Università Ca'Foscari Venezia

Master's Degree

in Economics and Finance

Evaluating Fama&French asset pricing model with Bayesian Additive Regression Trees

Supervisor

Prof. Roberto Casarin

Graduand

Rigers Behluli

Matriculation number

868244

Academic Year

2022/2023

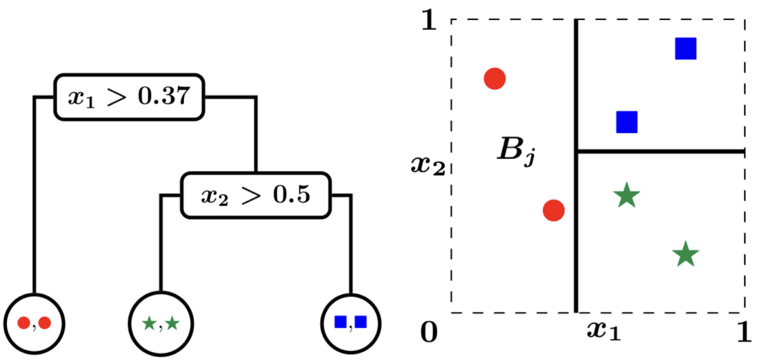
Abstract

Bayesian additive regression tree (BART) models are becoming increasingly popular in literature thank to their flexibility in accounting for many different data features. Many improvements to the simplest BART model have been proposed, such as sparse and smooth BARTs and BART with causal effects. Theoretical properties have been investigated as well. In this thesis I propose an application to economics, through BART estimation of the five factor asset pricing model by Fama & French. The aim is to give evidence in favor or against the correct specification of the model.

1 Introduction

Decision trees are a machine learning technique that has widespread applications in databases, switching theory and many others. Trees work well in problems that need to be evaluated sequentially (see Moret [1982]). To understand how they work, let us introduce a response vector Y and a collection of predictors $X_1, \ldots, X_k$, each of dimension n. A decision tree provides a partition $\mathcal{A} = \{A_1, \ldots, A_j\}$ of the sampling space X such that $Y = j$ if $X \in A_j$. The procedure is performed sequentially and partitions are of the form: if $X < C_j$, then $X \in C_j$. Shifting to trees is straightforward: starting from a node, we can divide it in two child nodes using the aforementioned splitting rule. An example is shown in Figure 1. See Loh [2011] for a more de-

Figure 1: Representation of a simple tree (left) and of the induced partition of the space $[0,1]^2$ (right)



tailed presentation of classification and regression trees.

Literature has shown that introducing randomness in splitting rules and tree structure is a good way to greatly improve prediction performances. Another way to reach this goal is to average the results of many small trees instead of having just one big tree (see for example Breiman [2001] or Lakshminarayanan et al. [2014]). Interesting formulations have been proposed for analysis of survival data (Ishwaran et al. [2008]).

Early combinations of Bayesian analysis and decision trees are called Bayesian CART algorithms and have been separately proposed by Denison et al. [1998] and Chipman et al. [1998]. These simple models put uniform priors on splitting value and splitting rules in such a way that stochastic search guides the model towards more promising trees and then select them according to various criteria as marginal likelihood or misclassification rate. These models have proved to outperform pre-existing alternatives.

Chipman et al. [2010] first introduced and labeled Bayesian additive regression trees (BART). This model, as it's name suggests, works by averaging many regression trees. Choosing suitable prior distributions, each tree in the average is just a part of the predictive function and ensures great flexibility. By randomly assigning parameters to tree structure and terminal nodes, the model is able to efficiently explore parameter space and reaches high predictive performance.

The basic BART model has been extended along many directions. An example are modifications in model specification, aimed at making it suitable for heteroskedasticity (Pratola et al. [2020]) or for survival

data analysis (Sparapani et al. [2016], Basak et al. [2021], Linero et al. [2021]). These models perform BART estimation to compute survival and hazard functions usually used in medical research studies.
Other research has focused on modification of priors. For example, choosing a Dirichlet prior on the probability of selecting a predictor for a new split in the tree, highly improves robustness of the model when noise is introduced in predictor's space (Linero and Yang [2018]). To address the same issue other researchers have proposed to keep track of the probabilities that a certain predictor is selected to compute Metropolis Importance ratios and perform variable selection (Luo and Daniels [2021]). Another interesting formulation is that of Linero [2018] which replaces the deterministic splitting rule $X < C_j$ with a probabilistic one. This induces smoothness in BART predictions, but nevertheless it increases the computational burden. Further research along this direction addressed this issue with targeted smoothing (Starling et al. [2020]), with distributed trees (Ran and Bai [2021a]) or longitudinal regression analysis (Ran and Bai [2021b]).
Some efforts have been done in deriving theoretical properties of BART models, for example explaining how do they adapt to an unknown level of smoothness or how can they perform model reduction when $p > n$ (Ročková and van der Pas [2020]). Other studies proposed models that are better suited for interaction detection among predictors, mainly used to search for causal inference, see for example Du and Linero [2019] or Woody et al. [2020].
Readers can have a look to Linero [2017] for a review of binary tree methods until 2017 and Hill et al. [2020] for the same thing until 2020

and some software implementations.

Most recent work has focused on analysis of longitudinal data (Josefsson et al. [2020]), density regression (Orlandi et al. [2021]), environment interaction (Sarti et al. [2021]), adaptive conditional distribution estimation (Li et al. [2022a]), generalization of the basic model (Linero [2022]).

BART models have had successful applications all over the world and in a wide range of fields. The algorithm has been used in education panel datasets to overcome the issue of temporary and permanent dropout of respondents of the surveys (Zinn and Gnambs [2022]) and to predict impact of remedial support programs on children's reading skills (de Oca et al. [2022])(Syria). Thank to its prediction accuracy, BART had applications also in genomic prediction (GP) studies, a field of study concerned with breeding different species of (for example) crops, in order to create a plant that is more productive and less threatened by insects (Li et al. [2022b])(Australia).

Different studies evaluated government policies and climate impact: Jin et al. [2022] performed a spatio-temporal analysis on noise pollution and environmental risk in urban areas (China); Henderson and Follett [2022] evaluated the effectiveness of targeted social safety programs (Indonesia); Frondel et al. [2022] evaluated government policy on market premium schemes (MPS)(Germany); Obringer et al. [2022] projected the increasing demand of air conditioning in the future in order to quantify associated climate risk (USA); Wong et al. [2022] conducted a significance analysis of factors included in a green government policy (India).

BART models are appropriate also for economic and econometric analysis (Athey and Imbens [2019]). It has been used to perform estimations of efficiency and productivity (Tsionas [2022])(Chile) and to investigate causal effects of diversification strategies on risk-adjusted portfolios Faraji [2022].

2 Model specification

This section presents a detailed specification of the BART model given in Chipman et al. [2010] and the modification called SoftBART by Linero and Yang [2018]. The two models will be used in following sections to perform simulations and applications on real dataset.

2.1 Baseline BART

Consider the problem of estimating an unknown function f_0 given output variables Y and using a p dimensional vector of predictors $X = (x_1, \ldots, x_p)$ when:

$$Y = f_0(x) + \epsilon, \qquad \epsilon \sim N(0, \sigma^2) \tag{1}$$

Chipman et al. [2010] express it as:

$$Y = \sum_{j=1}^{m} g(x; T_j, M_j) + \epsilon, \qquad \epsilon \sim N(0, \sigma^2) \tag{2}$$

where T_j represents a binary regression tree, M_j is the set of terminal node parameters associated to the j-th tree and $g(x; \cdot)$ is the function that assigns $\mu_{ij} \in M_j$ to x, that is:

$$g(x; T_j, M_j) = \sum_{i=1}^{n_j} \mathbb{I}(x \in A_{ji}) \mu_{ij}$$

where A_{ji} is the i-th element of the partition A_j induced by the j-th tree, and n_j is the number of elements of the partition.

Each element in the summation represents just a small part of the whole function f_0 and the influence of single trees is ensured to be small thank to opportune regularization prior specification.

To complete the model we need to set priors for all the parameters in the sum-of-trees model, namely, $(T_1, M_1), \ldots, (T_m, M_m)$ and σ. For T_j three priors are needed: (i) probability that a node is terminal/is a leaf (meaning that it does not split into other child nodes); (ii) a distribution for the splitting variable; (iii) a distribution for the splitting rule to adopt given the splitting value. Chipman et al. [2010] choose uniform priors on (ii) and (iii). For (i) they use

$$\alpha(1+d)^{-\beta}, \qquad \alpha \in (0,1), \beta \in [0,\infty), \tag{3}$$

where d is the depth of a node. In order to keep the individual tree components small they set $\alpha = 0.95$ and $\beta = 2$. The probability of a node to be non-terminal, given its depth is given in Table 1.

Table 1: Probability of a node to be non terminal given Equation 3, $\alpha = 0.95$ and $\beta = 2$

d	1	2	3	4	≥ 5
p	0.05	0.55	0.28	0.09	0.03

For terminal node parameters, $p(\mu_{ij}|T_j)$, they use a conjugate normal distribution $N(\mu_\mu, \sigma_\mu^2)$. Further, for convenience they shift and rescale Y s.t. the observed transformed values y range between $y_{min} = -0.5$ and $y_{max} = 0.5$ and then set

$$\mu_{ij} \sim N(0, \sigma_\mu^2) \qquad \text{where } \sigma_\mu = 0.5/k\sqrt{m}$$

where m is the total number of trees in the ensemble and k has to be tuned.

Last, they assume an inverse chi-square distribution for $\sigma^2 \sim \nu\lambda/\chi^2_\nu$ and m is generally set to 200. See Appendix A for details on prior independence and symmetry.

2.2 SoftBART specification

Linero [2018] proposed a modification of the baseline BART to make it more suitable for variable selection. Linero and Yang [2018] integrated this with a formulation that allows BART to smooth predictions. This model is called SoftBART and it will be used it for simulations and real data analysis through this thesis.

To make BART adapting to sparsity the authors change the distribution of $s = (s_1, \ldots, s_P)$, where s_j is the probability that predictor j is chosen for a given split. In the basic BART $s_j = P^{-1}$, instead s is given a sparsity-inducing Dirichlet distribution:

$$(s_1, \ldots, s_P) \sim \mathcal{D}\left(\frac{\alpha}{P}, \ldots, \frac{\alpha}{P}\right) \quad \text{with} \quad \frac{\alpha}{\alpha + \rho} \sim \text{Beta}(a, b).$$

The prior on α is given to have a fully Bayesian variable selection within the model and ρ is a parameter that controls for a priori beliefs on f_0 (usually set equal to P). Results show that DART (BART with Dirichlet prior) performs much better than BART when noise predictors are included (see Appendix A for some simulation results).

To induce smoothness instead, the step-wise function $\mathbb{I}(X \leq C_b)$ is substituted by a smooth function

$$\psi(x; \mathcal{T}, b) = \psi\left(\frac{x_j - C_b}{\tau_b}\right)$$

which allows for a probabilistic interpretation, where b denotes a branch of the tree and τ_b a bandwidth parameter associated with branch b. This approach makes the path that observation x follows down the tree no more deterministic, rather probabilistic. The result is that trees include information also from different covariate regions and smoothing is determined by local bandwidth parameters tuned from the data. The resulting probability of x ending up in leaf l is

$$\phi(x; \mathcal{T}, l) = \prod_{b \in A(l)} \psi(x; \mathcal{T}, b)^{1-R_b} \{1 - \psi(x; \mathcal{T}, b)\}^{R_b},$$

where $A(l)$ is the set of ancestor nodes of leaf l and $R_b = 1$ if the path to l goes right at b. Authors then propose the use of a logistic gating function $\psi(x) = \{1 + \exp(-x)\}^{-1}$ to complete the model (see Appendix A for simulations on model smoothness).

2.3 Posterior computation

Given observed data $y = (y_1, \ldots, y_n)$, BART induces a posterior of the form

$$p((T_1, M, 1), \ldots, (T_m, M_m), \sigma | y) \tag{4}$$

on the set of parameters of the sum-of-trees model (Equation 2). The algorithm used for computation of Equation 4 is a modification of the basic Gibbs sampler (see Casella and George [1992] and Ritter and Tanner [1992] for more details on the Gibbs sampler).

The sampler used to fit BART is called Bayesian backfitting algorithm and was first introduced by Hastie and Tibshirani [2000]. For notation convenience, define $T_{(j)}$ as the set of all trees in the sum except T_j, and the same for $M_{(j)}$. Thus, $T_{(j)}$ will be a set of $m-1$ trees, and $M_{(j)}$ the associated terminal node parameters. The algorithm iterates the following steps:

i) sample (T_j, M_j) conditionally on $(T_{(j)}, M_{(j)}, \sigma)$:

$$(T_j, M_j)|T_{(j)}, M_{(j)}, \sigma, y \qquad \text{for } j = 1, \dots, m, \tag{5}$$

ii) draw of σ from its full conditional distribution:

$$\sigma|T_1, \dots, T_m, M_1, \dots, M_m, y.$$

Sampling σ is straightforward since it corresponds to a draw from an inverse gamma distribution. To sample (T_j, M_j) instead we observe that the full conditional (Equation 5) depends on $(T_{(j)}, M_{(j)}, y)$ only through

$$R_j \equiv y - \sum_{k \neq j} g(x; T_k, M_k),$$

the n-vector of partial residuals of a simulation that excludes the j-th tree. Thus Equation 5 can be rewritten as

$$(T_j, M_j)|R_j, \sigma, \qquad \text{for } j = 1, \dots, m. \tag{6}$$

Each draw of Equation 6 can be obtained sequentially in two steps

$$T_j|R_j, \sigma \qquad \text{and} \qquad M_j|T_j, R_j, \sigma.$$

For the former, since the model includes a conjugate prior for M_j,

$$p(T_j|R_j,\sigma) \propto p(T_j)\int p(R_j|M_j,T_j,\sigma)p(M_j|T_j,\sigma)dM_j$$

can be obtained in close form by using the Metropolis-Hastings (MH) algorithm of Chipman et al. [1998](see Appendix A for details on that). See Metropolis et al. [1953] and Hastings [1970] for a more general introduction to MH algorithm. Finally the draw of M_j corresponds to independent draws for each of the μ_{ij} leaf node parameters.

This backfitting MCMC algorithm mixes dramatically better compared to single tree models which tend to stabilize to a local mode and its neighborhood. To overcome this issue these models need to be restarted many times and then to average partial results. The Bayesian backfitting instead allows for running a single long chain.

2.4 Priors

To conclude the model specification, I propose in Table 2 an overview of all priors used in the SBART model.

2.5 Simulation results

This section proposes a simulation test of SBART with comparison to other machine learning algorithms. For the comparison I use: SBART[†], BART[‡], Random Forests [§], Lasso [¶] and Gradient-Boosted decision trees[‖].

