Tail forecasting with multivariate Bayesian additive regression trees

Todd E. Clark[1] Florian Huber[2] Gary Koop[3]
Massimiliano Marcellino[4, 5, 6] Michael Pfarrhofer[2]

[1]Federal Reserve Bank of Cleveland [2]University of Salzburg [3]University of Strathclyde [4]Bocconi University [5]IGIER [6]CEPR

Florian Huber
Department of Economics
florian.huber@sbg.ac.at
sites.google.com/view/fhuber7

Introduction: Tail Risks to Economic Activity

- Policy makers and practitioners have a strong interest in modeling the tails of predictive densities
- Academic interest in modeling the tails of predictive densities:
  - Vulnerable growth: Adrian, Boyarchenko and Giannone (2019, AER)
  - Nowcasting tail risks to economic activity with many indicators: Carriero, Clark and Marcellino (2020)
  - Capturing macroeconomic tail risks with Bayesian Vector Autoregressions: Carriero, Clark and Marcellino (2020)
  - And many more
- Statistically interesting:
  - Regressions/VARs heavily in use but want models which allow for predictive densities which may be very non-Gaussian (fat-tails, skewness, multi-modality)
  - By definition, few observations in the tails which calls for regularization

Introduction: Non-parametric Modelling of Tail Risk

- Nowcasting in a pandemic using non-parametric mixed frequency VARs: Huber, Koop, Onorante, Pfarrhofer and Schreiner (JoE, in press)
  - Non-parametric methods nowcast well in extreme times
  - Quickly adjust to strong outliers far out of the range of the data
  - Predictive densities very non-Gaussian (fat tails, skewness, multi-modality)
- Non-parametric methods we use:
  - **B**ayesian **A**dditive **R**egression **T**rees (BART, see Chipman, George and McCulloch, 2010, AoAS),
  - Inference in Bayesian Additive Vector Autoregressive Tree models: Huber and Rossini (AoAS, in press)
- Could such methods be useful for forecasting tail risk?

What We Do

Methodological

- Develop various BART-based semi- and non-parametric VARs for tail risk forecasting
- BART treatments of both conditional means (VAR coeffs) and conditional variances (VAR error variances, labeled HeteroBART)
- Develop necessary MCMC methods for Bayesian inference

Empirical

- Real-time tail risk forecasting exercise with 23-dimensional VARs
- Compare various BART-based approaches to benchmark Bayesian VAR with SV
- We find BART methods tend to forecast tail risk better than BVAR-SV

Econometric framework

A general non-parametric multivariate regression

$$\boldsymbol{y}_t = F(\boldsymbol{x}_t) + \eta_t, \quad \eta_t = G(\boldsymbol{z}_t) + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(\boldsymbol{0}_M, \Sigma_t).$$

- $\{\boldsymbol{y}_t\}_{t=1}^T$ is M-dimensional with i^{th} element y_{it}
- $\boldsymbol{x}_t = (\boldsymbol{y}'_{t-1}, \ldots, \boldsymbol{y}'_{t-p})'$ is $K(= Mp)$-dimensional
- $F(\boldsymbol{x}_t) = (f_1(\boldsymbol{x}_t), \ldots, f_M(\boldsymbol{x}_t))'$ and $G(\boldsymbol{z}_t) = (g_1(\boldsymbol{z}_t), \ldots, g_M(\boldsymbol{z}_t))'$
- f_j and g_j are equation-specific (possibly) non-linear functions
- $\boldsymbol{z}_t$ to be defined later
- Σ_t is a $M \times M$ dimensional variance-covariance matrix

