Nowcasting with Mixed Frequency Data Using Gaussian Processes

OESTERREICHISCHE NATIONALBANK
EUROSYSTEM

Niko Hauzenberger,[1] Massimiliano Marcellino,[2]
Michael Pfarrhofer,[3] Anna Stelzer[4]

[1]U Strathclyde [2]Bocconi U [3]WU Vienna [4]OeNB

Anna Stelzer
Monetary Policy Section

Frankfurt am Main,
28 November 2024

Introduction: Combining Three Literatures

1. **MI**xed **DA**ta **S**ampling (**MIDAS**)

   Ghysels *et al.* (2007), with additional results in Ghysels (2016)

   Leverage HF information *efficiently* for predicting LF target variable

2. (*Macro*) **Big Data**, many time-series with few observations

   More information is better than less — but: overfitting; penalized or shrinkage estimators (MF-context: Babii *et al.*, 2022; Mogliani and Simoni, 2021, 2024)

3. (*Bayesian*) **Machine Learning**

   Flexible algorithms and computational power to uncover complex economic relationships (nonlinear and/or unknown)

Mixed Data Sampling (MIDAS): Intuition

- We have LF target $\{y_t\}_{t=1}^{T_L}$, e.g., GDP growth and a HF predictor $\{z_t\}_{t=1}^{T_H}$, e.g., IP

  $m = T_H/T_L$ indicates how many HF observations there are for each LF unit

  lag operator $L^{p/m} z_t = z_{t-p/m}$

$$y_t = \mathfrak{B}(L^{1/m}, \tilde{\boldsymbol{b}}) z_t + \epsilon_t \quad \longrightarrow \quad y_t = \sum_{p=0}^{P_H-1} \mathrm{B}(p, \tilde{\boldsymbol{b}}) L^{p/m} z_t + \epsilon_t$$

  distributed lag model with MFs, $\mathrm{B}(p, \tilde{\boldsymbol{b}}) = \sum_{l=0}^{\mathbb{L}-1} \tilde{b}_l \varphi_l(p)$

- $\varphi_l(p)$'s are basis functions — store in $\mathbf{w}_p = (\varphi_0(p), \ldots, \varphi_{\mathbb{L}-1}(p))'$

MIDAS: Intuition

- Using $W = (\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_{P_H-1})'$ and $\tilde{\boldsymbol{b}} = (\tilde{b}_0, \ldots, \tilde{b}_{\mathbb{L}-1})'$ and $\boldsymbol{x}_t = W'\tilde{\boldsymbol{z}}_t$:

$$y_t = \tilde{\boldsymbol{z}}_t' W \tilde{\boldsymbol{b}} + \epsilon_t = \boldsymbol{x}_t' \tilde{\boldsymbol{b}} + \epsilon_t, \qquad \boldsymbol{b} = W\tilde{\boldsymbol{b}} \quad \longrightarrow \quad b_p = \sum_{l=0}^{\mathbb{L}-1} \tilde{b}_l \varphi_l(p)$$

$$\underbrace{y_t}_{\texttt{2024-Q1}} \quad \text{and} \quad \tilde{\boldsymbol{z}}_t = (\underbrace{z_t}_{\texttt{2024-Mar}}, \underbrace{z_{t-1/3}}_{\texttt{2024-Feb}}, \underbrace{z_{t-2/3}}_{\texttt{2024-Jan}}, \underbrace{z_{t-1}}_{\texttt{2023-Dec}}, \ldots)'$$

- We break this linearity — nonlinear combination of basis functions

$$y_t = f(\boldsymbol{x}_t) + \epsilon_t = f(W'\tilde{\boldsymbol{z}}_t) + \epsilon_t,$$