[†]available at `https://github.com/theodds/SoftBART`
[‡]available at `https://cran.r-project.org/web/packages/BART/index.html`
[§]available at `https://cran.r-project.org/web/packages/randomForest/index.html`
[¶]available at `https://cran.r-project.org/web/packages/glmnet/index.html`
[‖]available at `https://cran.r-project.org/web/packages/xgboost/index.html`

Table 2: Overview of priors used for analysis

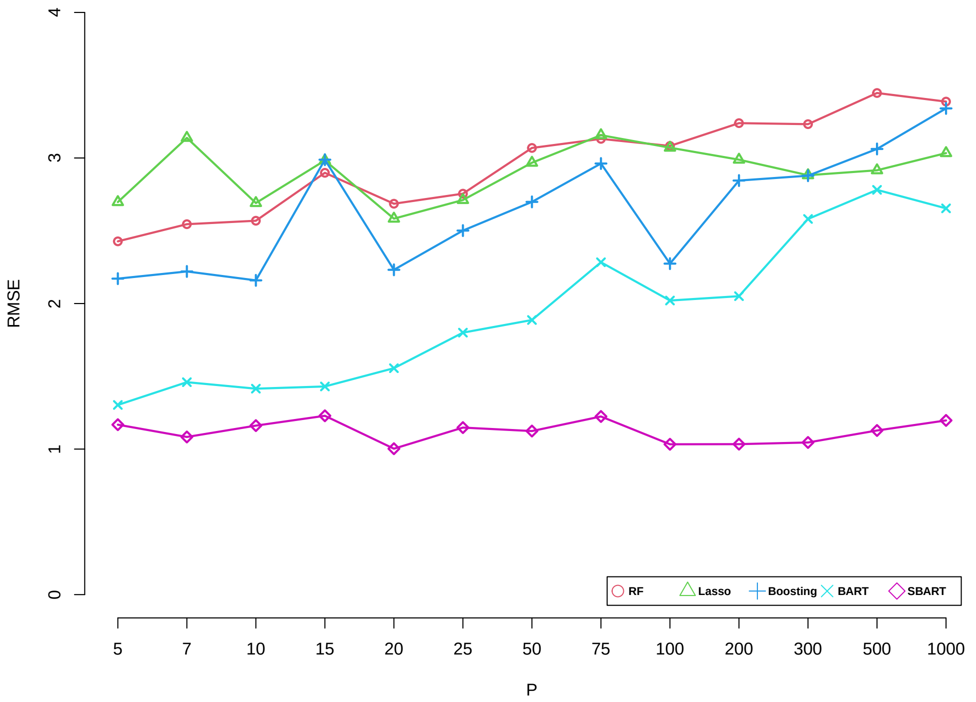
Prior	Values	Explanation
Node Depth $\sim \alpha(1+d)^{-\beta}$	$\alpha = 0.95$, $\beta = 2$	Distribution of Node Depth. Values of α and β are chosen to generate a distribution with high probability of small trees
$\mu_{ij} \sim N(0, \sigma_\mu^2)$	$\sigma_\mu \sim \text{Cauchy}_+(0, 0.25)$	Distribution of leaves of the trees. A Gaussian distribution is chosen to simplify posterior computation. σ_μ is given a default truncated Cauchy distribution with mean $= 0$ and variance $= 0.25$
$\sigma \sim \text{Cauchy}_+(0, \sigma^2_{\text{lasso}})$	σ^2_{lasso}	Distribution of σ of the model. The distribution is a truncated Cauchy with mean 0 and volatility tuned from the data using a lasso estimation
$s \sim \mathcal{D}(a/P, \ldots, a/P)$	a, P	Distribution of the probability of a predictor to be chosen to perform a split on a tree. It is a sparsity inducing Dirichlet distribution, with all parameters $= a/P$. a is given a distribution, P is the total number of predictors
$a/(a+\rho) \sim \text{Be}(c, d)$	ρ =P, $c = 0.5$, $d = 1$	Distribution of parameter a of the Dirichlet distribution above. ρ is a hyperparameter that controls for a priori beliefs on f_0: ρ <P if there is a strong prior belief that f_0 is sparse; ρ =P otherwise.
$\tau_b \overset{\text{iid}}{\sim} \text{Exp}(\nu)$	$\nu = 0.1$	Distribution of the bandwidth parameter. τ_b controls for smoothness of decision, approximating constant models as $\tau_b \to \infty$ and hard decision trees as $\tau_b \to 0$

To compare algorithms I use Friedman [1991] formula and run simulations at increasing number of noise predictors. I use thirteen different levels of total predictors for each model, for a total of 65 simulations. To compare performances RMSE is used as a goodness of fit measure. It is computed as:

$$RMSE = \left\{ \int \{f(x) - \hat{f}(x)\}^2 dx \right\}^{1/2},$$

and is approximated by Monte Carlo integration. All models are used with default settings. Results are shown in Figure 2.

Figure 2: RMSE comparison of BART, SBART, RF, Lasso and Boosting (different lines), at increasing P (horizontal axis)



As we can see from Figure 2, most of considered algorithms lose prediction accuracy at increasing values of P. BART seems to outperform other methods, for each P. SBART results can be compared to those of BART when P is near to the real number of predictors, but is much better when noise predictors are included. Overall, SBART performs much better than any other model considered for each level of P.

3 Asset pricing theory

Asset pricing is one of the main branches of study in economic theory. Investment theory, which is a synonym, entails the knowledge used to support the decision-making process of choosing investments. Models developed so far, can be divided in two main groups: general equilibrium asset pricing and rational pricing. Under the latter, derivatives are priced in a way that they are arbitrage-free with respect to fundamental security prices determined by equilibrium.

Under general equilibrium asset pricing instead prices are determined through the usual market dynamics of supply and demand. The most famous result of market equilibrium is probably the *Proposition I* of Modigliani and Miller [1958] which states that in an environment of efficient markets (i.e. markets in which demand always meets offer) and in absence of taxes, bankruptcy costs and asymmetric information, firm value doesn't depend on how the firm gets financed. The theorem holds thank to some basic assumptions on people's investment behavior on choosing "mean-variance-efficient" portfolio returns: 1) maximize returns, given variance and 2) minimize variance, given returns. This concept was developed by Markowitz [1959]. In his model, the investor chooses a portfolio of investments in $t-1$ and then collects stochastic returns at t and using the criteria stated above for portfolio composition. A branch of literature has studied most efficient ways to construct such portfolios (see for example Clarke et al. [2011]).

One of the main issues was to implement risk in such investment decisions. First, we need to define risk as the impossibility of predicting future market movements, which leads directly to uncertainty in portfolio returns. Agents are assumed to be risk averse, meaning that, if two assets have the same expected return, the one with lower volatility will be included in the portfolio. As a consequence, if the two assets were to be sold at the same price, no one would buy the one with higher volatility, resulting in an impossibility to reach a market equilibrium (markets won't clear since all the demand will be towards the first asset and none for the second). Many authors found a possible solution to this with the concept of risk premium and developing a model named capital asset pricing model (CAPM)(see Treynor [1962], Sharpe [1964], Lintner [1965], Mossin [1966]).

This model solves the issue of defining risk premium by computing it as the excess return of an asset compared to a risk-free asset. The formula is the following:

$$\mathbb{E}(R_t) = R_{ft} + \beta(\mathbb{E}(R_{mt}) - R_{ft}) \tag{7}$$

where $\mathbb{E}(R_t)$ is the expected return of a certain asset at time t, R_{ft} is the return of a risk-free asset in period t, $\mathbb{E}(R_{mt})$ is the expected return of the market at time t. The analysis is carried out through a regression of the form

$$R_t = R_{ft} + a + \beta(R_{mt} - R_{ft}) + \epsilon_t \qquad \text{with } \epsilon_t \sim N(0, \sigma^2).$$

The model states that the return of an asset is equal to the return of a risk-free asset raised by a β factor that multiplies the market risk premium $(R_m - R_f)$. The resulting β is nowadays a commonly used tool of financial analysts for interpreting firm profitability, in fact a higher β denotes that a certain firm is performing better than others with a lower value.

Since the model was first formulated, many critiques have been raised from literature. A first example is that a risk free rate is controversial to determine. To overcome this issue, Black [1972] propose a formulation of the CAPM that does not include the risk-free asset.

However the problems do not stop here, finding a proxy for market returns is also difficult and results of CAPM contrast with empirical evidence (intercept too high and β too small). Estimations in general resulted to be imprecise and residuals presented serial autocorrelation (Fama and French [2004]). All these issues come from a more fundamental critique on the basic assumptions of CAMP. Assumptions in question state that agents share the same expectations and that exists a risk-free rate at which people can lend/borrow money. There is empirical evidence that actually agents to not have homogeneous expectations, which results in a violation of the rationality assumption (Elbannan [2015]). Another big limitation of CAPM is that the agents are not allowed to sell short on assets. Readers can have a look to Campbell [2000] and Barillas and Shanken [2018] for more detailed surveys on CAPM status.

To overcome this issues Ross [2013] proposed a model called Arbitrage

Pricing Theory (APT) which is a factor model of the form

$$R_t = a + b_1 R_{1,t} + \cdots + b_K, R_{K,t} + \epsilon_t$$

where $R_1, \ldots, R_K$ are expected returns of some risk factors usually linked to macroeconomic variables (e.g. the oil price). See also Roll and Ross [1980] for an introduction to APT.

Others have attempted to modify CAPM to make it suitable for analysis in continuous time (Intertemporal-CAPM, Merton [1973]), or linking it to consumption. The latter embodies in the model the fact that agents usually prefer to smooth consumption over time and thus to hedge against that risk. The model in question is called consumption-based CAPM (CCAPM) and was developed by Breeden [2005].

More recent studies proposed a relaxation of the linearity assumptions behind the basic model. For example Tauchen and Hussey [1991] proposed a quadrature-based model that is used to approximate different asset pricing models. Dittmar [2002] use non-linear pricing kernels to create a model that works like non-parametric ones. Chen et al. [2011] use a Bayesian-GARCH algorithm to model residuals of the basic CAPM model.

Another branch of literature developed models which adjust predictions of CAPM adding some factors. In order, we have a two-factor model by Black et al. [1972], a three-factor model by Fama and French [1992], a four-factor model by Carhart [1997] and a five-factor model by Fama and French [2015]. The model that I use for my analysis is the last one. In next section I discuss more in detail the factors added by the authors to the basic CAPM.

3.1 Fama & French five-factor model

In Fama and French [1992] authors test CAPM on US stock returns data in the period 1963-1990. Their tests move against the central prediction of CAPM that market β is positively related to average stock returns. Following empirical evidence of a positive link between average returns and size and book-to-market (B/M) equity, they propose adding two other factors to the basic model:

· SMB "Small Minus Big [market capitalisation]" which is a measure of the historical difference between returns of small caps and big caps;

· HML "High Minus Low [book-to-market ratio]", difference between returns of firms with high and low book-to-market ratios.

The model performs better than the basic CAPM, but still presents some criticisms. Titman et al. [2004] and Novy-Marx [2013] pointed out that the three-factor model is incomplete because it doesn't consider the relation between average returns and operating profitability and investment. To overcome this issue Fama and French [2015] propose the addition of two further variables to the three-factor model:

· RMW "Robust Minus Weak [operating profitability]" which is a measure of the historical difference between returns of high vs. low operating profitability stocks;

· CMA "Conservative Minus Aggressive [investment]", difference between returns of conservative (low investment) vs. aggressive (high investment) stocks.

The five-factor model then can be represented as:

$$R_t - R_{ft} = a + b(R_{mt} - R_{ft}) + c(SMB_t) + d(HML_t) + \tag{8}$$

$$+e(RMW_t) + f(CMA_t) + \epsilon_t \qquad \epsilon_t \sim N(0, \sigma^2)$$

The authors propose different ways of combining portfolios to construct factors above, details can be found in their paper. Their main results show that the *GRS* test rejects the five-factor model, but they estimate that it can explain between 71% and 94% of the cross-section they examine. They also find that the way in which factors are constructed, doesn't affect much the outcome of the regressions. Last, according to their analysis, HML is a redundant factor and consequently a four-factor model which drops it, is effective as the five-factor one (see following section for regressions using CAPM, the three and the five-factor model).

3.2 Regressions

To get a better insight of CAPM and Fama&French factor models, I propose a regression analysis of the data listed in Appendix B. I present results for regressions using CAPM, F&F three and five factor, F&F five factor + Momentum factor, and a series of regressions in which SMB is always included and one of the other factors not, as in Fama and French [2015]. Here is a list of models used in the analysis:

Model 1 (CAPM)

$$R_t - R_{ft} = a + \beta(R_{mt} - R_{ft}) + \epsilon_t \tag{9}$$

Model 2 (F&F three factors)

$$R_t - R_{ft} = a + \beta(R_{mt} - R_{ft}) + c(SMB_t) + d(HML_t) + \epsilon_t \quad (10)$$

Model 3

$$R_t - R_{ft} = a + \beta(R_{mt} - R_{ft}) + c(SMB_t) + d(HML_t) + e(RMW_t) + \epsilon_t \quad (11)$$

Model 4

$$R_t - R_{ft} = a + \beta(R_{mt} - R_{ft}) + c(SMB_t) + d(HML_t) + e(CMA_t) + \epsilon_t \quad (12)$$

Model 5

$$R_t - R_{ft} = a + \beta(R_{mt} - R_{ft}) + c(SMB) + d(RMW_t) + e(CMA_t) + \epsilon_t \quad (13)$$

Model 6 (F&F five factors)

$$R_t - R_{ft} = a + \beta(R_{mt} - R_{ft}) + c(SMB_t) + d(HML_t) + e(RMW_t) + \quad (14)$$

$$+ f(CMA_t) + \epsilon_t$$

Model 7 (F&F five factors + Momentum factor)

$$R_t - R_{ft} = a + \beta(R_{mt} - R_{ft}) + c(SMB_t) + d(HML_t) + \quad (15)$$

$$+ e(RMW_t) + f(CMA_t) + g(Mom_t) + \epsilon_t$$

I apply each model to each of the five industry portfolios listed in

Data source section, for a total of 35 regressions. Results are shown in following tables. For each factor is reported the coefficient and significance, with sign. codes: (***) 0, (**) 0.01, (*) 0.05, (.) 0.1, ( ) 1. For each model, also R^2, adj-R^2 and F-statistic values are included.

Table 3: Regressions for Cnsmr returns data. For each factor is reported the coefficient and significance, with sign. codes: (***) 0, (**) 0.01, (*) 0.05, (.) 0.1, ( ) 1. R2, adj-R2 and F-statistic

Factor	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Interc.	0.02 ***	0.02 ***	0.01 ***	0.01 ***	0.01 ***	0.01 ***	0.01 ***
$R_m - R_f$	0.83 ***	0.85 ***	0.87 ***	0.87 ***	0.89 ***	0.89 ***	0.89 ***
SMB	- -	0.33 ***	0.38 ***	0.34 ***	0.39 ***	0.39 ***	0.39 ***
HML	- -	0.08 ***	0.07 ***	0.02 ***	- -	0.02 ***	-0.01 *
RMW	- -	- -	0.25 ***	- -	0.25 ***	0.25 ***	0.25 ***
CMA	- -	- -	- -	0.17 ***	0.18 ***	0.16 ***	0.18 ***
Mom	- -	- -	- -	- -	- -	- -	-0.05 ***
R^2	0.779	0.818	0.827	0.820	0.829	0.829	0.830
adj-R^2	0.779	0.818	0.827	0.820	0.829	0.829	0.830
F-stat.	1.0e+05 ***	4.4e+04 ***	3.5e+04 ***	3.4e+04 ***	3.6e+04 ***	2.9e+04 ***	2.4e+04 ***

Tables 3 - 7 show that for each dataset and for each model, all factors are significant at a level <0.001 with just three exceptions: HML of model 7 in Cnsmr data, the intercept of models 3 to 7 in Manuf data and Mom of model 7 in Other data.

Fama&French three factors (Model 2) performs much better than

Table 4: Regressions for Manuf returns data. For each factor is reported the coefficient and significance, with sign. codes: (***) 0, (**) 0.01, (*) 0.05, (.) 0.1, ( ) 1. R2, adj-R2 and F-statistic

Factor	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Interc.	0.02 ***	0.01 ***	0.005 .	0.006 *	0.002	0.002	0.004
$R_m - R_f$	0.89 ***	0.93 ***	0.95 ***	0.95 ***	0.97 ***	0.97 ***	0.96 ***
SMB	- -	0.32 ***	0.38 ***	0.33 ***	0.41 ***	0.38 ***	0.38 ***
HML	- -	0.31 ***	0.30 ***	0.23 ***	- -	0.23 ***	0.19 ***
RMW	- -	- -	0.24 ***	- -	0.25 ***	0.24 ***	0.24 ***
CMA	- -	- -	- -	0.23 ***	0.42 ***	0.22 ***	0.25 ***
Mom	- -	- -	- -	- -	- -	- -	-0.06 ***
R^2	0.762	0.826	0.833	0.830	0.826	0.837	0.839
adj-R^2	0.762	0.826	0.833	0.830	0.826	0.837	0.839
F-stat.	9.5e+04 ***	4.7e+04 ***	3.7e+04 ***	3.6e+04 ***	3.5e+04 ***	3.1e+04 ***	2.6e+04 ***

Table 5: Regressions for HiTec returns data. For each factor is reported the coefficient and significance, with sign. codes: (***) 0, (**) 0.01, (*) 0.05, (.) 0.1, ( ) 1. R2, adj-R2 and F-statistic

Factor	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Interc.	0.02 ***	0.02 ***	0.03 ***	0.03 ***	0.03 ***	0.03 ***	0.04 ***
$R_m - R_f$	1.06 ***	1.05 ***	1.02 ***	1.02 ***	0.98 ***	0.99 ***	0.99 ***
SMB	- -	0.43 ***	0.35 ***	0.42 ***	0.31 ***	0.34 ***	0.34 ***
HML	- -	-0.33 ***	-0.32 ***	-0.24 ***	- -	-0.23 ***	-0.26 ***
RMW	- -	- -	-0.41 ***	- -	-0.41 ***	-0.40 ***	-0.40 ***
CMA	- -	- -	- -	-0.27 ***	-0.46 ***	-0.26 ***	-0.24 ***
Mom	- -	- -	- -	- -	- -	- -	-0.05 ***
R^2	0.772	0.824	0.839	0.829	0.835	0.843	0.844
adj-R^2	0.772	0.824	0.839	0.829	0.835	0.843	0.844
F-stat.	1.0e+05 ***	4.6e+04 ***	3.9e+04 ***	3.6e+04 ***	3.7e+04 ***	3.2e+04 ***	2.7e+04 ***

Table 6: Regressions for Hlth returns data. For each factor is reported the coefficient and significance, with sign. codes: (***) 0, (**) 0.01, (*) 0.05, (.) 0.1, ( ) 1. R2, adj-R2 and F-statistic

Factor	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Interc.	0.03 ***	0.03 ***	0.03 ***	0.03 ***	0.03 ***	0.03 ***	0.03 ***
$R_m - R_f$	0.87 ***	0.86 ***	0.86 ***	0.88 ***	0.87 ***	0.88 ***	0.88 ***
SMB	- -	0.37 ***	0.36 ***	0.38 ***	0.32 ***	0.37 ***	0.37 ***
HML	- -	-0.29 ***	-0.29 ***	-0.36 ***	- -	-0.36 ***	-0.36 ***
RMW	- -	- -	-0.03 ***	- -	-0.05 ***	-0.04 ***	-0.04 ***
CMA	- -	- -	- -	0.23 ***	-0.08 ***	0.23 ***	0.23 ***
Mom	- -	- -	- -	- -	- -	- -	0.004
R^2	0.634	0.680	0.680	0.684	0.660	0.684	0.684
adj-R^2	0.634	0.680	0.680	0.684	0.660	0.684	0.684
F-stat.	5.1+04 ***	2.1e+04 ***	1.6e+04 ***	1.6e+04 ***	1.4e+04 ***	1.3e+04 ***	1.1e+04 ***

Table 7: Regressions for Other returns data. For each factor is reported the coefficient and significance, with sign. codes: (***) 0, (**) 0.01, (*) 0.05, (.) 0.1, ( ) 1. R2, adj-R2 and F-statistic

Factor	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Interc.	0.02 ***	0.01 ***	0.008 ***	0.01 ***	0.01 ***	0.01 ***	0.01 ***
$R_m - R_f$	0.91 ***	0.96 ***	0.96 ***	0.94 ***	0.96 ***	0.94 ***	0.94 ***
SMB	- -	0.40 ***	0.41 ***	0.39 ***	0.46 ***	0.41 ***	0.41 ***
HML	- -	0.39 ***	0.39 ***	0.46 ***	- -	0.46 ***	0.41 ***
RMW	- -	- -	0.07 ***	- -	0.09 ***	0.07 ***	0.08 ***
CMA	- -	- -	- -	-0.20 ***	0.20 ***	-0.20 ***	-0.17 ***
Mom	- -	- -	- -	- -	- -	- -	-0.07 ***
R^2	0.772	0.869	0.870	0.872	0.830	0.873	0.875
adj-R^2	0.772	0.869	0.870	0.872	0.830	0.873	0.875
F-stat.	1.0e+05 ***	6.6e+04 ***	5.0e+04 ***	5.1e+04 ***	3.6e+04 ***	4.1e+04 ***	3.5e+04 ***

CAPM (Model 1) for every dataset in terms of R^2, increasing it's value by roughly 0.05, meaning that SMB and HML add explanatory power to the basic model. Adding other factors instead, seem to have a really poor effect on the goodness of fit indicator, leaving it's value almost unchanged.