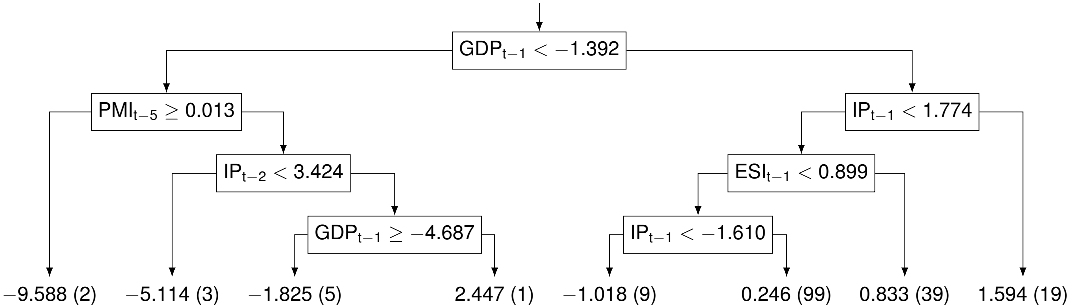
Approximating F and G with BART

BART approximation of $f_j(\boldsymbol{x}_t)$ and $g_j(\boldsymbol{z}_t)$

$$f_j(\boldsymbol{x}_t) \approx \sum_{s=1}^{S} h^f_{js}(\boldsymbol{x}_t | \mathcal{T}^f_{js}, \boldsymbol{\mu}^f_{js}), \quad g_j(\boldsymbol{z}_t) \approx \sum_{s=1}^{S} h^g_{js}(\boldsymbol{z}_t | \mathcal{T}^g_{js}, \boldsymbol{\mu}^g_{js})$$

- h^i_{js} is a tree function which depends on
  - tree structures $\mathcal{T}^i_{js}$
  - tree-specific terminal nodes $\boldsymbol{\mu}^i_{js}$
  - Dimension of $\boldsymbol{\mu}^i_{js}$ is denoted by b^i_{js} which depends on the complexity of the tree
- S denotes the total number of trees used.
- j denotes equations in the VAR
- Illustrate using a single tree for a VAR with 6 variables (do not worry about details of empirical application)

Example of a Regression Tree

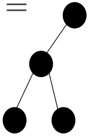


Intuition of BART

- The tree on the previous slide was quite complex $\rightarrow$ what about overfitting issues?
- BART prunes the trees to make them simpler
- But instead of using a single tree, BART uses S (which is a big number) of trees

Intuition

$$E(\boldsymbol{y}|\boldsymbol{x}) =$$

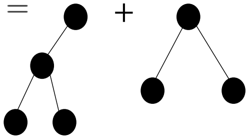


Intuition of BART

- The tree on the previous slide was quite complex $\rightarrow$ what about overfitting issues?
- BART prunes the trees to make them simpler
- But instead of using a single tree, BART uses S (which is a big number) of trees

Intuition

$E(\boldsymbol{y}|\boldsymbol{x}) =$

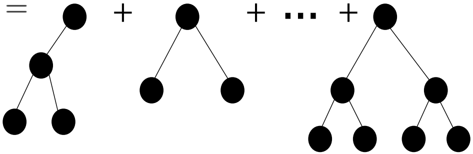


Intuition of BART

- The tree on the previous slide was quite complex $\rightarrow$ what about overfitting issues?
- BART prunes the trees to make them simpler
- But instead of using a single tree, BART uses S (which is a big number) of trees

Intuition

$$E(\boldsymbol{y}|\boldsymbol{x}) =$$



Models for the Conditional Mean

A general non-parametric multivariate regression

$$\mathbf{y}_t = F(\mathbf{x}_t) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t = G(\mathbf{z}_t) + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}_M, \boldsymbol{\Sigma}_t).$$

Different choices for $\mathbf{z}_t$, F and G give wide range of flexible models:

Models for the Conditional Mean

A general non-parametric multivariate regression

$$\boldsymbol{y}_t = F(\boldsymbol{x}_t) + \eta_t, \quad \eta_t = G(\boldsymbol{z}_t) + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(\mathbf{0}_M, \boldsymbol{\Sigma}_t).$$

Different choices for $\boldsymbol{z}_t$, F and G give wide range of flexible models:

BVAR

F is linear and G omitted

Models for the Conditional Mean

A general non-parametric multivariate regression

$$\boldsymbol{y}_t = F(\boldsymbol{x}_t) + \eta_t, \quad \eta_t = G(\boldsymbol{z}_t) + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(\mathbf{0}_M, \boldsymbol{\Sigma}_t).$$

Different choices for $\boldsymbol{z}_t$, F and G give wide range of flexible models:

BART

F estimated using BART, G omitted (standard case in Huber and Rossini)

Models for the Conditional Mean

A general non-parametric multivariate regression

$$\boldsymbol{y}_t = F(\boldsymbol{x}_t) + \eta_t, \quad \eta_t = G(\boldsymbol{z}_t) + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(\mathbf{0}_M, \Sigma_t).$$

Different choices for $\boldsymbol{z}_t$, F and G give wide range of flexible models:

mixBART

F is linear, $\boldsymbol{z}_t = \boldsymbol{x}_t$, G estimated using BART

- Shocks η_t, have non-linear regression specification
- Aims to control for any non-linear effects that persist after controlling for linear relations

