Here the nonlinearity is even more important due the high Sharpe ratio of its tracking portfolio. Inflation and the long-term interest rate are significant as well.

7 Conclusion

By applying a workhorse algorithm from the field of machine learning we address a fundamental econometric challenge in empirical asset pricing: having many assets and short time series. The component-wise L_2-Boosting algorithm is computationally feasible and appears to produce sensible tracking portfolios for macroeconomic variables. Empirically the methodology appears useful for studying the asset pricing implications of macroeconomic economic factors. Theoretically, boosting is known to be consistent for projections even

Table 9: Spanning regressions

	CV				BIC			
	CAPM	C+	macro	all	CAPM	C+	macro	all
Consumption	0.25	-0.34	0.09	-0.56	0.24	0.11	0.22	0.11
	(1.7)	(1.5)	(0.7)	(2.4)	(3.7)	(1.2)	(2.8)	(1.2)
Neg. growth		1.52		1.83		0.28		0.32
		(2.9)		(3.3)		(2.2)		(2.4)
Inflation			0.19	0.27			0.29	0.31
			(1.7)	(2.4)			(3.9)	(4.3)
Long-term rate			-0.11	-0.12			-0.24	-0.26
			(1.0)	(1.1)			(2.4)	(2.7)
Credit spread			-0.50	-0.29			-0.19	-0.08
			(1.9)	(1.1)			(0.9)	(0.4)
Unemployment							-0.05	-0.51
							(0.4)	(2.7)
Neg. unemp.				2.31				6.53
				(3.0)				(6.2)
α	0.59	0.58	0.58	0.57	0.56	0.56	0.55	0.53
	(41.8)	(41.6)	(41.4)	(40.5)	(39.9)	(39.9)	(39.2)	(37.0)
R^2								

The table reports parameter estimates of the time series factor regression $\hat{1}_t = \alpha + \beta' \hat{x}_t$ where $\hat{1}$ is the mean variance efficient portfolio and $\hat{x}_t$ contains the tracking portfolios for the macroeconomic factors. The left side of the table uses tracking portfolio returns based on cross-validation; the right side of the table is based on the BIC for instrument selection. T-statistics are in parentheses.

when N greatly exceeds T under standard regularity conditions.

Using tracking portfolios as optimal instruments we can estimate the risk prices in a stochastic discount factor model. It provides an alternative to the Fama-MacBeth two-stage procedure by applying optimally chosen weights to different assets in the cross section. Assets are weighted according to their power in explaining the time series variation is macro factors. We complement the tracking portfolio methodology by the Hansen-Jagannathan distance for evaluating alternative asset pricing models.

We have focused on the machine learning aspects and explored the empirical properties of tracking portfolios formed by boosting. Standard macro factors remain largely unsuccessful in explaining a mean-variance efficient portfolio. On a more positive note, consumption and other macro factors appear to be priced.

Various issues remain on the agenda. First, since the few standard macro factors do not explain the cross section, there is an obvious interest to extend the factor space to alternative macro factors. Both on the monthly as well as quarterly frequency many more factors are available, for example using the McCracken and Ng (2015) database. The number of factors will quickly proliferate when we also allow interactions and nonlinearities. From our limited experience in this paper, nonlinearities may prove important.

Second, when more macro factors are introduced, there will be a need for a second learning procedure for selecting the most promising factors or transformations and combinations of factors. An example of a transformation would be a habit factor (Campbell and Cochrane, 1999). In addition, Merton (1973) introduces macroeconomic factors as state indicators that provide the pricing kernel with the flexibility to accommodate various nonlinearities and a potentially large number of M potential variables. Our focus in this paper has been on the big-N aspect. When M and N can both be large, there is an interesting double statistical learning element to the model selection.

Third, weak instruments could be a problem when estimating our factor prices. We have paid much attention to the correlation between our macro variables and their tracking portfolios. In the empirics, cross-validation turned out to provide clear warning signs when returns do not correlate with a macro factor, but we did not do a more formal check on the instrument validity of the tracking portfolios.

Fourth, although the asymptotics of the factor price estimator have been established in the literature, a simulation check on the sampling properties in a setting that is relevant for the asset pricing application will be informative, especially in the case of many assets that are weakly correlated with the macro factor. Dependent on the simulation evidence, bootstrap standard error may give a better indication of the strength of a factor. Fifth, as part of the simulation study, it is of interest to compare the tracking portfolio methodology with the two-stage Fama-MacBeth procedure for estimating the risk prices.