Econometrics: (Nonlinear) MIDAS

- Direct specification, $\{x_t\}_{t=1}^{T_L}$ are M predictors

$$y_t = f(x_t) + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma_t^2)$$

- Let $\tilde{z}_{kt} = (z_{kt}, z_{kt-1/m}, z_{kt-2/m}, \ldots, z_{kt-(P_H-1)/m})'$ with P_H HF-lags
  $\boldsymbol{W}$ is a matrix of weights with dimension $P_H \times \mathbb{L}$
- We also use P_L lags of the target variable:

$$x_t = (y_{t-1}, \ldots, y_{t-P_L}, \overbrace{\underbrace{\tilde{z}'_{1t}}_{1\times P_H} \underbrace{\boldsymbol{W}}_{P_H\times\mathbb{L}}, \ldots, \underbrace{\tilde{z}'_{Kt}\boldsymbol{W}}_{1\times\mathbb{L}}}^{1\times K\mathbb{L}})', \qquad M = P_L + K\mathbb{L} \text{ predictors}$$

Econometrics: U-MIDAS & R-MIDAS

- **Unrestricted** (U-MIDAS) uses $\boldsymbol{W} = \boldsymbol{I}_{P_H}$, huge parameter space (particularly relative to T_L), with $M = P_L + KP_H$

  e.g., we have $M = 4 + 116 \cdot 12 \approx 1,400$ with $T_L \approx 100$ in our application

- **Restricted** MIDAS introduces HF lag-polynomials to reduce dimensionality
  - $\mathbb{L} = 1$ aggregates vectors to scalars, $M = P_L + K$ predictors
  - $\mathbb{L} > 1$ aggregates vectors to smaller vectors, $M = P_L + K\mathbb{L}$ predictors

Unrestricted (`u`), bridge (`br`); exponential Almon (`xalm`), Almon power (`alm`), Legendre (`leg`), Bernstein (`ber`), Fourier (`fou`) polynomials

Econometrics: Gaussian Processes (GPs)

- GP prior on the conditional means (see Williams and Rasmussen, 2006)

$$f(\boldsymbol{x}_t) \sim \mathcal{GP}\left(0, \mathcal{K}_{\boldsymbol{\kappa}}(\boldsymbol{x}_t, \boldsymbol{x}_t)\right)$$

  using $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{T_L})'$, this prior is a multivariate Gaussian:

$$\boldsymbol{f} \sim \mathcal{N}\left(\mathbf{0}_{T_L}, \mathcal{K}_{\boldsymbol{\kappa}}(\boldsymbol{X}, \boldsymbol{X})\right), \quad \boldsymbol{f} = (f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_{T_L}))'$$

- Kernel $\mathcal{K}_{\boldsymbol{\kappa}}(\boldsymbol{X}, \boldsymbol{X})$ with typical $[t, \tilde{t}]$ element $\mathcal{K}_{\boldsymbol{\kappa}}(\boldsymbol{x}_t, \boldsymbol{x}_{\tilde{t}})$
  - Covariance matrix which describes the shape of the joint distribution of $\boldsymbol{f}$
    Defines the prior shapes of functions (stationarity, smoothness)

Econometrics: Kernel (Hyperparameters)

- Squared exponential kernel, Euclidean distance:

$$\mathcal{K}_{\boldsymbol{\kappa}}(\boldsymbol{x}_t, \boldsymbol{x}_{\tilde{t}}) = \xi \cdot \exp\left(-\frac{1}{2}(\boldsymbol{x}_t - \boldsymbol{x}_{\tilde{t}})'\boldsymbol{\Lambda}(\boldsymbol{x}_t - \boldsymbol{x}_{\tilde{t}})\right)$$

  signal variance ξ and inverse length-scales λ_i, $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_M)$