The adj-R^2 indicator values are the same as R^2 since the weighting effect of adding new variables to the model is overtaken by the large number of observations (29662).

The F-statistic shows that all models are statistically significant. However the value of the statistic decreases when adding factors to the model, indicating that the significance of the models is decreasing. To test this, I performed F tests using the usual formula

$$F_{test} = \frac{(RRS_r - RRS_u)/r}{RRS_u/(n-k)} = \frac{((1-R_r^2)-(1-R_u^2))/r}{(1-R_u^2)/(n-k)} \quad \sim F_{r,n-k} \tag{16}$$

where RRS_r and RRS_u are the residual sum of squares of the restricted and unrestricted model respectively, r is the number of restrictions, n is the number of observations and k is the number of predictors of the unrestricted model. I performed the tests keeping Model 7 as the unrestricted one and then alternating Models 1 to 6 as the restricted ones. I omit reporting results since all tests reject the null hypothesis of non-significance of the unrestricted model at a significance level <0.001.

However, given the behaviour of the F-statistic and Fama and French [2015] arguing that, according to their analysis, a four-factor model which drops HML is as effective as the five-factor model, a more detailed test on significance of the factors is needed and I do that in this thesis using SBART model.

4 Model application and results

The analysis conducted in previous section raises the necessity to test the significance of Fama&French factors in a context different from basic linear regression. To do so, I propose an analysis of posterior probabilities on inclusion in the SBART model of factors, with inclusion of noise variables (in a fashion similar to the analysis of Appendix A, Figure 11).
The data I use are the same described in Appendix B. For the analysis I use the routine setting of SBART described in Linero [2018], with: n. of trees = 50, n. of burn-in iterations = 500 and n. of posterior iterations = 500 (this choice is made to keep acceptable the computational time required for the model to carry out the analysis). To run the model I leave the last 1000 observations as a test and use the others for training. To each model I attach ten noise variables drawn from a uniform distribution between the min and max value that can be found in the set of factors belonging to the model.
Tables 8 to 14 report the posterior probability of inclusion of factors and noise variables for each Model and dataset.
Table 8 shows that in Model 1, the factor $(R_m - R_f)$ has a substantially higher probability of inclusion compared to noise variables. However there is a negative trend as other factors are added. Table 14 shows that the probability of inclusion of factors is slightly higher than the one of noise variables for most of them, meaning that they lost their

explanatory power. This can be due to the fact that the model splits probability of inclusion into more variables, however simulations presented in Appendix A show that BART with Dirichlet prior, that is included also in SBART, adapts very well to sparsity in the data, bringing the conclusion that the more factors are included in the model, the more it gets overfitted. This result is in line with findings about F-statistic in previous section.

Results vary a lot between datasets and Models, making it difficult to identify which factors are the most important. The only one that always keeps a high probability of inclusion compared to noise variables and that many times overtakes other factors is the excess market return $(R_m - R_f)$. Evidence therefore is in favor of the basic CAPM and against Fama&French models that add factors to the basic setting. See Appendix C for RMSE computation and plots of results.

Table 8: Asset pricing application. Model 1. For each factor and for all noise variables is reported the posterior probability of inclusion in the SBART model

Factor	Cnsmr	Manuf	HiTec	Hlth	Other
$R_m - R_f$	**0.405**	**0.321**	**0.599**	**0.617**	**0.347**
Noise 1	0.085	0.068	0.003	0.044	0.042
Noise 2	0.063	0.069	0.044	0.098	0.071
Noise 3	0.048	0.048	0.042	0.002	0.086
Noise 4	0.041	0.070	0.020	0.021	0.074
Noise 5	0.057	0.079	0.000	0.006	0.080
Noise 6	0.034	0.095	0.096	0.058	0.041
Noise 7	0.010	0.061	0.022	0.023	0.064
Noise 8	0.068	0.073	0.024	0.060	0.048
Noise 9	0.058	0.057	0.058	0.029	0.075
Noise 10	0.130	0.060	0.090	0.043	0.072

Table 9: Asset pricing application. Model 2. For each factor and for all noise variables is reported the posterior probability of inclusion in the SBART model

Factor	Cnsmr	Manuf	HiTec	Hlth	Other
$R_m - R_f$	**0.204**	**0.204**	**0.160**	**0.251**	**0.196**
SMB	0.179	0.105	0.144	0.157	0.112
HML	0.152	0.168	0.193	0.187	0.148
Noise 1	0.025	0.032	0.032	0.055	0.069
Noise 2	0.049	0.034	0.074	0.054	0.059
Noise 3	0.042	0.054	0.043	0.026	0.044
Noise 4	0.066	0.056	0.034	0.022	0.054
Noise 5	0.041	0.067	0.044	0.014	0.081
Noise 6	0.070	0.028	0.074	0.023	0.046
Noise 7	0.062	0.083	0.073	0.031	0.076
Noise 8	0.026	0.079	0.057	0.036	0.065
Noise 9	0.028	0.023	0.023	0.052	0.018
Noise 10	0.057	0.059	0.014	0.091	0.033

Table 10: Asset pricing application. Model 3. For each factor and for all noise variables is reported the posterior probability of inclusion in the SBART model

Factor	Cnsmr	Manuf	HiTec	Hlth	Other
$R_m - R_f$	**0.149**	**0.217**	**0.174**	**0.210**	**0.193**
SMB	0.108	0.095	0.133	0.112	0.117
HML	0.115	0.092	0.116	0.148	0.107
RMW	0.116	0.137	0.134	0.174	0.064
Noise 1	0.055	0.049	0.020	0.058	0.052
Noise 2	0.046	0.070	0.041	0.056	0.041
Noise 3	0.058	0.046	0.024	0.024	0.050
Noise 4	0.079	0.045	0.078	0.015	0.056
Noise 5	0.043	0.024	0.034	0.026	0.043
Noise 6	0.037	0.040	0.052	0.039	0.051
Noise 7	0.055	0.049	0.050	0.035	0.059
Noise 8	0.044	0.084	0.061	0.015	0.054
Noise 9	0.061	0.029	0.031	0.047	0.054
Noise 10	0.035	0.021	0.052	0.042	0.058

Table 11: Asset pricing application. Model 4. For each factor and for all noise variables is reported the posterior probability of inclusion in the SBART model

Factor	Cnsmr	Manuf	HiTec	Hlth	Other
$R_m - R_f$	**0.163**	**0.178**	0.160	**0.199**	**0.180**
SMB	0.148	0.144	**0.193**	0.089	0.085
HML	0.075	0.143	0.119	0.168	0.132
CMA	0.115	0.089	0.133	0.120	0.114
Noise 1	0.054	0.028	0.016	0.043	0.025
Noise 2	0.056	0.039	0.056	0.067	0.029
Noise 3	0.019	0.053	0.035	0.034	0.057
Noise 4	0.050	0.041	0.039	0.050	0.044
Noise 5	0.060	0.061	0.026	0.046	0.063
Noise 6	0.029	0.038	0.043	0.055	0.081
Noise 7	0.060	0.062	0.058	0.044	0.041
Noise 8	0.040	0.045	0.037	0.006	0.055
Noise 9	0.064	0.027	0.030	0.042	0.036
Noise 10	0.067	0.052	0.056	0.037	0.058

Table 12: Asset pricing application. Model 5. For each factor and for all noise variables is reported the posterior probability of inclusion in the SBART model

Factor	Cnsmr	Manuf	HiTec	Hlth	Other
$R_m - R_f$	**0.172**	**0.199**	**0.179**	**0.197**	**0.153**
SMB	0.145	0.164	0.122	0.152	0.145
RMW	0.100	0.133	0.113	0.115	0.140
CMA	0.100	0.100	0.135	0.140	0.072
Noise 1	0.057	0.031	0.044	0.035	0.071
Noise 2	0.045	0.035	0.022	0.020	0.049
Noise 3	0.040	0.041	0.039	0.018	0.063
Noise 4	0.050	0.023	0.060	0.064	0.066
Noise 5	0.056	0.050	0.024	0.050	0.037
Noise 6	0.031	0.040	0.043	0.048	0.044
Noise 7	0.045	0.020	0.067	0.041	0.027
Noise 8	0.076	0.056	0.061	0.031	0.041
Noise 9	0.046	0.060	0.056	0.025	0.044
Noise 10	0.035	0.048	0.035	0.064	0.050

Table 13: Asset pricing application. Model 6. For each factor and for all noise variables is reported the posterior probability of inclusion in the SBART model

Factor	Cnsmr	Manuf	HiTec	Hlth	Other
$R_m - R_f$	**0.136**	0.131	**0.146**	0.143	0.152
SMB	0.115	**0.133**	0.143	0.129	**0.171**
HML	0.099	0.122	0.086	0.070	0.124
RMW	0.072	0.095	0.070	**0.165**	0.074
CMA	0.081	0.074	0.135	0.118	0.071
Noise 1	0.062	0.054	0.038	0.037	0.024
Noise 2	0.054	0.036	0.040	0.014	0.019
Noise 3	0.054	0.038	0.037	0.037	0.044
Noise 4	0.032	0.040	0.038	0.087	0.033
Noise 5	0.028	0.066	0.071	0.037	0.035
Noise 6	0.058	0.036	0.014	0.038	0.075
Noise 7	0.061	0.025	0.057	0.046	0.038
Noise 8	0.045	0.066	0.018	0.015	0.060
Noise 9	0.057	0.038	0.044	0.026	0.036
Noise 10	0.046	0.048	0.063	0.038	0.045

Table 14: Asset pricing application. Model 7. For each factor and for all noise variables is reported the posterior probability of inclusion in the SBART model

Factor	Cnsmr	Manuf	HiTec	Hlth	Other
$R_m - R_f$	**0.145**	**0.160**	**0.167**	**0.150**	**0.154**
SMB	0.076	0.094	0.132	0.108	0.0107
HML	0.101	0.082	0.092	0.093	0.091
RMW	0.075	0.074	0.095	0.117	0.068
CMA	0.065	0.095	0.090	0.129	0.069
Mom	0.078	0.093	0.063	0.102	0.078
Noise 1	0.045	0.018	0.053	0.033	0.040
Noise 2	0.037	0.022	0.037	0.030	0.075
Noise 3	0.066	0.028	0.022	0.023	0.056
Noise 4	0.045	0.034	0.021	0.024	0.030
Noise 5	0.034	0.072	0.047	0.009	0.035
Noise 6	0.031	0.037	0.038	0.030	0.041
Noise 7	0.037	0.031	0.029	0.023	0.041
Noise 8	0.081	0.054	0.041	0.060	0.025
Noise 9	0.041	0.046	0.037	0.031	0.052
Noise 10	0.042	0.059	0.037	0.037	0.045

4.1 Partial dependence analysis

Analysis so far identifies the excess market return $(R_m - R_f)$ as the most important factor among those considered. To better understand its influence on the estimated f_0, I propose a partial dependence analysis of it relative to each model considered. The analysis is carried out using a function already present in the SoftBART package, called "pdsoftbart".

I set number of trees, number of burn-in iterations and number of save iterations as the ones in previous section: 50, 500 and 500 respectively. For each model I indicate $(R_m - R_f)$ as the factor to consider when computing partial dependence. The function returns a plot of the partial dependence of estimated f_0 (y-axis) against $(R_m - R_f)$(x-axis) (see Appendix C for simulations on this using Friedman [1991] formula). Results are plotted in Figures 3 to 9.

Figure 3: Asset pricing application. For Model 1, plot of partial dependence of f_0 (y-axis) against $(R_m - R_f)$ (x-axis), for the following sectors: consumer (Cnsmr), manufacturing (Manuf), technology (HiTec), health (Hlth) and others (Other)

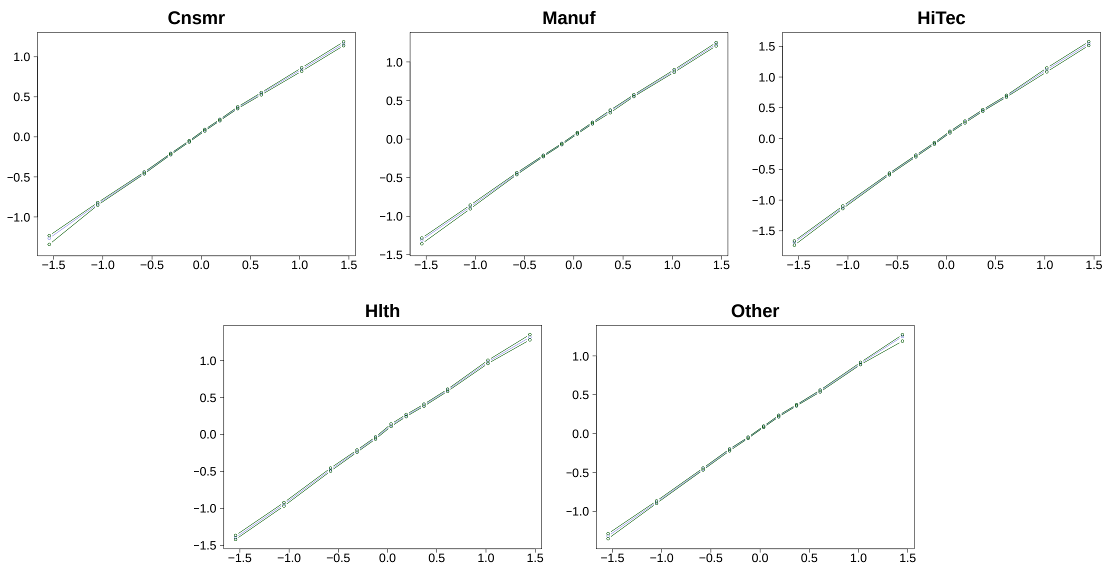


Figure 4: Asset pricing application. For Model 2, plot of partial dependence of f_0 (y-axis) against $(R_m - R_f)$ (x-axis), for the following sectors: consumer (Cnsmr), manufacturing (Manuf), technology (HiTec), health (Hlth) and others (Other)