Sixth, the full sample mean-variance efficient portfolio that we obtained from the L_2-Boosting algorithm seems a poor benchmark for testing the asset pricing implications. Its huge in-sample Sharpe ratio, many times bigger than any of the macro factor tracking portfolios, grossly overestimates the true Sharpe ratio. For this reason it may be more appropriate to compute the Hansen-Jagannathan distance and spanning alpha's on validation samples created during the cross validation.

References

Bansal, R. and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59, 1481–1509.

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, 2369–2429.

Breiman, L. (1998): "Arcing Classifier (with discussion)," *Annals of Statistics*, 26, 801–849.

Bryzgalova, S. (2016): "Spurious Factors in Linear Asset Pricing Models," Stanford University, working paper.

Bühlmann, P. (2006): "Boosting for High-Dimensional Linear Models," *Annals of Statistics*, 34, 559–583.

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

Campbell, J. and J. Cochrane (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107, 205–251.

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

Cochrane, J. (2011): "Discount Rates," *Journal of Finance*, 56, 1047–1108.

DeMiguel, V., L. Garlappi, and R. Uppal (2007): "Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy?" *Review of Financial Studies*, 22, 1915–1953.

DeMiguel, V., A. Martin-Utrera, F. Nogales, and R. Uppal (2017): "A Portfolio Perspective on the Multitude of Firm Characteristics," SSRN working paper 2912819.

Feng, G., S. Giglio, and D. Xiu (2017): "Taming the Factor Zoo," SSRN working paper 2934020.

Freyberger, J., A. Neuhierl, and M. Weber (2017): "Nonparametric Dissection of the Cross-Section of Expected Stock Returns," NBER working paper 23227.

Friedman, J. (2001): "Greedy Function Approximation: A Gradient Boosting Algorithm," *Annals of Statistics*, 29, 1189–1232.

Hansen, L. and R. Jagannathan (1997): "Assessing Specification Errors in Stochastic Discount Factor Models," *Journal of Finance*, 52, 557–590.

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

Huberman, G. and S. Kandel (1987): "Mean-Variance Spanning," *Journal of Finance*, 42, 873–888.

Jobson, D. and R. Korkie (1982): "Potential Performance and Tests of Portfolio Efficiency," *Journal of Financial Economics*, 10, 433–466.

Jobson, J. and B. Korkie (1983): "Statistical Inference in Two-Parameter Portfolio Theory with Multiple Regression Software," *Journal of Financial and Quantitative Analysis*, 18, 189–197.

Kan, R. and C. Zhang (1999): "Two-Pass Tests of Asset Pricing Models with Useless Factors," *Journal of Finance*, 54, 203–235.

Kelly, B., S. Pruitt, and S. Su (2018): "Characteristics Are Covariances: A Unified Model of Risk and Return," SSRN working paper 3032013.

Kleibergen, F. (2009): "Tests of Risk Premia in Linear Factor Models," *Journal of Econometrics*, 149, 149–173.

Kleibergen, F. and Z. Zhan (2018): "Identification-Robust Inference on Risk Premia of Mimicking Portfolios of Non-traded Factors ," *Journal of Financial Econometrics*, 16, 155–190.

Kozak, S., S. Nagel, and S. Santosh (2017): "Shrinking the Cross Section," SSRN working paper 2945663.

Lamont, O. (2001): "Economic Tracking Portfolios," *Journal of Econometrics*, 105, 161–184.

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

Merton, R. C. (1973): "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867–887.

