

# Combining Artificial and Expert Intelligence: Using Machine Learning to Improve the World Economic Outlook

Marijn (Maryn) A. Bolhuis

University of Toronto and University of Oxford

*[marijn.bolhuis@mail.utoronto.ca](mailto:marijn.bolhuis@mail.utoronto.ca)*

February 4, 2019

## Motivation

- Traditional macroeconomic forecasts (e.g. WEO) use country-specific expertise
- Key trade-off: using the past to predict the future...
  - bias ↓ ...
  - but variance ↑
- WEO forecasts tend to overfit (Hellwig, 2018) → variance ↑
- **This paper:** explore to what extent Machine Learning (ML) models can improve expert forecasts in WEO
  - without additional data
  - 1- to 5-year horizons
  - full sample of countries

# Approach

- Thought experiment:
  - give WEO data to data scientist at time of publication
  - how do predictions of non-expert compare to WEO?
- Goal is not to get most accurate prediction, rather:
  - compare performance of ML models with WEO
  - identify setting in which ML and WEO complement each other
- Study complements [Jung et al. \(2018\)](#) [▶ Literature](#)
  - no additional data → minimal information set
  - cover *all* countries

## Motivation: Current Approach

WEO forecast errors are consistently higher (r.t. average) in at least three different settings:

→ Ahir & Loungani (2014); Jalles et al. (2015); Eicher et al. (2018); Bluedorn & Leight (2018)

1. During periods of low growth ('optimism')

[» Details](#)

2. In low-income countries

[» Details](#)

3. During recessions

[» Details](#)

## Summary of Findings (2013-2016, 1-year)

Can we use ML models to improve macro-economic forecasting?

→ **14-19% ↓ RMSE overall**

- Gains are driven by
  1. countries w/ real GDP growth  $<$  median
  2. countries w/ real GDP p.c.  $<$  median
- ML models are not necessarily better at predicting recessions...
  - ...but help make the role of **(a)symmetric loss function** explicit

# Outline

1. Machine Learning Models
  - What, Why, When
2. Method
  - Data
  - Approach
3. Results
  - 3.1 Real economic growth
  - 3.2 Contractions ('recessions')
4. Conclusion

# ML Models for Macroeconomic Forecasting

- Algorithms purely focused on predicting  $y$  using data from  $x$
- Traditionally used for static, cross-section prediction
  - e.g. customer recommendations, image and speech recognition
- Practical challenges for dynamic forecasts in macroeconomic context
  1. Underlying process more volatile
  2. Lack of large, complete, training datasets
  3. Panel structure
  4. Data revisions are common

# Why and When Use ML Models for Macroeconomic Forecasting?

## Strengths

1. Generally outperform other quantitative models for prediction
2. Unbiased w.r.t. irrelevant context
3. Use *all* available data
4. Can be tuned to focus on predicting specific events (e.g. recessions)

## Weaknesses

1.  $\longleftrightarrow$  hard to understand mechanisms ('black box')
2.  $\longleftrightarrow$  vulnerable to relevant changes in context
3.  $\longleftrightarrow$  reliant on large, high-quality datasets
4.  $\longleftrightarrow$  'dumb' and reliant on model and variable selection by researcher



## What Are ML Models?

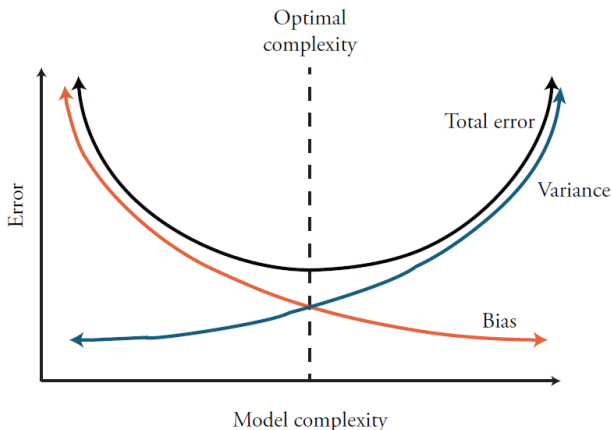
Prediction problem:

- $y_i = f(X_i, \beta) + \epsilon_i$
- $\min_{\beta, \alpha} \sum_{i=1}^N L(y_i - f(X_i, \beta)) \quad s.t. \quad \beta \in \Theta(\alpha)$ 
  - $\beta$  are parameters that determine specific function
  - $\alpha$  are 'tuning parameters' or 'regularizers' that determine model complexity

Function  $\tilde{f}(\alpha, \beta, X_i)$  maps input data ( $X_i$ ) to prediction ( $y_i$ ).