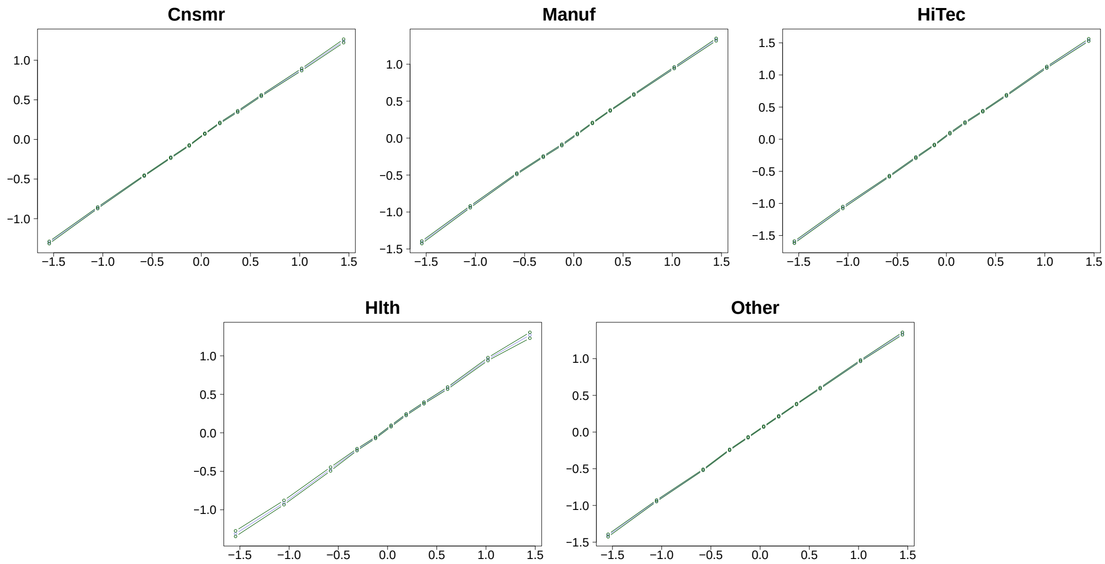


Figure 5: Asset pricing application. For Model 3, plot of partial dependence of f_0 (y-axis) against $(R_m - R_f)$ (x-axis), for the following sectors: consumer (Cnsmr), manufacturing (Manuf), technology (HiTec), health (Hlth) and others (Other)

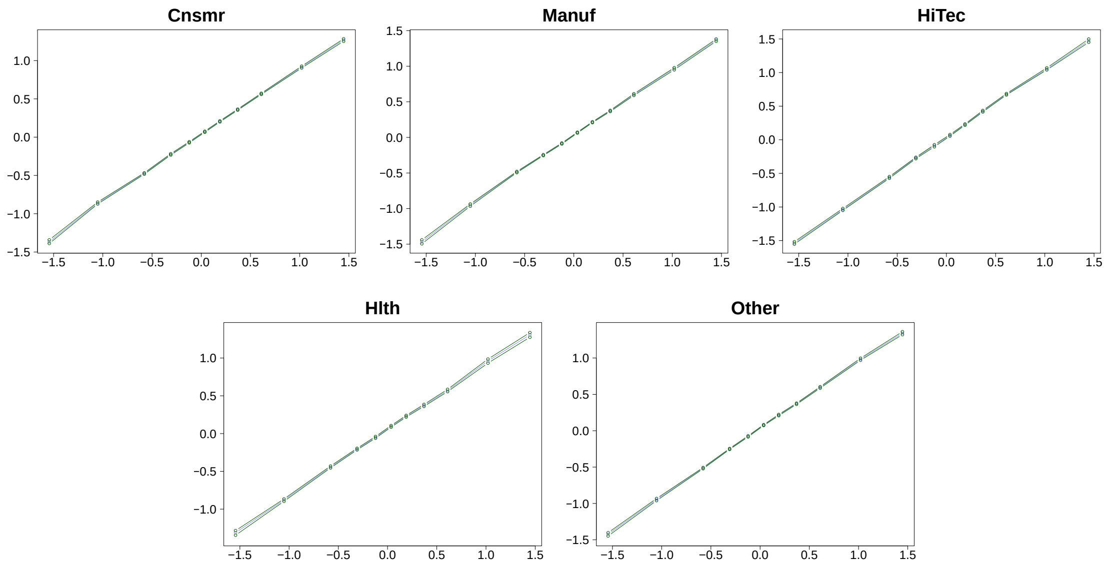


Figure 6: Asset pricing application. For Model 4, plot of partial dependence of f_0 (y-axis) against $(R_m - R_f)$ (x-axis), for the following sectors: consumer (Cnsmr), manufacturing (Manuf), technology (HiTec), health (Hlth) and others (Other)

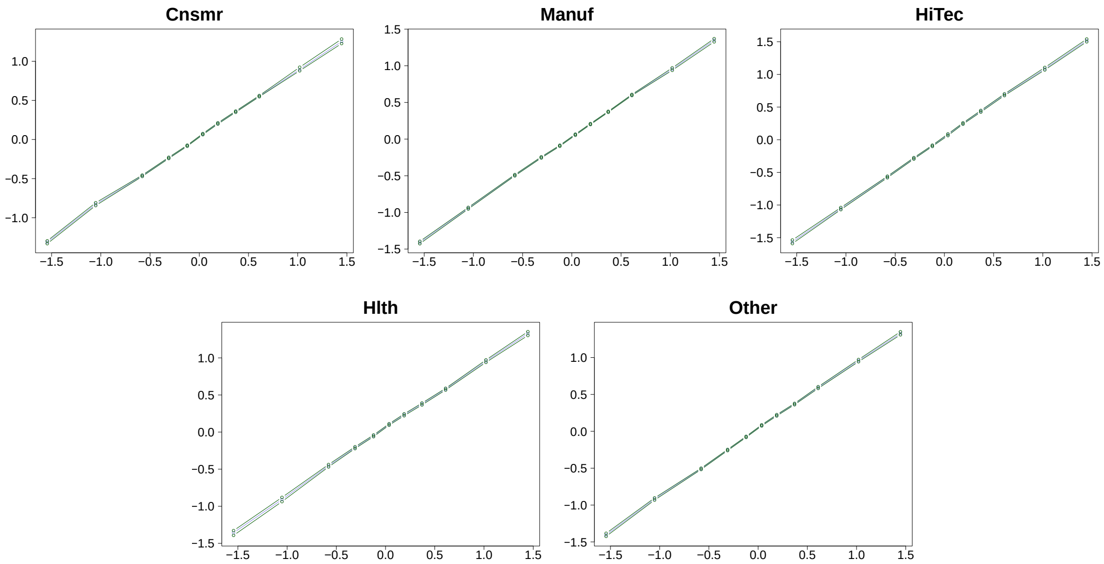


Figure 7: Asset pricing application. For Model 5, plot of partial dependence of f_0 (y-axis) against $(R_m - R_f)$ (x-axis), for the following sectors: consumer (Cnsmr), manufacturing (Manuf), technology (HiTec), health (Hlth) and others (Other)

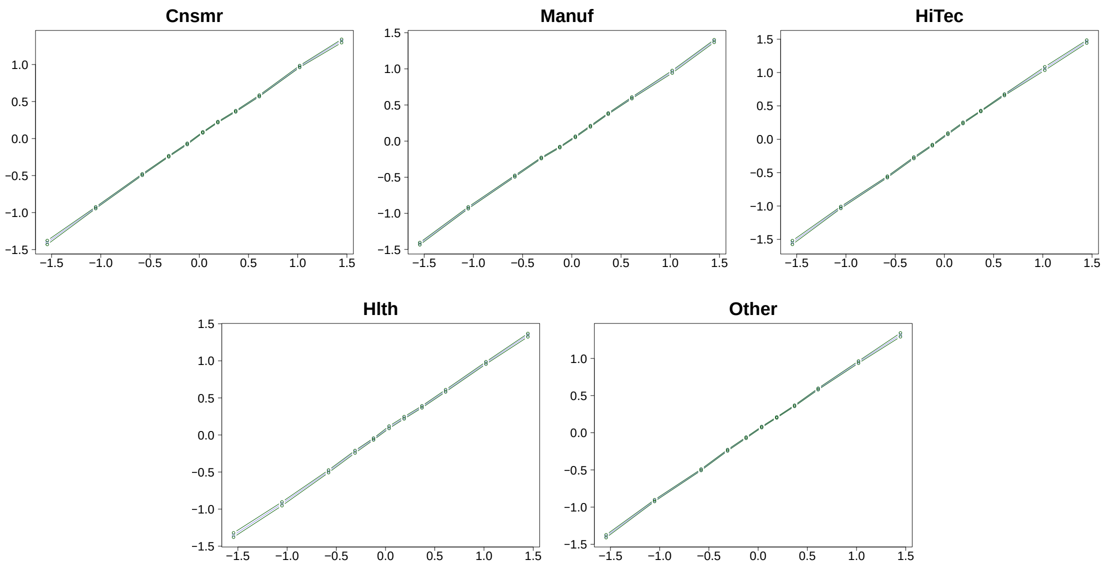


Figure 8: Asset pricing application. For Model 6, plot of partial dependence of f_0 (y-axis) against $(R_m - R_f)$ (x-axis), for the following sectors: consumer (Cnsmr), manufacturing (Manuf), technology (HiTec), health (Hlth) and others (Other)

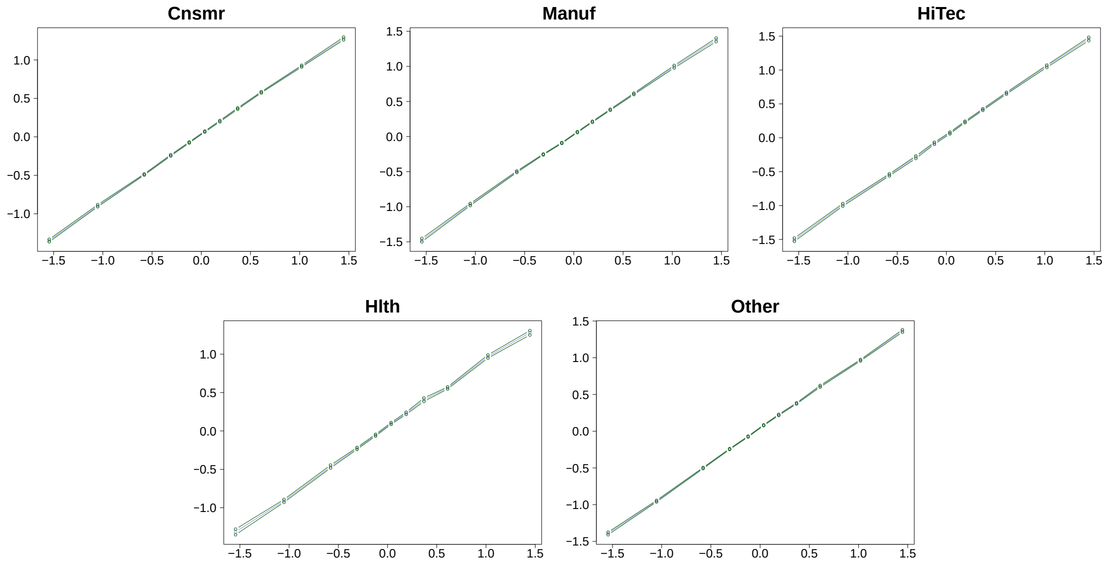
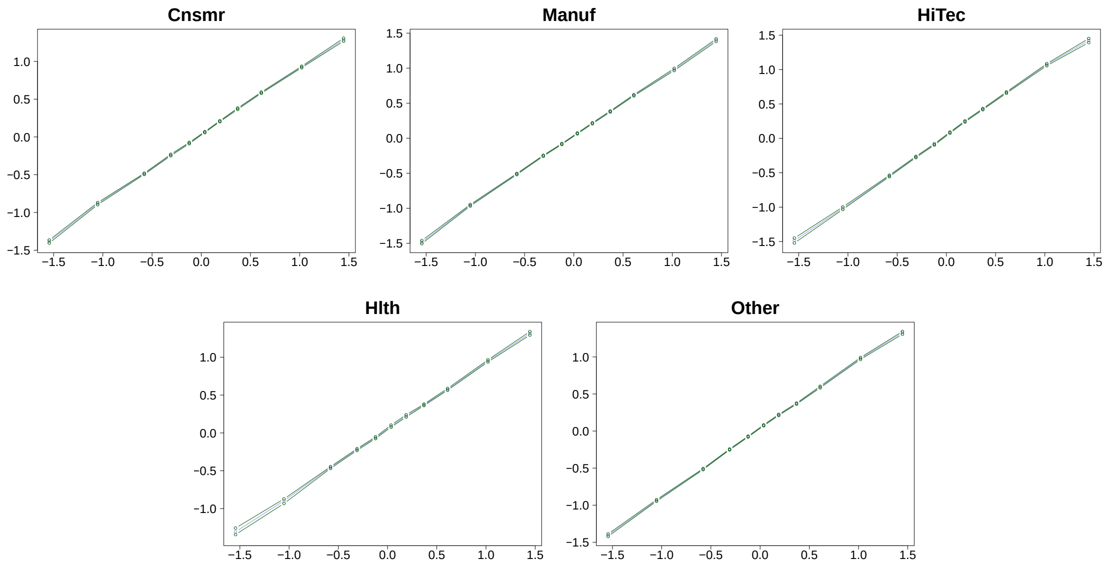


Figure 9: Asset pricing application. For Model 7, plot of partial dependence of f_0 (y-axis) against $(R_m - R_f)$ (x-axis), for the following sectors: consumer (Cnsmr), manufacturing (Manuf), technology (HiTec), health (Hlth) and others (Other)



Results show that for all models and for all sectors, the excess market return $(R_m - R_f)$ has a positive and linear relationship with f_0. Economically this is equivalent to stating that the expected rate of return of a certain asset will depend mainly on market performance.

This behavior induces and enforces an already existing investment strategy called buy-and-hold. According to the strategy, the best investment decision is to buy assets and then to keep them for the long run. Market returns, despite local spikes and plunges, in the long run present a positive trend. Therefore, the prediction of this analysis is that, since individual asset returns are linked by a linear and positive relation to market returns, buy-and-hold strategy applied to single assets will ensure a positive return to the investor in the long run.

5 Conclusions

CAPM is probably the most famous and widely used asset pricing model. One of its strengths is the simplicity of implementation and the robustness of its results. During the years it has been criticized for its simplicity and for its unrealistic assumptions of the world. Some studies have proposed the addition of other factors in the basic setting in order to capture a wider range of financial facts already existing in literature (for example Fama and French [2015]).

In order to give evidence in favor or against the relevance of these new factors, in this thesis I proposed an application of the Soft Bayesian Additive Regression Trees model by Linero [2018]. Regression trees have proven to be very effective statistical models and I chose this one given its capability to adapt to sparsity in predictor space and for its smoothness property.

My results give evidence against factors added to basic CAPM in two ways. According to this thesis analysis, $(R_m - R_f)$ factor of CAPM remains the most relevant factor among all those considered. In and out RMSE computations show that the basic model minimizes in average the misclassification error of predictions. In addition, as the number of factor increases, their explanatory power (measured as probability of inclusion in the model) lowers at a level that is slightly higher than the one of noise variables drawn from a uniform distribution.

In conclusion, according to my work, CAPM remains the best asset pricing model and the only one comparable in terms of results is the Fama&French five factor model that drops HML. Furthermore the relationship between returns of a certain asset and excess market return is simply positive and linear, characteristic that leads to straightforward and precise predictions about portfolio returns.

A On BART models

A.1 BART independence and symmetry

In their formulation of BART, Chipman et al. [2010] simplify the model considering only priors for which:

$$
\begin{aligned}
p((T_1, M_1), \ldots, (T_m, M_m), \sigma) &= \left[\prod_j p(T_j, M_j)\right] p(\sigma) \\
&= \left[\prod_j p(M_j|T_j)p(T_j)\right] p(\sigma)
\end{aligned}
$$

and

$$p(M_j|T_j) = \prod_j p(\mu_{ij}|T_j).$$

This setting involves using identical forms for all T_j and for all $p(\mu_{ij}|T_j)$. Under such priors, tree components (T_j, M_j) are mutually independent and every tree parameter μ_{ij} is independent.