- We assume a common inverse length-scale, such that $\boldsymbol{\Lambda} = \lambda \boldsymbol{I}_m$

$$\mathcal{K}_{\boldsymbol{\kappa}}(\boldsymbol{x}_t, \boldsymbol{x}_{\tilde{t}}) = \xi \cdot \exp\left(-\frac{\lambda}{2}(\tilde{\boldsymbol{z}}_t - \tilde{\boldsymbol{z}}_{\tilde{t}})'(\boldsymbol{I}_K \otimes \boldsymbol{W})(\boldsymbol{I}_K \otimes \boldsymbol{W}')(\tilde{\boldsymbol{z}}_t - \tilde{\boldsymbol{z}}_{\tilde{t}})\right).$$

  $\boldsymbol{I}_K \otimes (\lambda \boldsymbol{W}\boldsymbol{W}')$ is block diagonal, grouping high frequency lags

  - λ is a "global" parameter, $\boldsymbol{W}$ provides "local" adjustments

Econometrics: Conditional Mean, $f(\bullet)$

1. **Gaussian process** (GP) regression:

$$f(\boldsymbol{x}_t) \sim \mathcal{GP}\left(0, \mathcal{K}_{\boldsymbol{\kappa}}(\boldsymbol{x}_t, \boldsymbol{x}_t)\right)$$

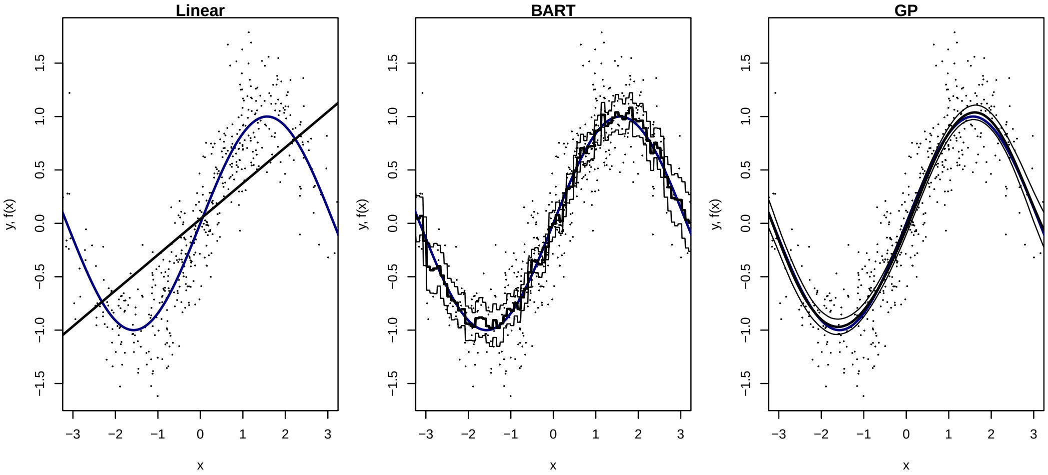
2. **Sum of regressions trees** (BART, Chipman *et al.*, 2010):

$$f(\boldsymbol{x}_t) \approx \sum_{s=1}^{S} \ell_s(\boldsymbol{x}_t | \mathcal{T}_s, \boldsymbol{\mu}_s)$$

3. **Bayesian Linear Regression** (BLR):

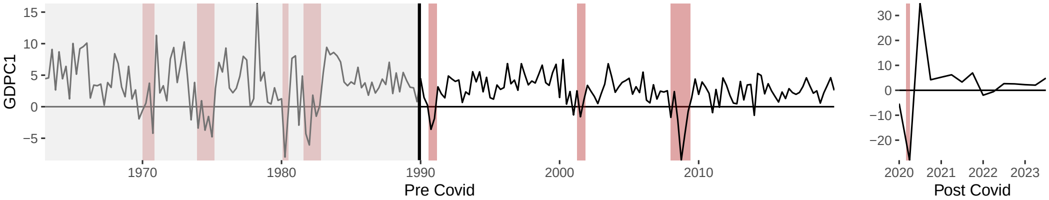
$$f(\boldsymbol{x}_t) = \boldsymbol{x}_t' \boldsymbol{\beta}$$