## Bias-Variance Trade-Off

$$L(z) = z^2 \rightarrow \mathbb{E}[L(\tilde{f}(\cdot) - f(\cdot))] = \text{Bias}[\tilde{f}(\cdot)]^2 + \text{Var}[\tilde{f}(\cdot)] + \text{Var}[\epsilon_i]$$



## What Are ML Models?

Prediction problem:

- $y_i = f(X_i, \beta) + \epsilon_i$
- $\min_{\beta, \alpha} \sum_{i=1}^N L(y_i - f(X_i, \beta)) \quad s.t. \quad \beta \in \Theta(\alpha)$ 
  - $\beta$  are parameters that determine specific function
  - $\alpha$  are 'tuning parameters' or 'regularizers' that determine model complexity

Function  $\tilde{f}(\alpha, \beta, X_i)$  maps input data ( $X_i$ ) to prediction ( $y_i$ ).

Estimation (i.e. picking  $\tilde{f}(\cdot)$ ) consists of 3 steps:

1. Given  $\alpha$ , pick loss-minimizing  $\hat{\beta}$  for subset (training) of data
2. Predict  $\hat{y}_i = f(x_i, \hat{\beta})|_{\alpha}$  for remaining (test) data
3. Pick  $\alpha$  with lowest loss for test data

# Linear Methods (OLS, LASSO, Ridge, Elastic Net)

- $\min_{\beta} \{ \sum_{i=1}^N (y_i - \beta_0 - \sum_{k=1}^K \beta_k x_{ik})^2 \}$ 
  - s.t.  $\sum_{k=1}^K \gamma |\beta_k| + (1 - \gamma)(\beta_k)^2 \leq \lambda$

## 1. OLS

- $\lambda \rightarrow \infty$

## 2. LASSO

- $\gamma = 1$

## 3. Ridge

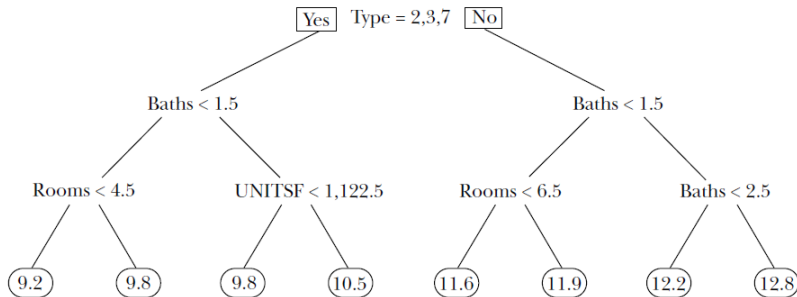
- $\gamma = 0$

## 4. Elastic Net

- $\gamma \in (0, 1)$

# Decision Trees

## A Shallow Regression Tree Predicting House Values

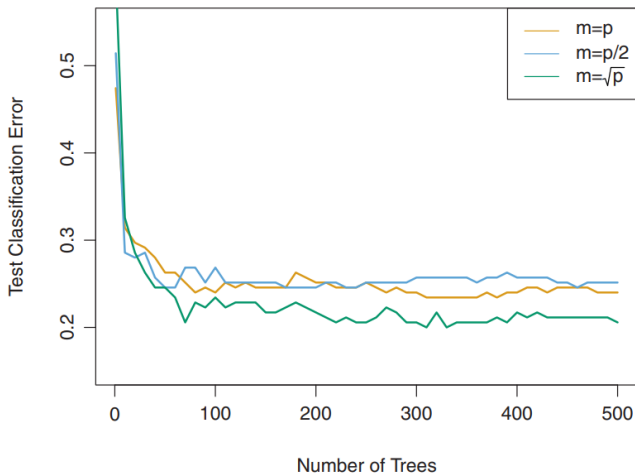


Source: Mullainathan & Spiess (2017)

## Random Forest (1)

- Exploits 'wisdom of the crowd' to reduce variance
  - Use different trees as building blocks
1. Take bootstrap sample
    - 'experience'
  2. At each split, use only randomly chosen  $m$  out of  $p$  predictors
    - 'expertise'
  3. Aggregate predictions of trees for each observation

## Random Forest (2)



**Source:** James et al. (2013). Gene classification with  $p = 500$  predictors. Single classification tree has test classification error of 0.457.