A Additional figures

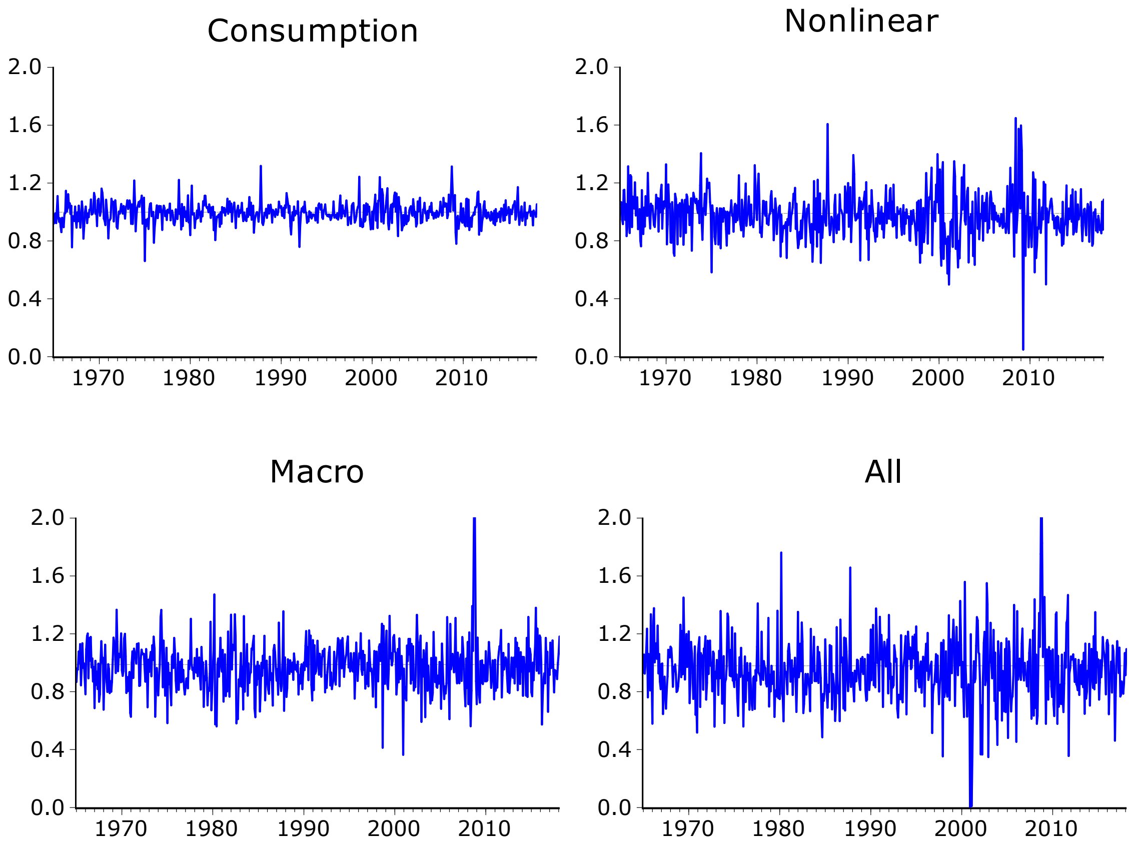


Figure 5: SDF time series

The graphs shows the estimated stochastic discount factors $m_t = 1 - \hat{\delta} x_t$ for the four specifications considered: (*i*) Only consumption growth; (*ii*) Consumption growth plus Negative Growth; (*iii*) Basic macro variables Consumption, Inflation, Credit Spread and Long Term Interest Rate; (*iv*) All variables.

B Simulation design

To study the properties of the L_2-boosting algorithm in our setting we consider the following design. Initially we assume that the asset pricing conditions are valid, meaning that the HJ distance is equal to zero. In the simplest possible setting of any interest we have a single factor model for the pricing kernel

$$m_t = 1 - x_t \tag{25}$$

The systematic risk factor x_t is generated from a distribution with mean zero and variance ω^2. With monthly data the linear approximation to the pricing kernel may always satisfy the condition that the pricing kernel must be positive. The volatility ω is an important design parameter, since the volatility of the pricing kernel determines the size of the premium in returns and with that the information content of average excess returns. The volatility of m is also the maximum Sharpe ratio for any trading strategy. In the data the Sharpe ratio of the market portfolio is about 0.4 for annual data. Assuming that an optimal portfolio may have a double Sharpe ratio equal to 0.8, a reasonable limiting Sharpe ratio for monthly data is $\omega = 0.8/\sqrt{12} = 0.23$.

To satisfy the pricing condition $\mathrm{E}[my] = 0$ excess returns are generated as

$$y_{it} = \beta_i(x_t + \omega^2) + u_{it}, \tag{26}$$

where β_i is the exposure to the macro factor x_t and u_{it} is idiosyncratic risk that is independent of x_t. The idiosyncratic errors have a covariance matrix Σ. With cross-sectional correlation among the u_{it} the term idiosyncratic is not entirely correct. The more precise notion is non-priced risk, meaning that u_{it} is not correlated with x_t and therefore does not affect the mean of y_{it}.