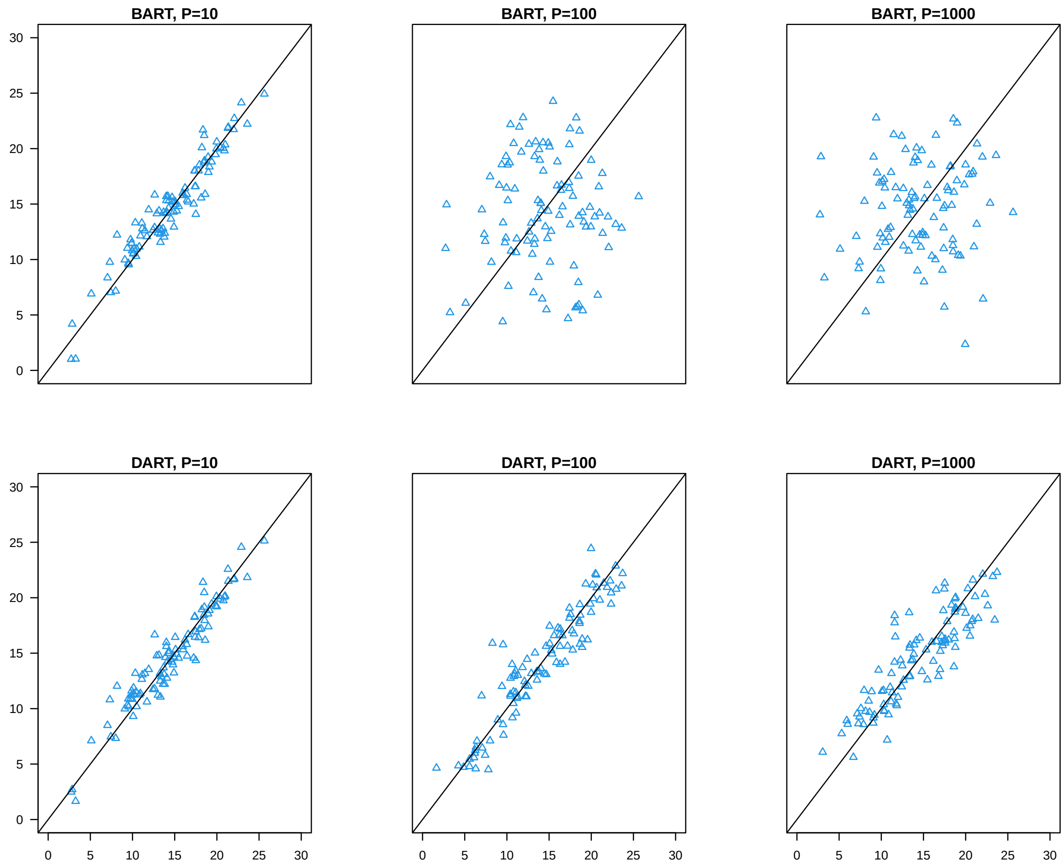
A.2 DART simulation results

In order to better understand implications of introducing noise in predictors I perform simulations of BART models at increasing levels of P = number of predictors considered. Simulations are performed following the method of Friedman [1991], namely estimating

$$f_0(x) = 10 \sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5$$

and introducing other predictors drawn from a uniform distribution. Simulations are run using the BART package available at `https://cran.r-project.org/web/packages/BART/index.html`.

Figure 10: True vs. estimated μ at increasing P

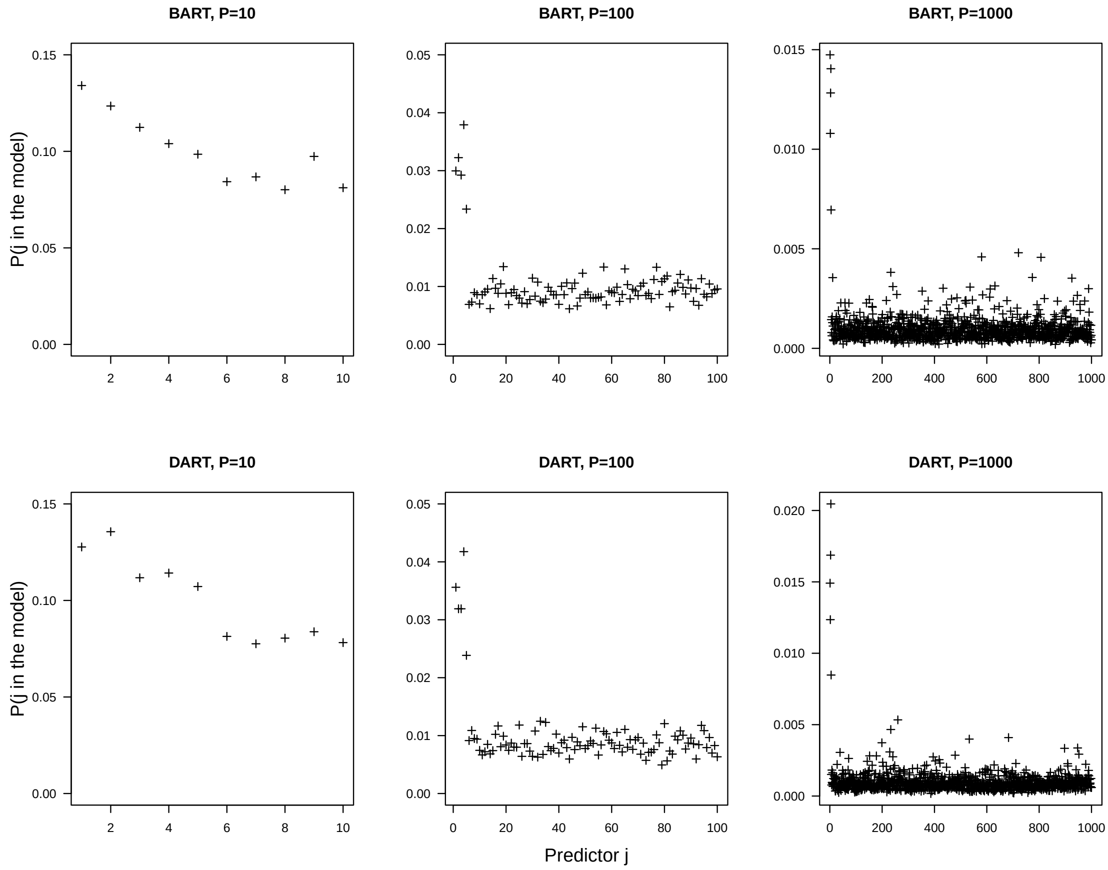


y-axis: estimated $\hat{\mu}$, x-axis: real μ

Results in Figure 10 show that BART model loses it's predictive power when too many irrelevant predictors are considered in the model, instead DART keeps it's performance independently of P.

We can see another hint of this behavior by looking at Figure 11. The plot shows the probability that predictor j is selected by the model and used for inference. Increasing P we can see that DART keeps the probability of including noise variables in the model lower than BART does.

Figure 11: Probability of inclusion of predictors in the model

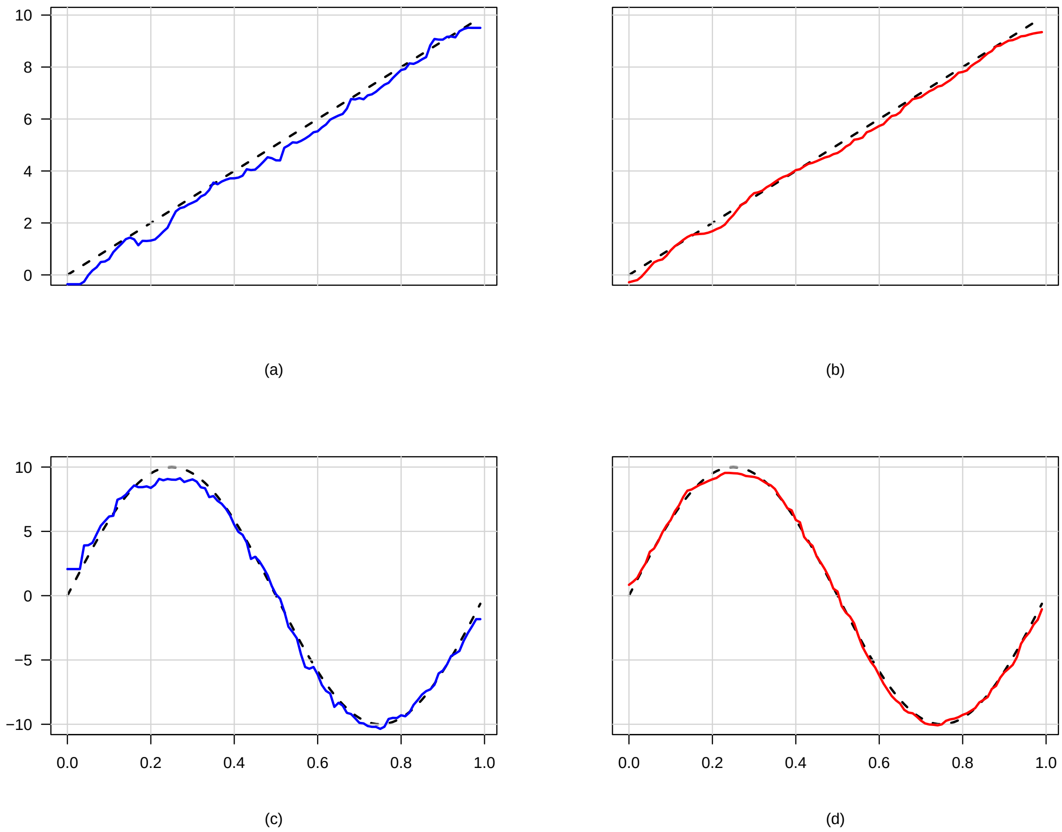


A.3 Smoothness simulation results

I conduct a simulation on posterior smoothness to compare BART and SBART with this setting: priors are kept at default; number of burn in simulations is 100; number of posterior samples is 1000; number of trees is 200. Results are shown in Figure 12. Simulations are run using BART package of previous section and SBART package available at `https://github.com/theodds/SoftBART`.

The graph shows the estimates of two different functions, namely $f(x) = 10x$ is panels (a) and (b) and $f(x) = 10sin(2\pi x)$ in panels (c) and (d). Results show that SBART produces a smoother fit.

Figure 12: BART vs. SBART smoothness simulation



(- - -) real $f_0(x)$ vs. (—) estimated $\hat{f}_0(x)$. Panels (a) and (c) are BART (blue), (b) and (d) are SBART (red)

A.4 The MH algorithm for drawing T_j

The algorithm of Chipman et al. [1998] for sampling trees, proposes a new one by applying one of the following modifications at each leaf node:

1. GROW: randomly pick a leaf node and split it in two child nodes, assigning to each a parameter μ_{ij} drawn from the prior;

2. PRUNE: randomly pick a parent of two terminal nodes and make it a leaf node, collapsing parameters below it;

3. CHANGE: randomly pick an internal node and reassign it's splitting rule according to the prior;

4. SWAP: randomly pick a parent-child pair and swap their splitting rule.

In the SBART formulation, authors consider only the first three rules: GROW, PRUNE and CHANGE. Transitions to new trees are performed computing the usual α probability of acceptance and weighting it by means of likelihood computations.

B On asset pricing

B.1 Data source

The data I'm using in the analysis section of this thesis is entirely taken from Fama&French database**. In particular I pick, under " U.S. Research Returns Data (Downloadable Files)", the "Fama/French 5 Factors (2x3) [Daily]" data for RF, Market Returns, SMB, HML, RMW and CMA. The dataset is composed by daily entries of these seven variables, ranging from July 1, 1963 to May 31, 2022, for a total of $7 \cdot 14.831 = 103.817$ observations. In "Details" there is an explanation of how the dataset in constructed, formulas used by the authors for factors computation and reference to stocks analyzed.
For stock returns I pick "5 industry portfolios [Daily]". This dataset is composed by daily returns of five portfolios composed by assets categorized by the authors through the four-digit SIC codes. The five categories are:

Cnsmr Consumer Durables, Nondurables, Wholesale, Retail, and Some Services (Laundries, Repair Shops);

**available at `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

Manuf Manufacturing, Energy, and Utilities;

HiTec Business Equipment, Telephone and Television Transmission;

Hlth Healthcare, Medical Equipment, and Drugs;

Other Other — Mines, Constr, BldMt, Trans, Hotels, Bus Serv, Entertainment, Finance.

Data range from July 1, 1926 to July 31, 2022. Further details can be found in "Details" section near dataset link of the same webpage.
For completeness, I use also "Momentum Factor (Mom) [Daily]", which is a factor introduced in Carhart [1997] and can be found in Fama&French webpage.
Since factor data include a narrower time span of observations compared to the other datasets, I drop all years of observations that are not complete or that come with missing data.

C SBART application details

I report in Table 15 RMSE of the simulations in section 4. RMSE measures the variance between prediction and real value and is computed as

$$RMSE = \left(\frac{\sum (\hat{y} - y)}{n} \right)^2$$

For every model and dataset, there are two values: the first is RMSE of training data; the second is the one of test data. Results show that Model 1 (CAPM) and Model 5 (Fama&French five factors without HML) generally perform better than the others. We can see that for Models 3, 4, 6 and 7, test (prediction) RMSE is always higher that the one of Models 1 and 5. Model 2 (Fama&French three factors) perform good just with Cnsmr and Other data.

Table 15: RMSE of models in section 4

Y	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Cnsmr	0.418	0.359	0.344	0.352	0.349	0.342	**0.338**
	0.761	0.764	0.910	0.805	0.774	0.828	0.836
Manuf	0.443	0.413	0.406	0.411	**0.399**	0.406	0.409
	0.777	0.874	0.952	0.936	**0.748**	0.923	1.064
HiTec	**0.511**	0.534	0.575	0.557	0.578	0.586	0.589
	0.735	0.960	1.057	0.958	1.013	0.983	1.041
Hlth	0.442	0.446	0.454	0.445	**0.427**	0.460	0.460
	0.733	0.914	1.027	0.952	0.842	0.964	0.957
Other	0.459	0.450	0.449	0.455	**0.428**	0.450	0.450
	0.752	**0.734**	0.940	0.929	0.784	0.977	0.937

These results support the the idea of Fama and French [2015] that HML is redundant, in fact, models that include it have a substantially higher prediction error. The best model for prediction seems to be the simple CAPM, which has the lower prediction RMSE in three out of five datasets and still performs good in the remaining two.

I report here the graphs of posterior probability of inclusion, one for each of the 35 regressions in Section 3 and grouped by model (7 figures, each composed of 5 graphs). The red vertical line present in the graphs is used to distinguish factors (to the left) and noise variables (to the right). The factors are ordered as in their formulations in Equations 9 to 15. Results are presented in Figures 13 to 19.

I report also plots of y(x-axis) vs. $\hat{y}$(y-axis) of a part of the observations in Figures 20 to 26. We can see that generally HiTec and Hlth datasets present more skewed plots.

Figure 13: Asset pricing application. For Model 1, plot of posterior probability of inclusion of factors, for the following sectors: consumer (Cnsmr), manufacturing (Manuf), technology (HiTec), health (Hlth) and others (Other)

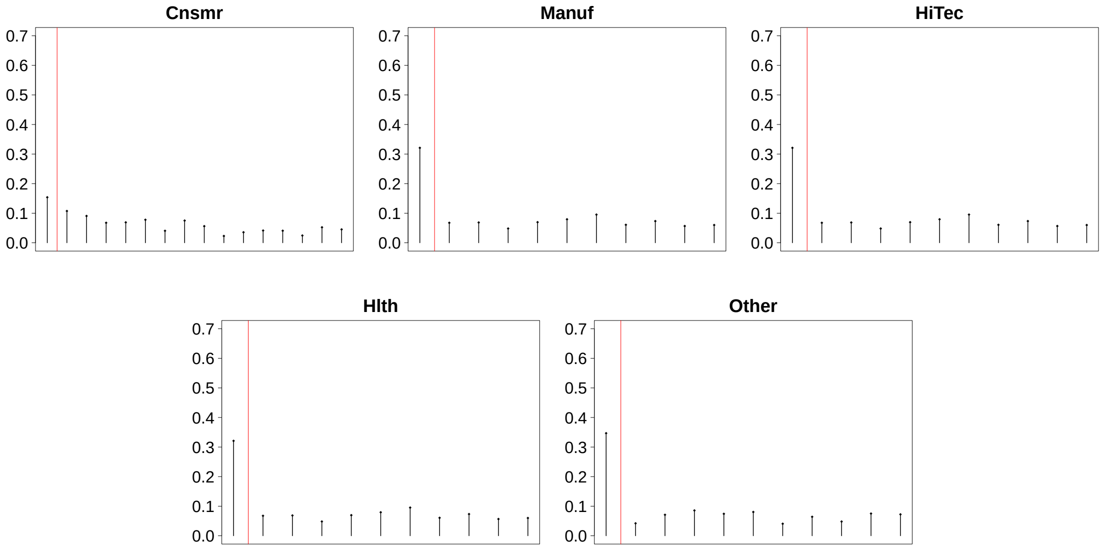


Figure 14: Asset pricing application. For Model 2, plot of posterior probability of inclusion of factors, for the following sectors: consumer (Cnsmr), manufacturing (Manuf), technology (HiTec), health (Hlth) and others (Other)

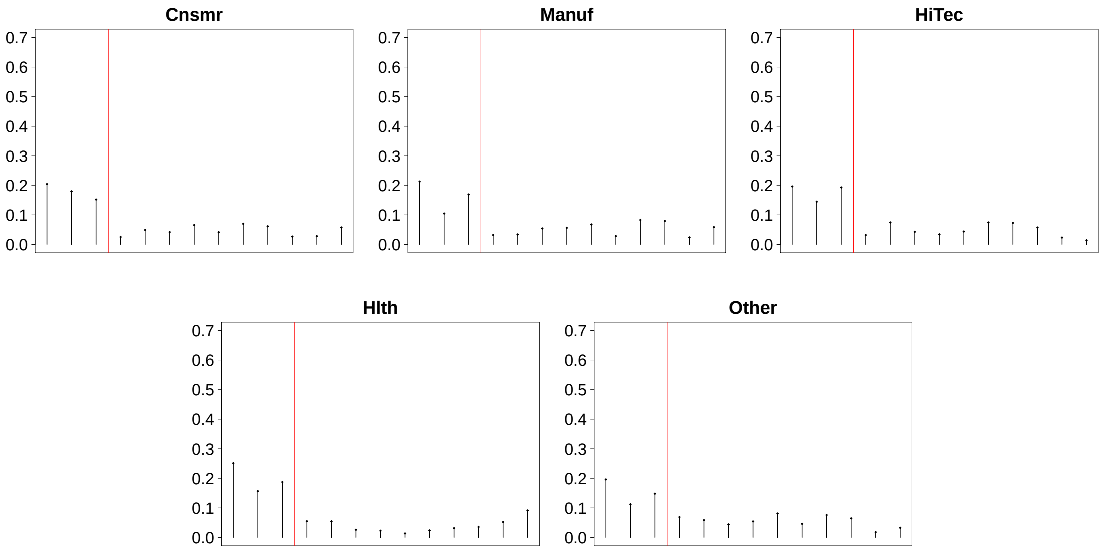


Figure 15: Asset pricing application. For Model 3, plot of posterior probability of inclusion of factors, for the following sectors: consumer (Cnsmr), manufacturing (Manuf), technology (HiTec), health (Hlth) and others (Other)