## Generalizing ML Models

Method	$\beta$	Most common $\alpha$
OLS	regression coefficients	-
LASSO	regression coefficients	$\sum_{k=0}^K  \beta_k  \leq \lambda_L$
Ridge	regression coefficients	$\sum_{k=0}^K (\beta_k)^2 \leq \lambda_R$
Elastic Net	regression coefficients	$\lambda$ , lasso weight
Decision tree	splits	depth, # of leaves, obs. per leaf
Random forest	aggregation rule	# vars, obs. per bootstrap



# Data

## Data

- Original 2004-2016 World Economic Outlook (April)
  - Impute missing data using random forests
- Merge w/ real GDP growth from 2018 World Economic Outlook (April)
  - take as actual outcomes

## Variable Selection

- Only country-level
- Only variables w/  $> 75\%$  coverage
- Log-level, log-difference, 1-5 year lags

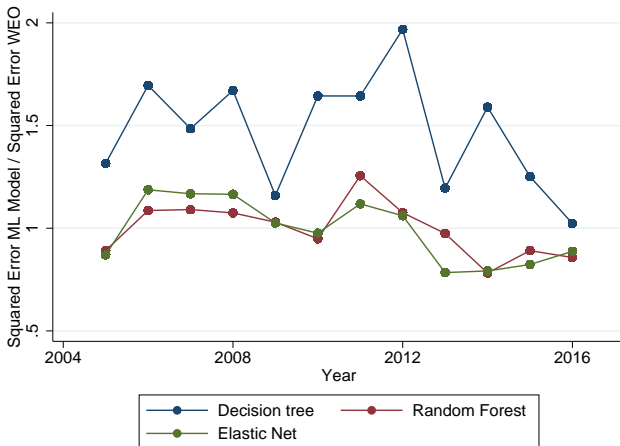
## Approach

- $y_{it} = f(x_{i,t-1}, x_{i,t-2}, x_{it,-3}, x_{it,-4}, x_{i,t-5}, \beta) + \epsilon_{it}$ 
  - $y_{it}$ : y-o-y real GDP growth in %
- Loss function:  $L(z) = z^2$  (squared error)
  - in addition, use  $L(z) = |z|$  (absolute error)
  - evaluate both mean and median
- Models:
  - Elastic Net
  - Individual decision tree
  - Random forest
- Focus on 2013-2016 period for 1-, 3-, 5-year ahead
  - results qualitatively robust

## Learning Rate of ML Models r.t. WEO

All three ML models improve over time r.t. WEO

» Extra



## 1-year Squared Error, 2013-2016

19% and 14% ↓ RMSE for Elastic Net and Random Forest



	WEO	Elastic Net	Decision Tree	Random Forest
<b>RMSE</b>	5.9	4.8	7.7	5.1
<b>RMedSE</b>	1.4	1.4	2.3	1.3
<b>SD SE</b>	248	176	593	198
<b>R-Max.SE</b>	63	60	123	57

RMSE: root-mean-squared error; RMedSE: root-median-squared error; SD SE: standard deviation of squared error; R-Max.SE: root of maximum squared error.

## 3-year Squared Error, 2013-2016

17% and 13% ↓ RMSE for Elastic Net and Random Forest



	WEO	Elastic Net	Decision Tree	Random Forest
<b>RMSE</b>	5.4	4.5	7.4	4.7
<b>RMedSE</b>	1.5	1.4	2.6	1.5
<b>SD SE</b>	174	126	257	135
<b>R-Max.SE</b>	45	41	64	41

RMSE: root-mean-squared error; RMedSE: root-median-squared error; SD SE: standard deviation of squared error; R-Max.SE: root of maximum squared error.