Given the design we can compute the projections $\hat{x} \equiv \mathrm{Proj}(x|y)$ and $\hat{1} \equiv \mathrm{Proj}(1|y)$. In the simulation we compare these to the projections identified by boosting. It will be useful to analyze a few properties of the projections analytically. First we have that

$$\mathrm{E}[xy] = \mathrm{E}[1y] = \beta\omega^2 \tag{27}$$

Next,

$$\mathrm{E}[yy'] = \omega^2(1+\omega^2)\beta\beta' + \Sigma \tag{28}$$

which has the inverse

$$\mathrm{E}[yy']^{-1} = \Sigma^{-1} - \frac{\omega^2(1+\omega^2)}{1+\omega^2(1+\omega^2)B}\Sigma^{-1}\beta\beta'\Sigma^{-1} \tag{29}$$

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

$$h = \mathrm{E}[yy']^{-1}\,\mathrm{E}[xy] = \frac{\omega^2}{1+\omega^2(1+\omega^2)B}\Sigma^{-1}\beta \tag{30}$$

Mean and standard deviation of this portfolio follow as

$$\mathrm{E}[h'y] = \frac{\omega^4 B}{1+\omega^2(1+\omega^2)B} \tag{31}$$

$$s(h'y) = \frac{\omega^2\sqrt{\omega^2B^2 + B}}{1+\omega^2(1+\omega^2)B}, \tag{32}$$

from which we find the Sharpe ratio

$$\mathsf{Sh} = \omega \times \left(\frac{B}{B+1/\omega^2}\right)^{1/2} \tag{33}$$

In normal cases the quadratic form $B = \beta'\Sigma^{-1}\beta$ increases with N.[5] The limiting Sharpe ratio, as $B \to \infty$, is therefore equal to ω, while for all finite N the Sharpe ratio is strictly less than ω.

[5] 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 B need not go to infinity with N, while in the latter case B will be infinite without N going to infinity.

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 $\mathrm{E}[x^2] = \omega^2$. The fitted values have second moment

$$\begin{aligned}\mathrm{E}\left((h'y)^2\right) &= \mathrm{E}[h'y]^2 + s^2(h'y) \\ &= \frac{\omega^4(1+\omega^2)(\omega^2 B^2 + B)}{(1+\omega^2(1+\omega^2)B)^2}\end{aligned} \tag{34}$$

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

$$\lim_{B\to\infty} \frac{\mathrm{E}\left((h'y)^2\right)}{\mathrm{E}[x^2]} = \frac{1}{1+\omega^2} \tag{35}$$

$$\lim_{B\to\infty} \frac{\mathrm{E}\left((h'y)^2\right)}{\mathrm{E}[1^2]} = \frac{\omega^2}{1+\omega^2} \tag{36}$$

In this simulation design these pseudo 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 factor mimicking portfolio.

To study the properties of the projection coefficients h_i we need an assumption on the structure of the covariance matrix of the idiosyncratic risk. A simple assumption is that all u_{it} are equicorrelated with correlation ρ with equal variance σ^2. In that case we have

$$\Sigma = \sigma^2\left((1-\rho)I + \rho\iota\iota'\right) \tag{37}$$

which has the inverse

$$\Sigma^{-1} = \frac{1}{\sigma^2(1-\rho)}\left(I - \frac{\rho}{1-\rho+\rho N}\iota\iota'\right) \tag{38}$$

As a result, by simply substituting and simplifying,

$$B = \frac{N}{\sigma^2}\left(\bar{\beta}^2 + \frac{1}{1-\rho}S_\beta^2\right) \tag{39}$$

$$h_i = \frac{\omega^2}{\sigma^2(1+\omega^2(1+\omega^2)B)}\left(\beta_i + \frac{\rho N}{1-\rho}(\beta_i - \bar{\beta})\right) \tag{40}$$

For $\rho \neq 0$ none of the elements in the projection will vanish. Moreover, they will all be of order $O(1)$. On the other hand, if $\rho = 0$, all elements will be asymptotically negligible. For the performance of the projection algorithm it will therefore be important to minimize the correlation among the non-priced noise components. This can be achieved by selecting a sufficient number of factors in the pricing kernel.

$$\lim_{N\to\infty} h_i = \frac{\rho(\beta_i - \bar{\beta})}{(1+\omega^2)(\bar{\beta}^2(1-\rho) + S_\beta^2)} \tag{41}$$

For the simulation design we consider a number of different cases.

Strong instruments

All beta's are non-zero. When all returns are related to the macro factor, all will be valid instruments. Estimating d should be easy, even though the tracking portfolio may be difficult to estimate. Even with large N, it will be easy to select instruments, if all elements are valid. This design is not empirically not very plausible, since it would mean that the macro factor could be explained very well by the financial returns. From the literature we would expect that much the macro news is uncorrelated with financial returns. . In this case $B \sim O(N)$, and the fit of the tracking portfolios will quickly converge to their limiting values. This normal case does not contain any sparsity, however, and may therefore cause problems for the learning algorithm. All elements in the projections $\hat{1}$ and $\hat{X}$ will be non-negligible.