Models for the Conditional Mean

A general non-parametric multivariate regression

$$\boldsymbol{y}_t = F(\boldsymbol{x}_t) + \eta_t, \quad \eta_t = G(\boldsymbol{z}_t) + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(\mathbf{0}_M, \Sigma_t).$$

Different choices for $\boldsymbol{z}_t$, F and G give wide range of flexible models:

errorBART

F is linear, $\boldsymbol{z}_t = (\eta'_{t-1}, \ldots, \eta'_{t-p})'$, G estimated using BART

- Flexible adjustments of the conditional mean in the presence of large past shocks.
- During recessions (e.g.Covid-19) could help to quickly adjust forecasts after large forecast errors

Models for the Conditional Mean

A general non-parametric multivariate regression

$$\boldsymbol{y}_t = F(\boldsymbol{x}_t) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t = G(\boldsymbol{z}_t) + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}_M, \boldsymbol{\Sigma}_t).$$

Different choices for $\boldsymbol{z}_t$, F and G give wide range of flexible models:

fullBART

$\boldsymbol{\Sigma}_t$ is diagonal matrix, thus system of independent regression models.

- i^{th} equation is given by:

$$y_{it} = f_i(\boldsymbol{x}_t) + g_i(\boldsymbol{z}_{it}) + \varepsilon_{it},$$

with $\boldsymbol{z}_{it} = (\varepsilon_{1t}, \ldots, \varepsilon_{i-1,t})'$

Models for the Conditional Variance

- For every model for conditional mean we try three different treatments of error variances
  1. Homoscedastic variances
  2. Stochastic volatility (SV)
  3. Heteroscedastic BART (heteroBART)

Decomposition of the VC matrix

$$\Sigma_t = \boldsymbol{Q}\boldsymbol{H}_t\boldsymbol{Q}'$$

- $\boldsymbol{Q}$ lower triangular matrix and $\boldsymbol{H}_t = \text{diag}(e^{v_1(\boldsymbol{w}_t)}, \ldots, e^{v_M(\boldsymbol{w}_t)})$
- v_j is a function approx. by BART which depends on covariates: $\boldsymbol{w}_t = (t, \boldsymbol{x}_t')'$

- Similar to Heteroscedastic BART via multiplicative regression trees: Pratola, Chipman, George and McCulloch (2020, JCGS)

Bayesian Inference

- Need prior plus MCMC algorithm
- See paper for details of both
- Prior features (in a nutshell):
  - Automatic choice of prior hyperparameters from BART literature
  - Horseshoe prior used for any linear conditional mean coefficients
- MCMC features (in a nutshell):
  - MCMC methods mostly standard, combining methods from Bayesian VAR literature with BART literature
  - Novel updating step for heteroBART
  - MCMC computationally fast, capable of scaling to large VARs

Data

- Real time quarterly data set of 23 major US macroeconomic and financial variables 1973Q2-2020Q4
- Variables transformed to stationarity
- We ran everything twice: using data through 2019 (excluding pandemic) and full sample through 2020
- All models feature five lags
- I will present results through 2020:
  - Forecast horizons $h \in \{1, 4, 8, 12\}$
  - Forecast evaluation period begins in 1997
  - Forecast inflation, unemployment and GDP growth
  - Metrics: CRPS, quantile-weighted CRPS (Gneiting and Ranjan, 2011, JBES) and quantile scores
  - Results benchmarked to BVAR-SV

Summary of Findings

- The empirical results of the paper can be summarized as follows:
  - Overall BART-based models improve upon the benchmark BVAR-SV (especially for longer forecast horizons)
  - Putting BART in conditional mean improves tail forecasts
  - Volatility: heteroBART is typically better than SV
  - Homoskedastic BART often forecasts very well (after putting nonlinearities in conditional mean, less important to allow for heteroskedasticity)
  - More complex BART specifications add only small improvements relative to the basic BART model
  - Little evidence of asymmetry (contrary to ”vulnerable growth” findings, but more consistent with ”capturing macroeconomic tail risks” work of Carriero, Clark and Marcellino)
- To illustrate these findings we present forecasting results for inflation and conditional forecasts of the unemployment rate