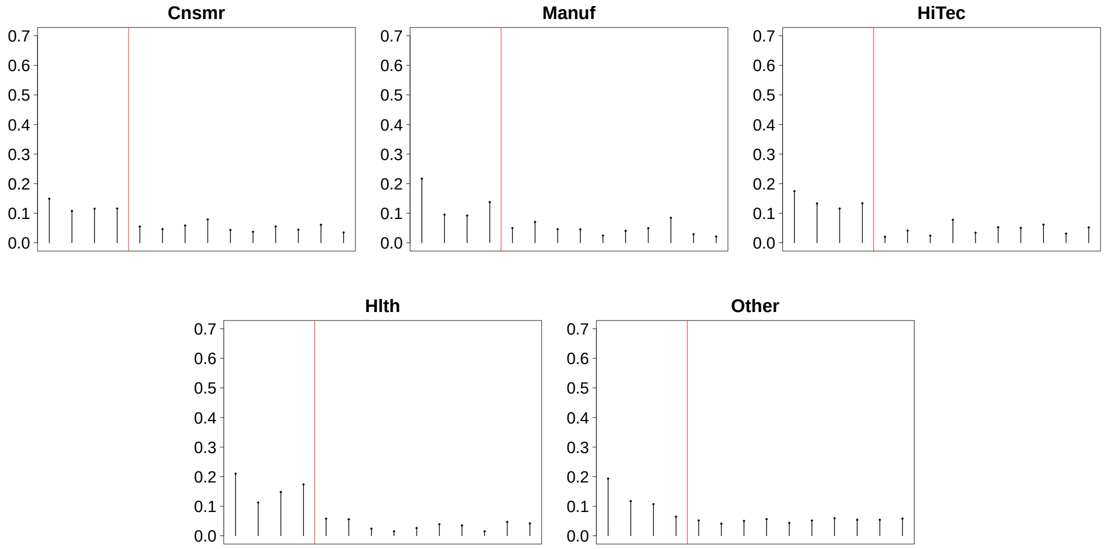


Figure 16: Asset pricing application. For Model 4, plot of posterior probability of inclusion of factors, for the following sectors: consumer (Cnsmr), manufacturing (Manuf), technology (HiTec), health (Hlth) and others (Other)

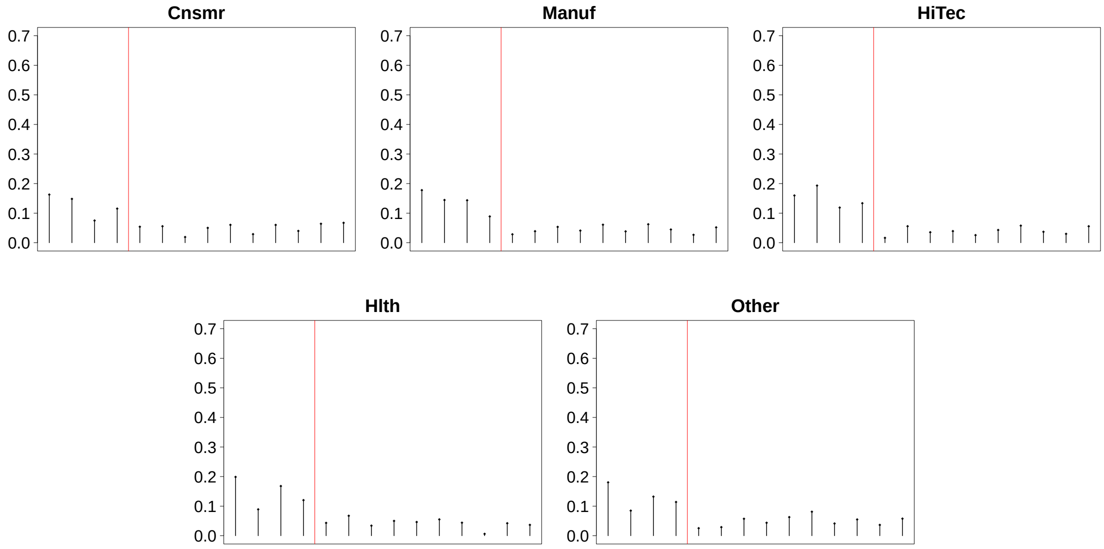


Figure 17: Asset pricing application. For Model 5, plot of posterior probability of inclusion of factors, for the following sectors: consumer (Cnsmr), manufacturing (Manuf), technology (HiTec), health (Hlth) and others (Other)

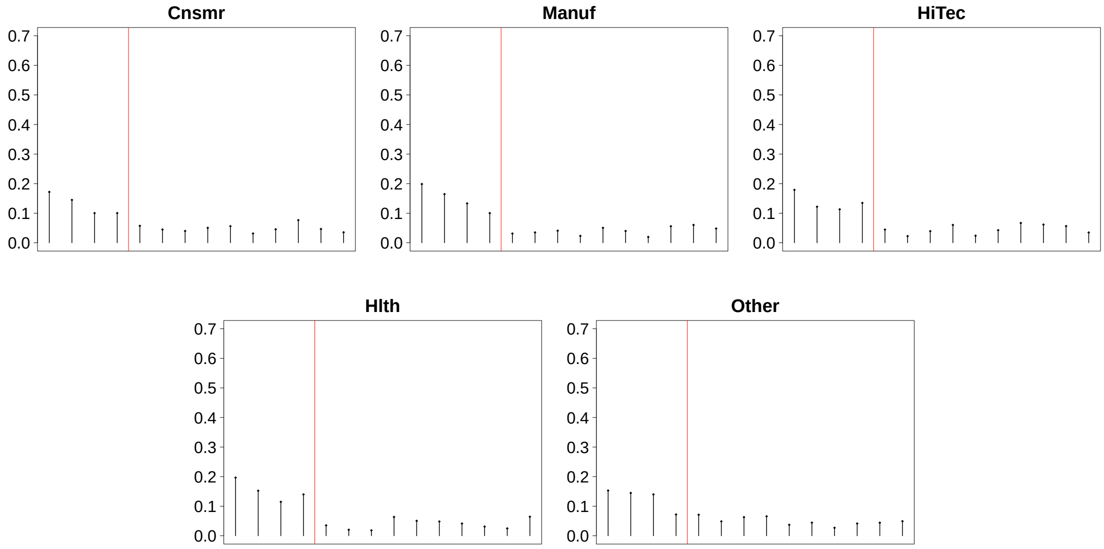


Figure 18: Asset pricing application. For Model 6, plot of posterior probability of inclusion of factors, for the following sectors: consumer (Cnsmr), manufacturing (Manuf), technology (HiTec), health (Hlth) and others (Other)

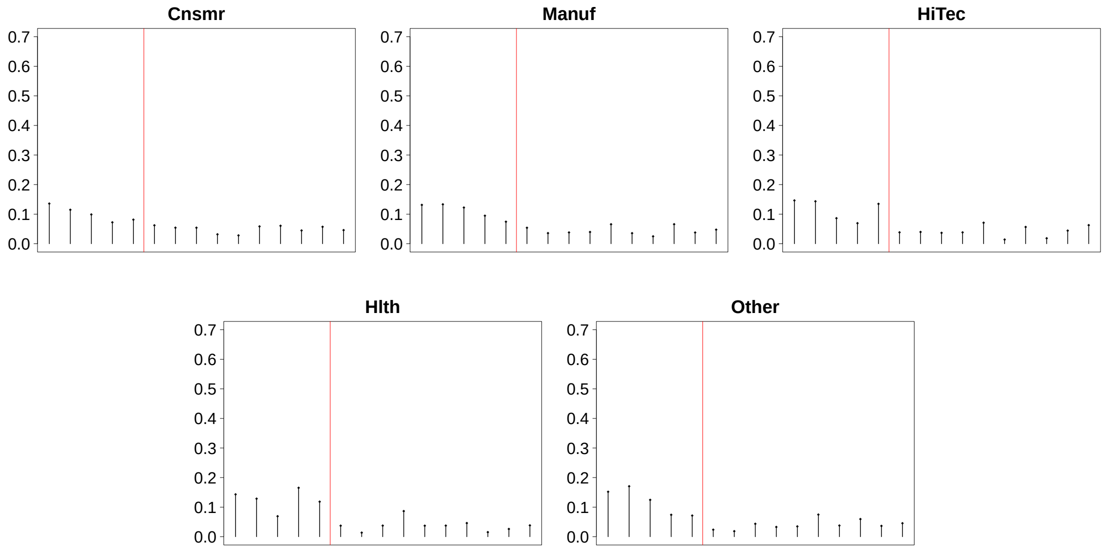


Figure 19: Asset pricing application. For Model 7, plot of posterior probability of inclusion of factors, for the following sectors: consumer (Cnsmr), manufacturing (Manuf), technology (HiTec), health (Hlth) and others (Other)

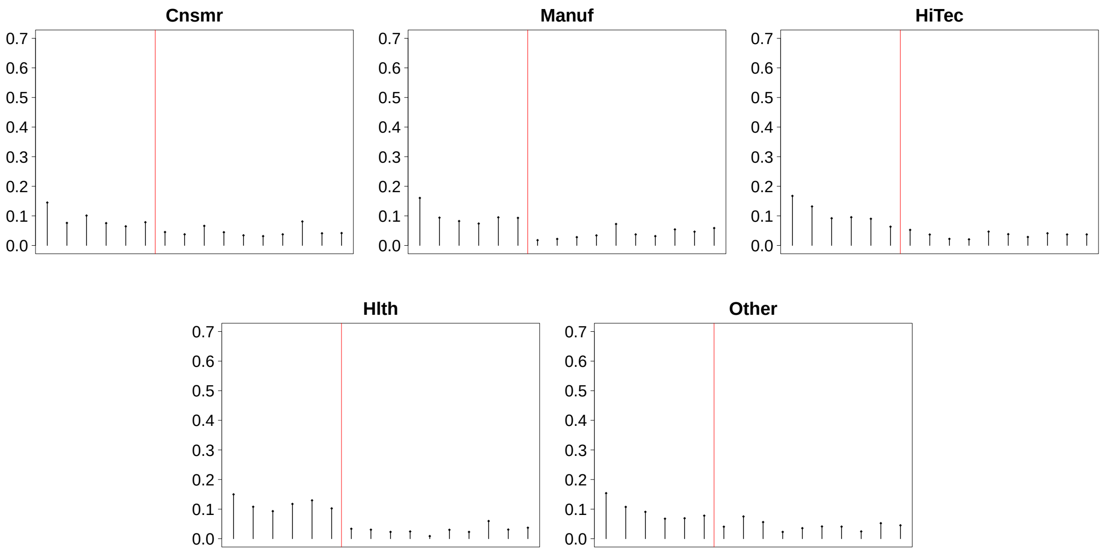


Figure 20: Asset pricing application. True y_i (horizontal axis) vs. estimated $\hat{y}_i$ (vertical axis), Model 1

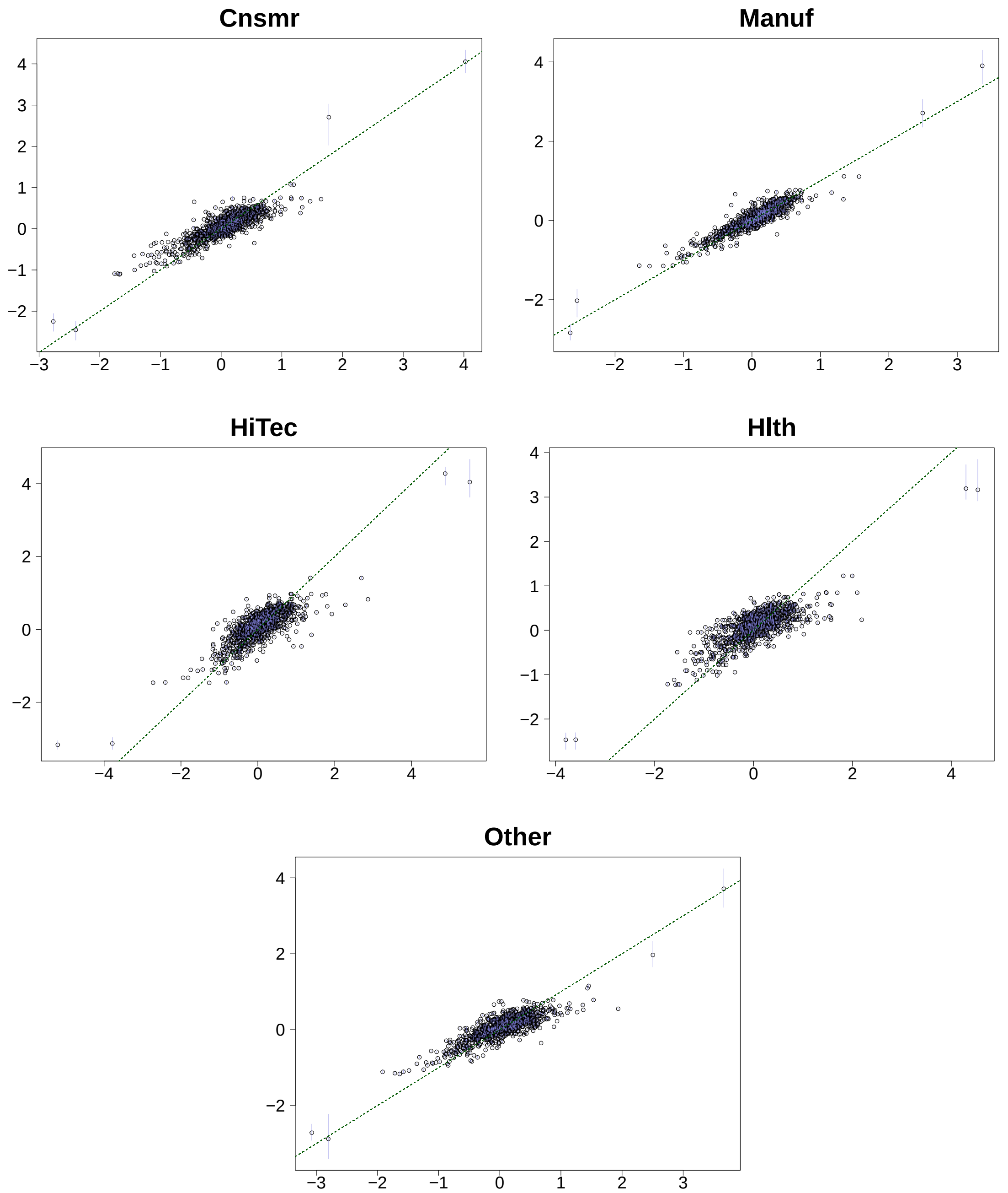


Figure 21: Asset pricing application. True y_i (horizontal axis) vs. estimated $\hat{y}_i$ (vertical axis), Model 2