Example: How Competitors Fit Data

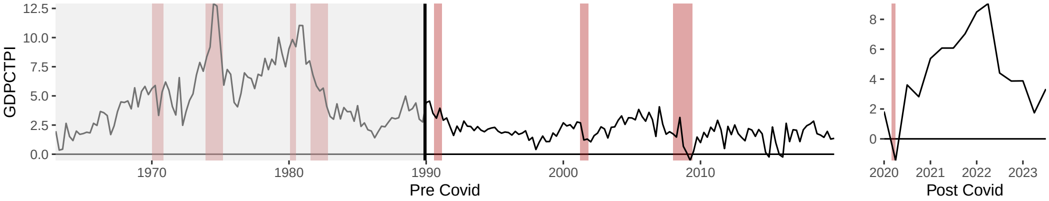


Data and Forecast Exercise

(a) *Output*



(b) *Inflation*



Data and Forecast Exercise

- Focus on nowcasts and short-horizon forecasts:

$$h \in \{0, \underbrace{\tfrac{1}{3}, \tfrac{2}{3}}_{\text{nowcast}}, 1, \underbrace{\tfrac{4}{3}, \tfrac{5}{3}}_{\text{forecast}}\}, \quad \text{monthly distance to target quarter}$$

- Losses for point, density and tail forecast performance (CRPSs)
- Small (`s`, $K = 12$), medium (`m`, $K = 23$), big data (`b`, $K = 116$) information sets

MIDAS	Mean	Variance
Unrestricted, $\mathbb{L} = P_H$ (`u`) $\mathbb{L} = 1$ (`br`, `xalm`) $\mathbb{L} = 3$ (`alm`, `ber`, `leg`, `fou`) $\mathbb{L} = 5$ (`ber`, `leg`)	Linear (`BLR`): $f(\boldsymbol{x}_t) = \boldsymbol{x}_t'\boldsymbol{\beta}$ `GP`: $f(\boldsymbol{x}_t) \sim \mathcal{GP}(0, \mathcal{K}_{\boldsymbol{\kappa}}(\boldsymbol{x}_t, \boldsymbol{x}_t))$ `BART`: $f(\boldsymbol{x}_t) \approx \sum_{s=1}^{S} \ell_s(\boldsymbol{x}_t \mid \mathcal{T}_s, \boldsymbol{\mu}_s)$	$\sigma_t^2 = \sigma^2$ (`hom`) σ_t^2 (`sv`)

Inclusion in Model Confidence Set (MCS)

GDPC1

MAE

		MAE
GP	sv−xalm−s	100% (2.95)
GP	sv−xalm−m	87% (2.98)
GP	sv−ber5−s	93% (3.00)
BLR	sv−xalm−s	97% (3.04)
BLR	sv−xalm−m	97% (3.04)
BLR	sv−xalm−b	83% (2.96)
BART	sv−xalm−s	73% (3.14)
BART	hom−xalm−s	80% (3.17)
BART	hom−u−m	77% (3.18)

CRPS

		CRPS
GP	sv−xalm−s	93% (2.28)
GP	sv−xalm−b	93% (2.31)
GP	hom−xalm−s	80% (2.32)
BLR	sv−xalm−s	90% (2.32)
BLR	sv−xalm−m	87% (2.32)
BLR	sv−xalm−b	93% (2.24)
BART	sv−xalm−s	63% (2.43)
BART	hom−xalm−s	70% (2.43)
BART	hom−u−m	87% (2.45)

GDPCTPI

MAE

		MAE
GP	sv−xalm−b	70% (0.97)
GP	hom−xalm−b	83% (0.96)
GP	hom−fou3−b	73% (0.99)
BLR	sv−xalm−m	93% (1.00)
BLR	sv−xalm−b	97% (0.94)
BLR	sv−u−s	90% (1.02)
BART	sv−xalm−m	67% (1.06)
BART	sv−xalm−b	63% (1.07)
BART	hom−xalm−b	67% (1.07)