Quantile weighted CRPS: GDPCTPI

Model	CRPS				qwCRPS-tails				qwCRPS-left			
	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12
BVAR cons	1.02	1.09**	1.08	1.08	1.05	1.14**	1.15**	1.15	1.00	1.01	0.97	0.96
BART cons	1.05	0.87**	0.75***	0.71***	1.04	0.86***	0.74***	0.70***	1.04	0.88**	0.77***	0.74***
mixBART cons	1.01	0.95	0.82**	0.79***	1.00	0.92	0.83**	0.85*	1.00	0.98	0.84*	0.82*
errorBART cons	0.99	1.03	0.93	0.84**	0.97	0.99	0.91	0.83***	0.98	1.02	0.94	0.90**
fullBART cons	1.02	0.87***	0.76***	0.71***	1.02	0.86***	0.75***	0.71***	1.02	0.88**	0.77***	0.75***
BVAR SV	0.57	0.69	0.89	1.03	0.06	0.07	0.09	0.10	0.09	0.11	0.13	0.14
BART SV	1.01	0.86***	0.78***	0.74***	1.00	0.85***	0.77***	0.73***	1.00	0.86**	0.78***	0.76***
mixBART SV	1.01	0.92	0.82**	0.78***	1.01	0.90**	0.82***	0.83**	1.01	0.95	0.83*	0.81*
errorBART SV	0.99	0.96	0.85*	0.77**	0.97	0.93	0.84**	0.78***	0.99	0.97	0.88	0.83*
fullBART SV	1.14**	0.87**	0.74***	0.71***	1.17**	0.87**	0.74***	0.70***	1.15*	0.88*	0.76***	0.74***
BVAR heteroBART	0.98	0.98	0.97*	0.96**	0.97	0.98	0.97	0.98	0.97	1.00	1.02	1.01
BART heteroBART	1.12*	0.88**	0.77***	0.73***	1.14**	0.87***	0.76***	0.72***	1.11	0.90*	0.79***	0.77***
mixBART heteroBART	0.98	0.91*	0.81***	0.78***	0.98	0.89**	0.81***	0.81***	0.97	0.94	0.83**	0.82**
errorBART heteroBART	0.99	1.03	0.93	0.85**	0.97	0.99	0.91	0.84***	0.98	1.02	0.95	0.92
fullBART heteroBART	1.09	0.87***	0.75***	0.71***	1.08	0.87***	0.75***	0.71***	1.08	0.88**	0.78***	0.75***

Table: Cumulative ranked probability score (CRPS) and quantile weighted CRPSs for GDPCTPI.

Quantile scores: GDPCTPI

Model	QS5				QS10				QS25			
	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12	h=1	h=4	h=8	h=12
BVAR cons	1.03	1.07	1.26**	1.06	0.95	1.00	1.09	0.97	1.03	0.96	0.84	0.84*
BART cons	1.15	0.98	0.87	0.87	1.03	0.90*	0.80**	0.77**	1.03	0.88*	0.76***	0.75***
mixBART cons	0.95	0.95	1.00	0.89	0.94	0.97	0.95	0.93	1.02	0.99	0.82	0.87
errorBART cons	0.91**	0.93	1.01	0.98	0.91*	0.96	0.98	0.98	0.98	1.01	0.93	0.91**
fullBART cons	1.13	0.97	0.89	0.87	1.02	0.91*	0.81**	0.79**	1.00	0.88*	0.76***	0.76***
BVAR SV	0.13	0.14	0.15	0.20	0.21	0.24	0.25	0.32	0.34	0.42	0.50	0.53
BART SV	1.08	0.94	0.87	0.85	0.97	0.87**	0.81**	0.76**	0.99	0.86**	0.76***	0.76***
mixBART SV	0.95	0.95	0.98	0.86	0.95	0.96	0.95	0.90	1.02	0.95	0.81*	0.85
errorBART SV	0.95	0.98	1.00	0.93***	0.93	0.96	0.95	0.93**	0.99	0.96	0.88	0.85*
fullBART SV	1.48**	1.00	0.88	0.84	1.24*	0.92	0.81**	0.76**	1.14	0.87*	0.75***	0.74***
BVAR heteroBART	1.00	1.09	1.12	1.16	0.94*	1.04	1.11	1.09*	0.96*	1.00	1.02	1.06
BART heteroBART	1.31*	1.05	0.91	0.91	1.13	0.95	0.82*	0.80**	1.09	0.89*	0.80***	0.79***
mixBART heteroBART	0.96	0.95	0.98	0.89	0.93*	0.95	0.93	0.89	0.98	0.95	0.81**	0.86
errorBART heteroBART	0.93**	0.93	1.05*	1.05	0.91**	0.97	0.99	1.01	0.99	1.01	0.95	0.94
fullBART heteroBART	1.22	1.03	0.93	0.87	1.07	0.93	0.85	0.79**	1.08	0.88**	0.77***	0.77***

Table: Quantile scores (QS) for GDPCTPI.