Many weak instruments

A big concern in studies with macro factors is the low correlation between financial returns and macroeconomic news. Kan and Zhang (1999), Kleibergen (2009) and Bryzgalova (2016) consider the case where the true beta's are either zero or close to zero if the sample size T increases. The latter is called the weak instruments case following the work of Staiger and Stock (1997). Bryzgalova (2016) considers weak instruments in a setting where the number of instruments N is large.

Misspecification

In the previous examples the HJ distance was zero. Misspecification arises if the true pricing kernel would read

$$m_t = 1 - x_{1t} - x_{2t}, \tag{42}$$

with x_1 and x_2 separate priced risk factors, whereas the model would only consider the single factor $x = x_1$. .

With misspecification of the discount factor, the IV estimator for the price of risk will be inconsistent. Assume the factors x_j are independent normal with mean zero and variance ω_j^2.

In our setting we consider the effects of many weak instruments by setting a limit on B as N increases. For a constant limiting B, we specify

$$\beta_i = \sqrt{b_i}\sqrt{N} \tag{43}$$

with b_i from a distribution with mean $\bar{b}$ and variance s_b^2. When N increases it will be harder to find good instruments, even though collectively the instruments explain an almost fixed fraction of the factor news.

For consistency of the boosting algorithm it is important that the DGP is sufficiently sparse. The sparsity condition is

$$\sum_i |h_i| = o\left(\sqrt{\frac{T}{\ln N}}\right) \qquad \text{as} \quad T \to \infty, \tag{44}$$

Another important design parameter is the relative size of systematic and idiosyncratic risk. We fix this by drawing R_i^2, the squared correlation between excess returns and the systematic risk, from a uniform distribution with support $[0, \kappa]$. This implies that the idiosyncratic risk is restricted to

$$\mathrm{E}[u_{it}^2] \equiv \sigma_i^2 = \beta_i^2 \omega^2 \frac{1 - R_i^2}{R_i^2} \tag{45}$$

With the total variance of the idiosyncratic noise fixed, it remains to specify the correlations between the term u_{it} and u_{jt}. The design parameters of the simulation are thus the volatility of the pricing kernel ω^2, the average exposure to systematic risk κ and the correlation of the residuals ρ.

Simulation results

The first experiment checks for biases and properties in a classic case; classic in the sense of small N and strong instruments. We set $N = 25$ and $T = 500$. As design parameters we set

	N	T	ω	σ	ρ	$\bar{\beta}$	s_β	Sh	R_1^2	R_x^2
Small N	25	500	0.231	0.231	0.3	1.0	3.0	0.229	0.05	0.93
Large N	500	500	0.5	0.144	0.3	1.0	3.0	0.230	0.02	0.98

With these choice there is substantial cross-sectional variation in beta's and thus expected returns. In this small N setting the number of boosting iterations is not an issue. The solution converges quickly. Estimates for δ have a slight downward bias, but this bias is the same as the bias in the optimal 2SLS or GMM estimator without instrument selection. Another result is the upward bias in the estimated Sharpe ratio for the estimated MV efficient portfolio h_1. From the population parameters the maximum Sharpe ratio is 0.229, almost equal to its upper bound ω, while the estimated Sharpe ratio from the sample mean and standard deviation of the estimated portfolio $\hat{h}_1' y$

is on average much higher at 0.33. This is the familiar bias analyzed long ago in Jobson and Korkie (1982).

The upward bias is already present after the first few iterations of the boosting algorithm. In the first step boosting selects the asset j which has the highest single asset Sharpe ratio. Even among 25 assets there is enough sampling variation in the excess returns such that the

$$\max_j \frac{\bar{y}_j}{s(y_j)} > \omega$$

To avoid overestimating the Sharpe ratio we need to resort to resampling techniques such as cross-validation. In population the factor tracking portfolio h_x should have the same Sharpe ratio as the MV portfolio. In the simulation they differ considerably. The main difference is that the Sharpe ratio of the factor tracking portfolio is estimated very precisely and almost free of bias.

For the next simulation we increase N to 500. Increasing N means more data to learn about the factor x, but also prohibits the direct least squares solutions for the MV- and tracking portfolios. The stopping criterion for boosting will now become important. The bias for δ drifts downwards to -0.35, underestimating the true price of risk, as the number of boosting iterations increases. As expected the Sharpe ratio for the MV-efficient portfolio drifts upwards. The Sharpe ratio for the factor tracking portfolio slowly decreases with the number of boosting iterations, and is always below the theoretically maximal Sharpe ratio.