CRPS

		CRPS
GP	sv−xalm−b	77% (0.67)
GP	sv−fou3−b	53% (0.69)
GP	hom−xalm−b	77% (0.68)
BLR	sv−xalm−m	93% (0.70)
BLR	sv−xalm−b	100% (0.66)
BLR	sv−u−s	83% (0.71)
BART	sv−xalm−m	60% (0.75)
BART	sv−xalm−b	57% (0.74)
BART	hom−xalm−b	50% (0.75)

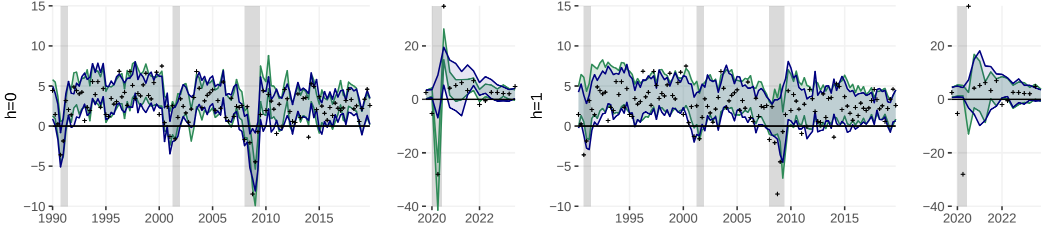
- Percentage of inclusion in superior model set across horizons and subsamples

Model-specific results: overview

- Simple `xalm` performs well across most specifications
- Controlling for heteroskedasticity improves forecasts in most cases
- Small dataset often sufficient with GP, best BLR is estimated with Big Data
- BLR/GP both perform well for nowcasts, GP offers some improvements BART typically outperformed
- Similar picture for both output growth and inflation predictions (bigger gains for output)

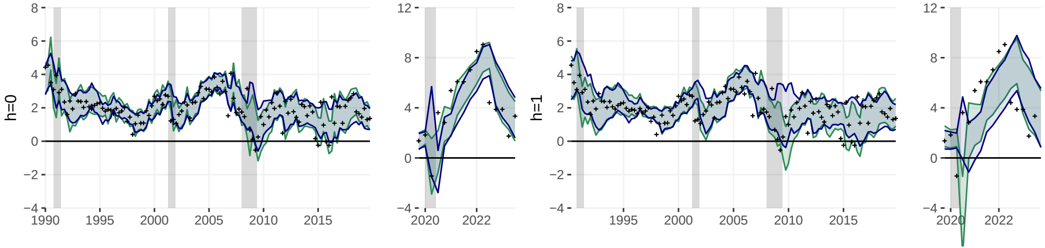
▸ Table model-specific results

Predictive Distributions (end-of-quarter): Output



- Blue for `GP-sv-xalm-s`, green for `BLR-sv-xalm-m`
- Credible sets narrow as more information flows in
- Lazy reactions of GP-version compared with BLR

Predictive Distributions (end-of-quarter): Inflation



- Blue for `GP-sv-xalm-s`, green for `BLR-sv-xalm-m`
- Similar picture as with output growth

Key Take-Aways

1. **Machine Learning**: Nonlinear means (especially GP) are relatively more important when trying to improve forecasts

   Improvements with GP for nowcasts; GP outperforms BART and SV helps

2. **Big Data/MIDAS**: Information set matters more when assuming linearity ("small" data often sufficient for flexible models — *unobserved* heterogeneity)

   Restricted lag polynomials are useful — direct `xalm` performs best

3. **Possibility of measuring variable importance:** variable selection is stable across horizons, denser models for inflation, sparser ones for output

   ▸ Variable importance

References i

Babii A, Ghysels E, and Striaukas J (2022), "Machine learning time series regressions with an application to nowcasting," *Journal of Business & Economic Statistics* **40**(3), 1094–1106.