The Role of Financial Conditions for Tail Forecasting

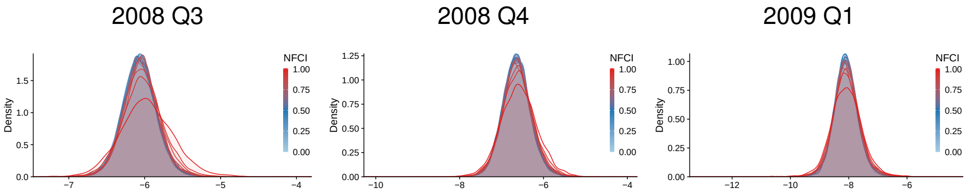
- Much interest in role of financial conditions (NFCI) in driving negative tail risks to economic activity
- Compare BART versus BVAR with same treatment of heteroskedasticity: BVAR-heteroBART and BART-heteroBART
- Conditional forecast of (negative of) unemployment rate
- Conditioning: NFCI paths over the forecast horizon different quantiles of the NFCI
- Step size 0.05 leads to 21 paths of the NFCI for which we produce conditional forecasts

A Closer Look at the Financial Crisis and the Pandemic

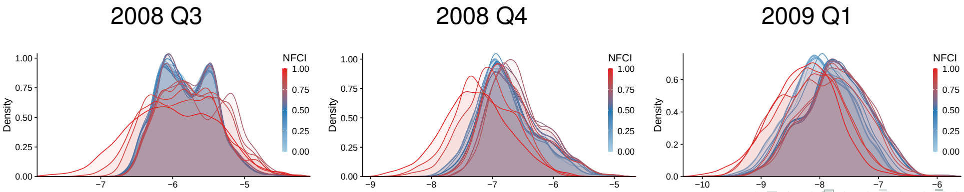
- Next figures plot one step ahead predictive densities conditional on different NFCI paths
- Blue = low values (loose financial conditions), red = high values (tight financial conditions)
- Conditioning on different financial settings has much larger effects on predictive distributions in the BART-heteroBART specification especially in financial crisis
- BART: Non-Gaussian distributions, with fat tails, asymmetries, or even multi-modality.

Conditional forecasts: Great Recession

BVAR heteroBART

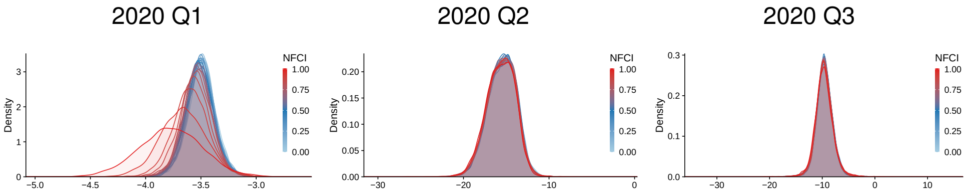


BART heteroBART

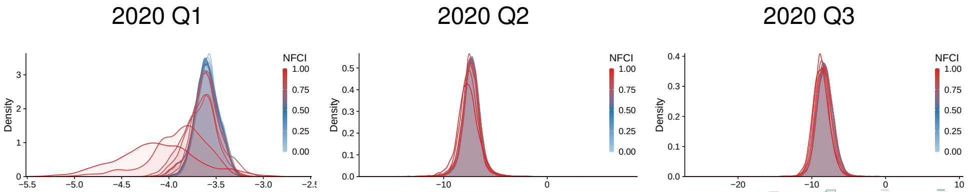


Conditional forecasts: Pandemic

BVAR heteroBART



BART heteroBART



Conclusions

- We have developed several non-parametric VARs using regression trees and associated scalable MCMC methods
- Can BART improve forecasts of tail risk and in extreme times such as pandemic?
  - Yes! But also improves entire predictive density and forecasting throughout sample
  - Once conditional mean is modeled using BART, less evidence for heteroskedasticity