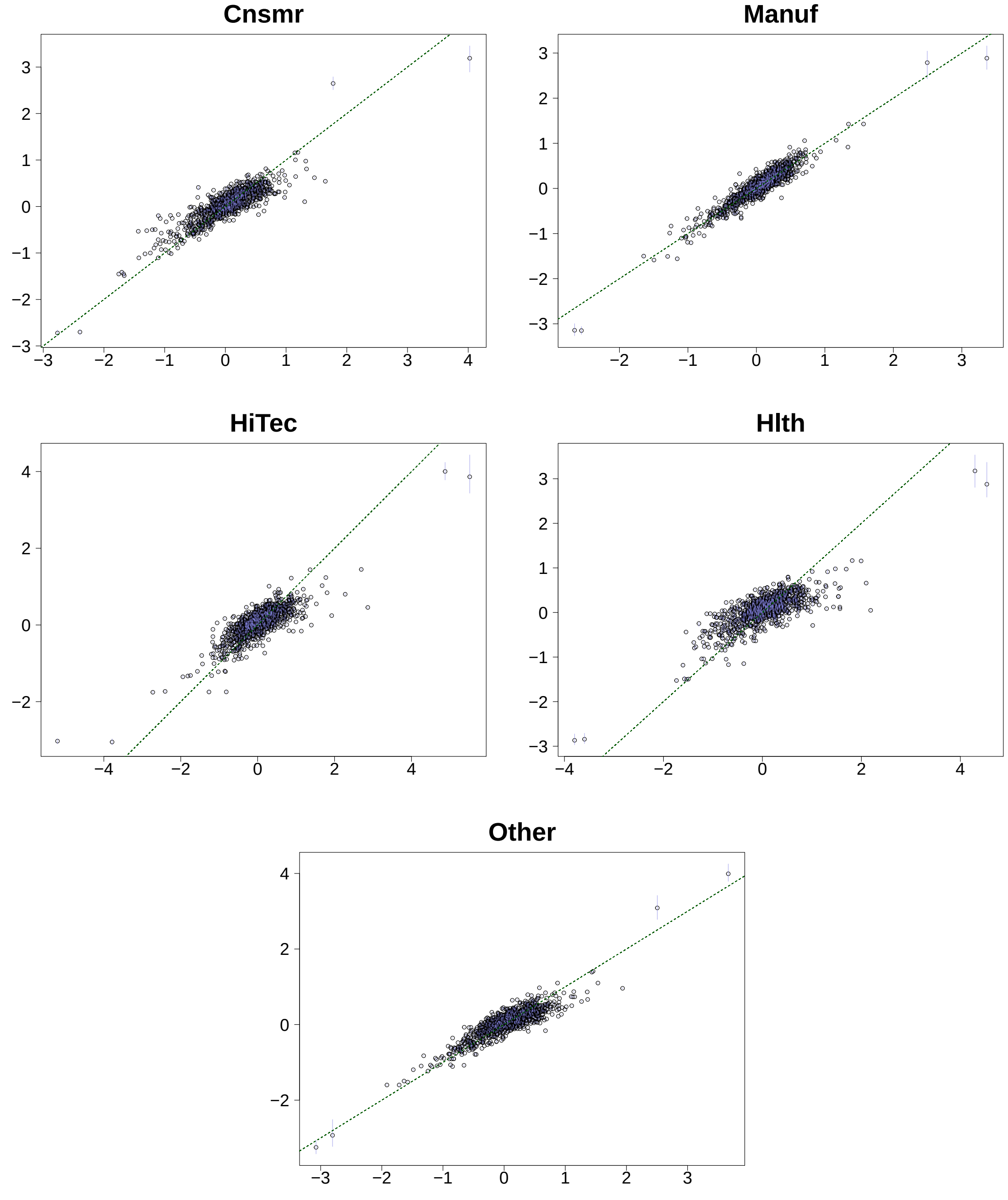


Figure 22: Asset pricing application. True y_i (horizontal axis) vs. estimated $\hat{y}_i$ (vertical axis), Model 3

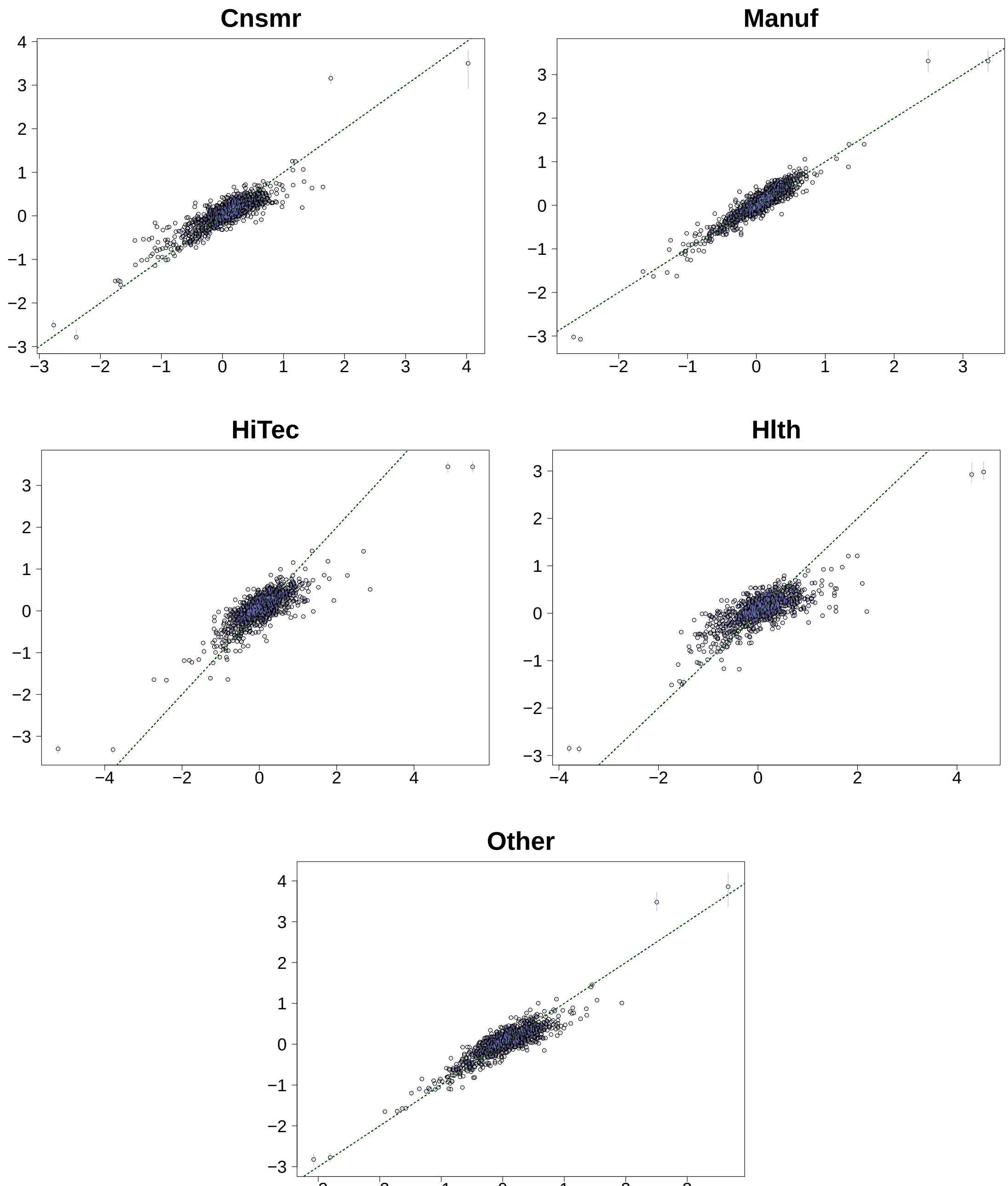


Figure 23: Asset pricing application. True y_i (horizontal axis) vs. estimated $\hat{y}_i$ (vertical axis), Model 4

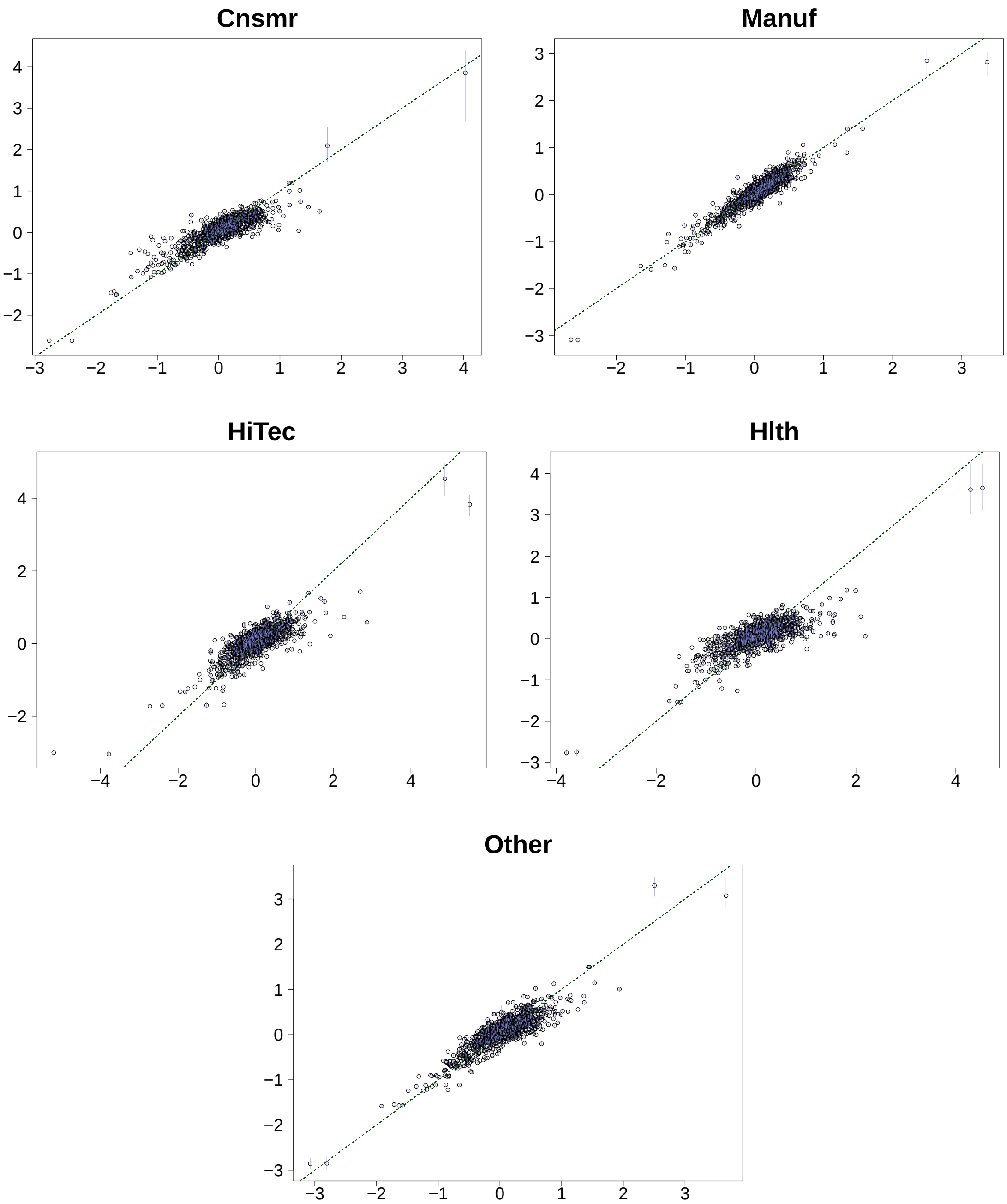


Figure 24: Asset pricing application. True y_i (horizontal axis) vs. estimated $\hat{y}_i$ (vertical axis), Model 5

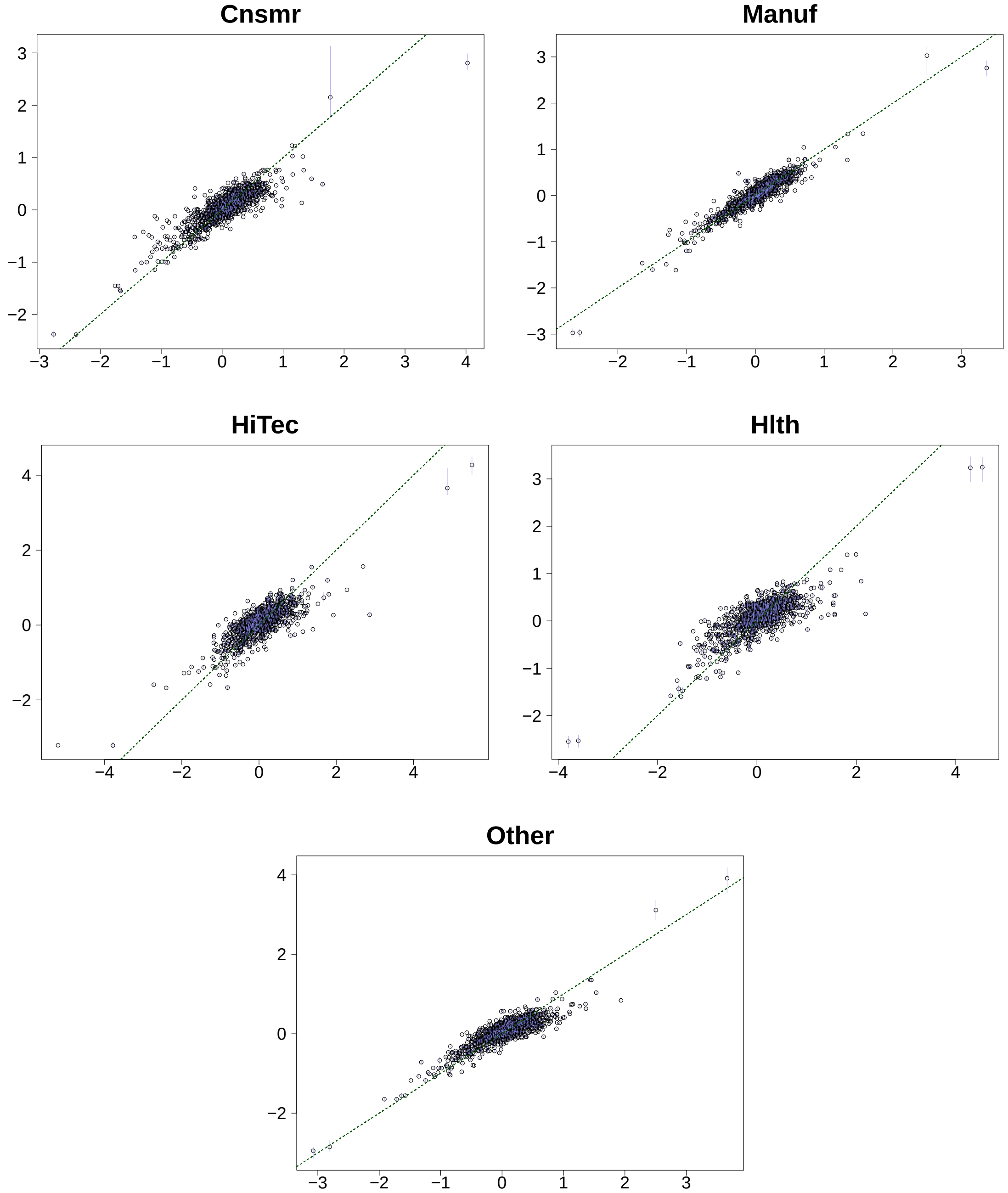


Figure 25: Asset pricing application. True y_i (horizontal axis) vs. estimated $\hat{y}_i$ (vertical axis), Model 6