Chipman HA, George EI, and McCulloch RE (2010), "BART: Bayesian additive regression trees," *The Annals of Applied Statistics* **4**(1), 266–298.

Clark TE, Huber F, Koop G, and Marcellino M (2024), "Forecasting US inflation using Bayesian nonparametric models," *Annals of Applied Statistics* **in-press**.

Ghysels E (2016), "Macroeconomics and the reality of mixed frequency data," *Journal of Econometrics* **193**(2), 294–314.

Ghysels E, Sinko A, and Valkanov R (2007), "MIDAS regressions: Further results and new directions," *Econometric Reviews* **26**(1), 53–90.

Mogliani M, and Simoni A (2021), "Bayesian MIDAS penalized regressions: estimation, selection, and prediction," *Journal of Econometrics* **222**(1), 833–860.

——— (2024), "Bayesian Bi-level Sparse Group Regressions for Macroeconomic Forecasting," *arXiv* **2404.02671**.

Williams CK, and Rasmussen CE (2006), *Gaussian processes for machine learning*, volume 2, Cambridge, MA, USA: MIT Press.

Woody S, Carvalho CM, and Murray JS (2021), "Model interpretation through lower-dimensional posterior summarization," *Journal of Computational and Graphical Statistics* **30**(1), 144–161.

Model-Specific Results: Output

GDPC1

MAE

Sample	Model	0	1/3	2/3	1	4/3	5/3
Full	GP−sv−xalm−s	0.77°	**0.79'**	0.92	**0.93'**	**0.95**	**0.98**
Full	GP−sv−xalm−m	0.77°	0.80'	0.92	0.93'	0.97	0.98
Full	GP−sv−xalm−b	0.79°	0.84	0.91	0.96	0.98	0.99
Full	BLR−sv−xalm−b	**0.64°**	0.85	**0.83**	1.01	1.02	1.02
Full	BART−hom−u−m	0.84'	0.87	0.93	0.97	1.01	1.01
Pre Covid	GP−sv−xalm−s	**0.76°**	**0.82'**	**0.94**	**0.93**	**0.98**	**1.01**
Pre Covid	GP−sv−xalm−m	0.76*	0.82'	0.94	0.94	1.00	1.02
Pre Covid	GP−sv−xalm−b	0.82°	0.89	0.99	0.96	1.02	1.02
Pre Covid	BLR−sv−xalm−b	0.76°	0.85'	0.98	0.99	1.03	1.07
Pre Covid	BART−hom−u−m	0.85'	0.90	0.98	0.98	1.01	1.02

h

CRPS

Sample	Model	0	1/3	2/3	1	4/3	5/3
Full	GP−sv−xalm−s	0.72°	**0.72'**	0.84'	0.92°	**0.94'**	**0.96**
Full	GP−sv−xalm−m	0.73°	0.73'	0.85'	**0.92°**	0.95'	0.96'
Full	GP−sv−xalm−b	0.72°	0.76'	0.84'	0.93°	0.95'	0.96
Full	BLR−sv−xalm−b	**0.57°**	0.77'	**0.74**	0.97	1.00	1.00
Full	BART−hom−u−m	0.80°	0.82'	0.88'	0.94'	0.97	0.98
Pre Covid	GP−sv−xalm−s	0.69*	**0.72***	**0.85°**	**0.88***	**0.92°**	**0.95'**
Pre Covid	GP−sv−xalm−m	0.70*	0.74*	0.86°	0.88*	0.93°	0.95'
Pre Covid	GP−sv−xalm−b	0.73*	0.80*	0.91'	0.89*	0.94°	0.96
Pre Covid	BLR−sv−xalm−b	**0.69***	0.76*	0.90'	0.91°	0.95'	1.01
Pre Covid	BART−hom−u−m	0.80*	0.84*	0.91°	0.92°	0.96	0.97

h

- Simple `xalm` performs well across most specifications
- Small dataset often sufficient with GP, best BLR is estimated with Big Data

Model-Specific Results: Inflation

GDPC1

MAE

Sample	Model	0	1/3	2/3	1	4/3	5/3
Full	GP−sv−xalm−s	0.91'	0.92'	0.91°	**0.91'**	**0.93'**	**0.94**
Full	GP−sv−xalm−m	0.90°	0.88'	0.90°	0.95	0.95	0.98
Full	GP−sv−xalm−b	0.89'	0.86'	**0.85°**	0.94	0.97	0.98
Full	BLR−sv−xalm−m	**0.85***	**0.82***	0.89°	0.96	1.03	0.98
Full	BART−sv−xalm−s	0.96	0.93	0.93	1.01	1.00	0.98
Pre Covid	GP−sv−xalm−s	**0.85***	0.82*	**0.86°**	**0.84°**	**0.87°**	**0.87°**
Pre Covid	GP−sv−xalm−m	0.86°	**0.79***	0.86°	0.90'	0.91'	0.94
Pre Covid	GP−sv−xalm−b	0.93	0.86'	0.92	0.97	1.00	0.99
Pre Covid	BLR−sv−xalm−m	0.87°	0.79*	0.90'	0.90	0.91	0.94
Pre Covid	BART−sv−xalm−s	0.93	0.90'	0.91'	0.98	0.97	0.94

h

CRPS

Sample	Model	0	1/3	2/3	1	4/3	5/3
Full	GP−sv−xalm−s	0.87*	0.88*	0.88*	**0.87***	**0.88***	**0.89°**
Full	GP−sv−xalm−m	0.85*	0.85*	0.87*	0.91°	0.90°	0.93'
Full	GP−sv−xalm−b	0.86°	0.84°	**0.83°**	0.90'	0.91'	0.93
Full	BLR−sv−xalm−m	**0.82***	**0.80***	0.87*	0.91	0.96	0.91'
Full	BART−sv−xalm−s	0.93	0.92	0.93	0.97	0.96	0.94
Pre Covid	GP−sv−xalm−s	**0.82***	0.80*	0.84*	**0.81***	**0.83***	**0.83***
Pre Covid	GP−sv−xalm−m	0.82*	**0.77***	**0.84***	0.86*	0.87*	0.90°
Pre Covid	GP−sv−xalm−b	0.90'	0.86°	0.91'	0.93	0.94	0.95
Pre Covid	BLR−sv−xalm−m	0.85*	0.78*	0.88*	0.84*	0.84*	0.87°
Pre Covid	BART−sv−xalm−s	0.92'	0.92'	0.93'	0.94	0.92'	0.91'

h

- Similar picture as with output growth predictions

Back

Which Predictors are Important?

- Black box — due to nonlinear/nonparametric conditional means
  Approximation: Clark *et al.* (2024) inspired by Woody *et al.* (2021)

- Sparse linear approximation for predictive median $\hat{y}_{t+h}$ over holdout

$$\hat{\mathfrak{b}}_h = \min_{\mathfrak{b}_h} \sum_{t=T_0}^{T} (\hat{y}_{t+h} - \boldsymbol{x}_t' \mathfrak{b}_h)^2 + \varrho \sum_{i=1}^{M} |\mathfrak{b}_{hi}| \qquad \text{(LASSO)}$$

  Cross-validation (CV) for penalty ϱ — selects which variables affect different parts of the predictive distribution

Variable Importance: Summary

- Variable selection is rather stable across horizons
- Denser models for inflation, sparser ones for output
  - Rather symmetric across quantiles for inflation
  - Even more sparsity for tails of output

- BLR and GP often choose similar predictors

  Output: industrial production (durable materials, nondurable consumer goods), personal consumption expenditures, manufacturing/trade sales

  Inflation: labor market (hours, earnings); housing; real money stock (M2)

Back