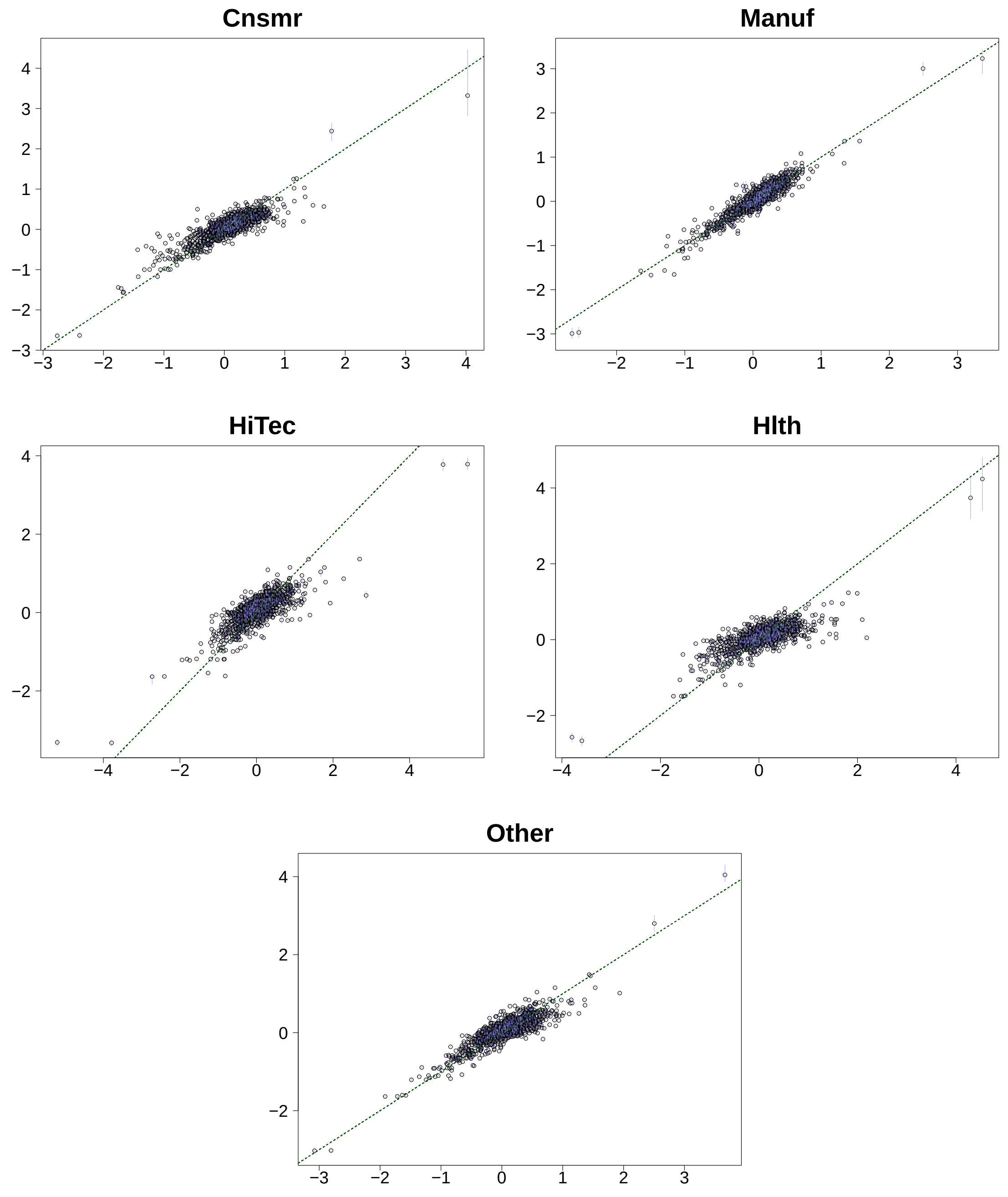
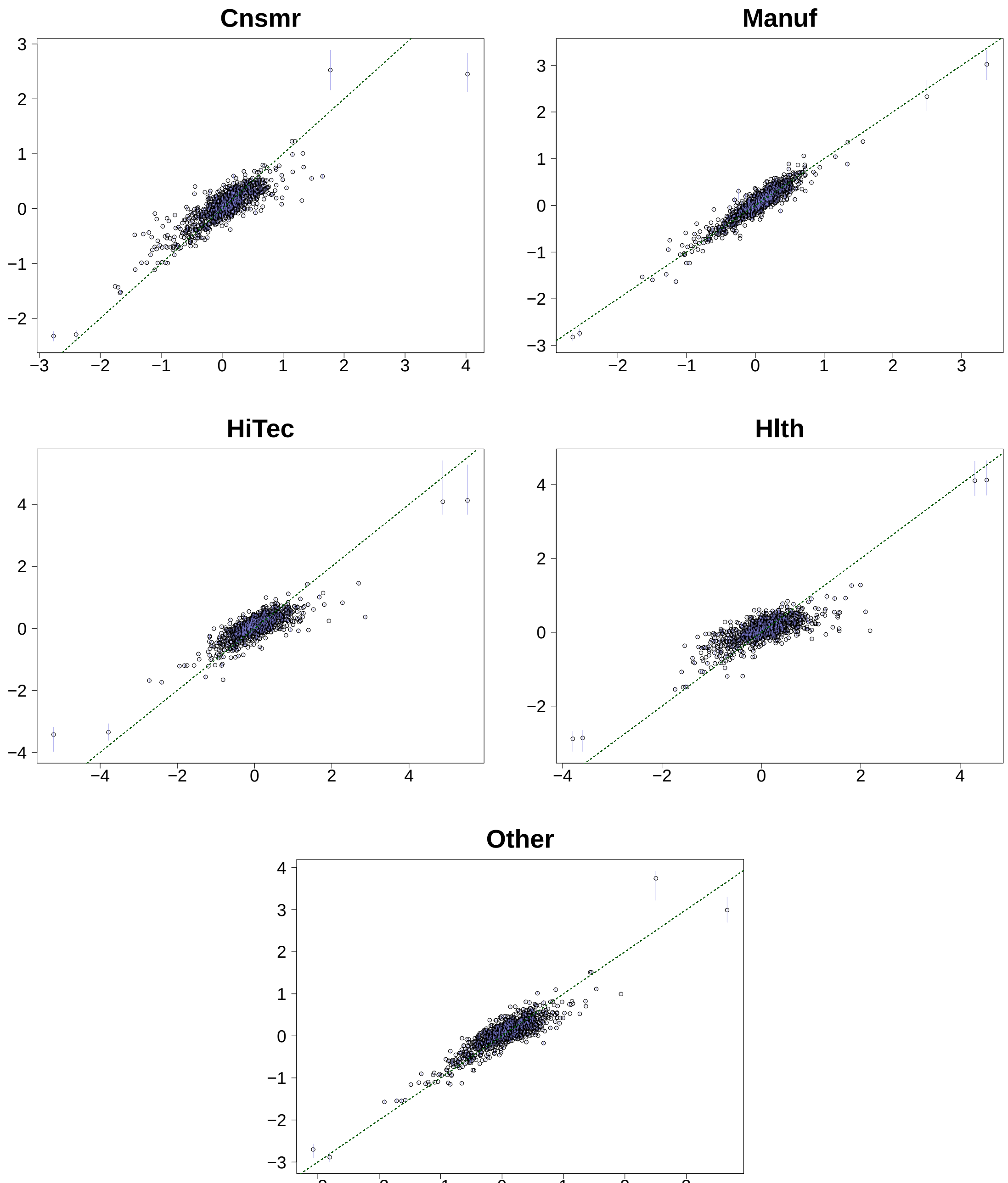


Figure 26: Asset pricing application. True y_i (horizontal axis) vs. estimated $\hat{y}_i$ (vertical axis), Model 7



C.1 Partial dependence simulations using Friedman [1991] formula

Recall Friedman [1991] formula:

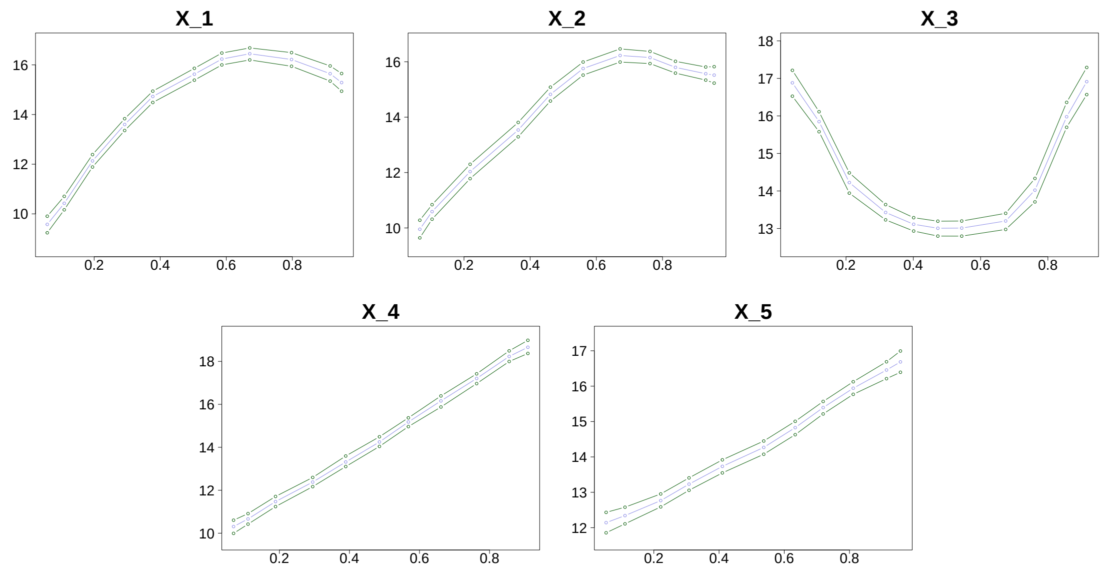
$$f_0(x) = 10 \sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5. \qquad (17)$$

Given this formula, we expect that the partial dependence of estimated f_0 is: sinusoidal against x_1 and x_2; concave against x_3; linear against x_4 and x_5.

To perform this simulation, the default setting of "pdsoftbart" is used: n. of trees = 20; n. of burn-in iterations = 2500; n. of save iterations = 2500.

Results presented in Figure 27 confirm our expectations: partial dependence shapes respect the formulation that we have in Equation 17.

Figure 27: Partial dependence of f_0 against predictive variables x_1, x_2, x_3, x_4 and x_5



References

Athey, S. and Imbens, G. W. (2019). Machine Learning methods that Economists should know about. *Annual Review of Economics*, 11(1):685–725.

Barillas, F. and Shanken, J. (2018). Comparing asset pricing models. *The Journal of Finance*, 73(2):715–754.

Basak, P., Linero, A., Sinha, D., and Lipsitz, S. (2021). Semiparametric analysis of clustered interval-censored survival data using soft bayesian additive regression trees (sbart). *Biometrics*.

Black, F. (1972). Capital market equilibrium with restricted borrowing. *The Journal of Business*, 45(3):444–455.

Black, F., Jensen, M. C., Scholes, M., et al. (1972). The capital asset pricing model: Some empirical tests. *Praeger New York*.

Breeden, D. T. (2005). An intertemporal asset pricing model with stochastic consumption and investment opportunities. In *Theory of valuation*, pages 53–96. World Scientific.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Campbell, J. Y. (2000). Asset pricing at the millennium. *The Journal of Finance*, 55(4):1515–1567.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82.

Casella, G. and George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46(3):167–174.

Chen, C. W., Gerlach, R. H., and Lin, A. M. (2011). Multi-regime nonlinear capital asset pricing models. *Quantitative Finance*, 11(9):1421–1438.

Chipman, H. A., George, E. I., and McCulloch, R. E. (1998). Bayesian cart model search. *Journal of the American Statistical Association*, 93(443):935–948.

Chipman, H. A., George, E. I., and McCulloch, R. E. (2010). Bart: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1):266–298.

Clarke, R., De Silva, H., and Thorley, S. (2011). Minimum-variance portfolio composition. *The Journal of Portfolio Management*, 37(2):31–45.

de Oca, M. M., Hill, J., Aber, L., Dolan, C. T., and Gjicali, K. (2022). The impact of attending a remedial support program on Syrian children's reading skills: Using BART for Causal Inference. *arXiv preprint arXiv:2208.13906*.

Denison, D. G., Mallick, B. K., and Smith, A. F. (1998). A Bayesian CART algorithm. *Biometrika*, 85(2):363–377.

Dittmar, R. F. (2002). Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *The Journal of Finance*, 57(1):369–403.

Du, J. and Linero, A. R. (2019). Interaction detection with Bayesian decision tree ensembles. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 108–117. PMLR.

Elbannan, M. A. (2015). The capital asset pricing model: An overview of the theory. *International Journal of Economics and Finance*, 7(1):216–228.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (2004). The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives*, 18(3):25–46.

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

Faraji, Z. (2022). Investigate the causal effect of diversification strategy on risk-adjusted performance using Bayesian additive regression trees. *International Journal of Information and Decision Sciences*, 14(2):97–114.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1):1–67.

Frondel, M., Kaeding, M., and Sommer, S. (2022). Market premia for renewables in Germany: The effect on electricity prices. *Energy Economics*, 109:105874.

Hastie, T. and Tibshirani, R. (2000). Bayesian backfitting (with comments and a rejoinder by the authors. *Statistical Science*, 15(3):196–223.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Oxford University Press*.

Henderson, H. and Follett, L. (2022). Targeting social safety net programs on human capabilities. *World Development*, 151:105741.

Hill, J., Linero, A., and Murray, J. (2020). Bayesian additive regression trees: A review and look forward. *Annual Review of Statistics and Its Application*, 7(1).

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., and Lauer, M. S. (2008). Random survival forests. *The Annals of Applied Statistics*, 2(3):841–860.

Jin, M., Ren, Z., and Shi, X. (2022). Spatiotemporal characteristics and drivers of Chinese urban total noise pollution from 2007 to 2019. *Environmental Science and Pollution Research*, pages 1–15.

Josefsson, M., Daniels, M. J., and Pudas, S. (2020). A Bayesian semiparametric approach for inference on the population partly conditional mean from longitudinal data with dropout. *arXiv preprint arXiv:2011.12345*.

Lakshminarayanan, B., Roy, D. M., and Teh, Y. W. (2014). Mondrian forests: Efficient online random forests. *Advances in Neural Information Processing Systems*, 27.

Li, Y., Linero, A. R., and Murray, J. (2022a). Adaptive conditional distribution estimation with Bayesian decision tree ensembles. *Journal of the American Statistical Association*, pages 1–14.

Li, Z., Liu, S., Conaty, W., Zhu, Q.-H., Moncuquet, P., Stiller, W., and Wilson, I. (2022b). Genomic prediction of cotton fibre quality and yield traits using Bayesian regression methods. *Heredity*, pages 1–10.

Linero, A. R. (2017). A review of tree-based Bayesian methods. *Communications for Statistical Applications and Methods*, 24(6):543–559.

Linero, A. R. (2018). Bayesian regression trees for high-dimensional prediction and variable selection. *Journal of the American Statistical Association*, 113(522):626–636.

Linero, A. R. (2022). Generalized Bayesian Additive Regression Trees models: Beyond conditional conjugacy. *arXiv preprint arXiv:2202.09924*.

Linero, A. R., Basak, P., Li, Y., and Sinha, D. (2021). Bayesian survival tree ensembles with submodel shrinkage. *Bayesian Analysis*, 1(1):1–24.

Linero, A. R. and Yang, Y. (2018). Bayesian regression tree ensembles that adapt to smoothness and sparsity. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(5):1087–1110.

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *The Journal of Finance*, 20(4):587–615.

Loh, W.-Y. (2011). Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 1(1):14–23.

Luo, C. and Daniels, M. J. (2021). Variable selection using Bayesian Additive Regression Trees. *arXiv preprint arXiv:2112.13998*.

Markowitz, H. (1959). Portfolio selection: Efficient diversification of investments–John Wiley & Sons, Inc.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, pages 867–887.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

Modigliani, F. and Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *The American Economic Review*, 48(3):261–297.

Moret, B. M. (1982). Decision trees and diagrams. *ACM Computing Surveys (CSUR)*, 14(4):593–623.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*, 34(4):768–783.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, 108(1):1–28.

Obringer, R., Nateghi, R., Maia-Silva, D., Mukherjee, S., Cr, V., McRoberts, D. B., and Kumar, R. (2022). Implications of increasing household air conditioning use across the United States under a warming climate. *Earth's Future*, 10(1):e2021EF002434.

Orlandi, V., Murray, J., Linero, A., and Volfovsky, A. (2021). Density regression with Bayesian Additive Regression Trees. *arXiv preprint arXiv:2112.12259*.

Pratola, M. T., Chipman, H. A., George, E. I., and McCulloch, R. E. (2020). Heteroscedastic BART via multiplicative regression trees. *Journal of Computational and Graphical Statistics*, 29(2):405–417.

Ran, H. and Bai, Y. (2021a). Distributed Soft Bayesian Additive Regression Trees. *arXiv preprint arXiv:2108.11600*.

Ran, H. and Bai, Y. (2021b). On Soft Bayesian Additive Regression Trees and asynchronous longitudinal regression analysis. *arXiv preprint arXiv:2108.11603*.

Ritter, C. and Tanner, M. A. (1992). Facilitating the Gibbs sampler: the Gibbs stopper and the griddy-Gibbs sampler. *Journal of the American Statistical Association*, 87(419):861–868.

Ročková, V. and van der Pas, S. (2020). Posterior concentration for Bayesian regression trees and forests. *The Annals of Statistics*, 48(4):2108–2131.

Roll, R. and Ross, S. A. (1980). An empirical investigation of the arbitrage pricing theory. *The Journal of Finance*, 35(5):1073–1103.

Ross, S. A. (2013). The arbitrage theory of capital asset pricing. In *Handbook of the fundamentals of financial decision making: Part I*, pages 11–30. World Scientific.

Sarti, D. A., Prado, E. B., Inglis, A. N., Dos Santos, A. A., Hurley, C. B., Moral, R. A., and Parnell, A. C. (2021). Bayesian Additive Regression Trees for genotype by environment interaction models. *bioRxiv*.

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

Sparapani, R. A., Logan, B. R., McCulloch, R. E., and Laud, P. W. (2016). Nonparametric survival analysis using Bayesian additive regression trees (bart). *Statistics in Medicine*, 35(16):2741–2753.

Starling, J. E., Murray, J. S., Carvalho, C. M., Bukowski, R. K., and Scott, J. G. (2020). BART with targeted smoothing: An analysis of patient-specific stillbirth risk. *The Annals of Applied Statistics*, 14(1):28–50.

Tauchen, G. and Hussey, R. (1991). Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models. *Econometrica*, pages 371–396.

Titman, S., Wei, K. J., and Xie, F. (2004). Capital investments and stock returns. *Journal of Financial and Quantitative Analysis*, 39(4):677–700.

Treynor, J. L. (1962). Toward a theory of market value of risky assets'. *Available at SSRN 628187*.

Tsionas, M. G. (2022). Efficiency estimation using probabilistic regression trees with an application to Chilean manufacturing industries. *International Journal of Production Economics*, page 108492.

Wong, J. C. Y., Blankenship, B., Harish, S., and Urpelainen, J. (2022). Increasing microsolar technology adoption: Efficacy of vouchers, cash transfers, and microfinance schemes. *Energy Economics*, 110:105952.

Woody, S., Carvalho, C. M., Hahn, P. R., and Murray, J. S. (2020). Estimating heterogeneous effects of continuous exposures using

Bayesian tree ensembles: Revisiting the impact of abortion rates on crime. *arXiv preprint arXiv:2007.09845*.

Zinn, S. and Gnambs, T. (2022). Analyzing nonresponse in longitudinal surveys using Bayesian Additive Regression Trees: A nonparametric event history analysis. *Social Science Computer Review*, 40(3):678–699.