## 5-year Squared Error, 2013-2016

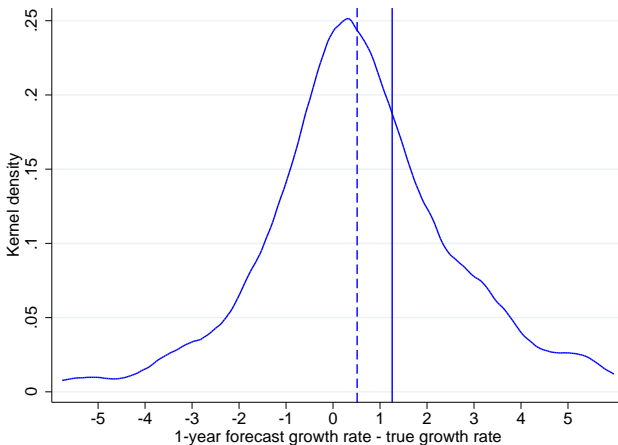
2% and 2% ↓ RMSE for Elastic Net and Random Forest

► Extra

	WEO	Elastic Net	Decision Tree	Random Forest
<b>RMSE</b>	5.2	5.1	6.3	5.1
<b>RMedSE</b>	1.7	1.5	2.4	1.5
<b>SD SE</b>	195	182	186	187
<b>R-Max.SE</b>	61	58	58	59

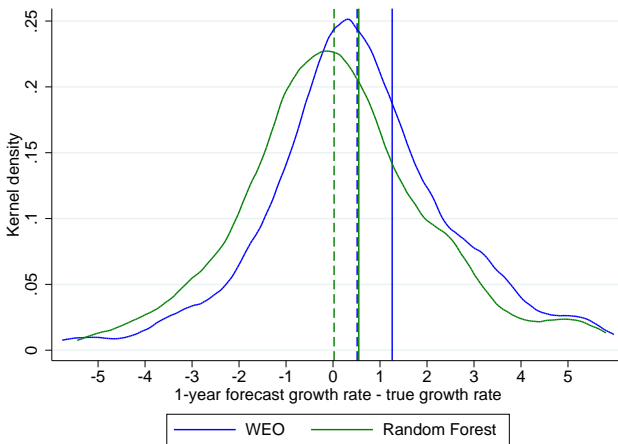
RMSE: root-mean-squared error; RMedSE: root-median-squared error; SD SE: standard deviation of squared error; R-Max.SE: root of maximum squared error.

## 1-year Error Rate: WEO



Vertical lines: (a) Solid blue: **mean** WEO error rate; (b) dashed blue: **median** WEO error rate.

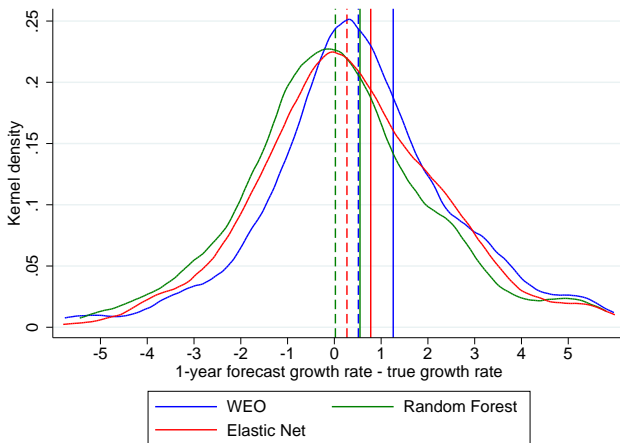
# 1-year Error Rate: WEO and Random Forest



Vertical lines: (a) Solid blue: **mean** WEO error rate; (b) dashed blue: **median** WEO error rate; (c) solid green: **mean** RF error rate; (d) dashed green: **median** RF error rate.

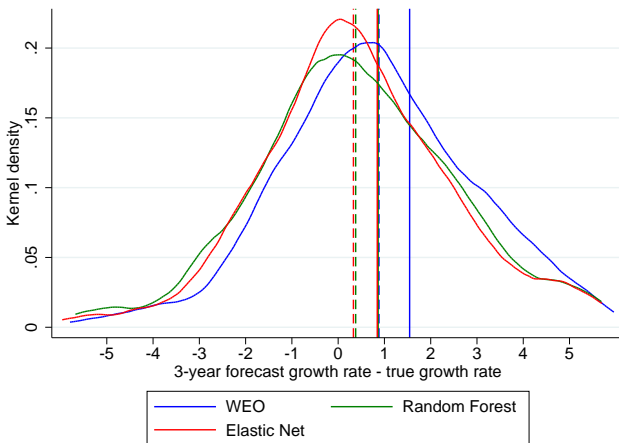


# 1-year Error Rate: WEO, Random Forest, and Elastic Net



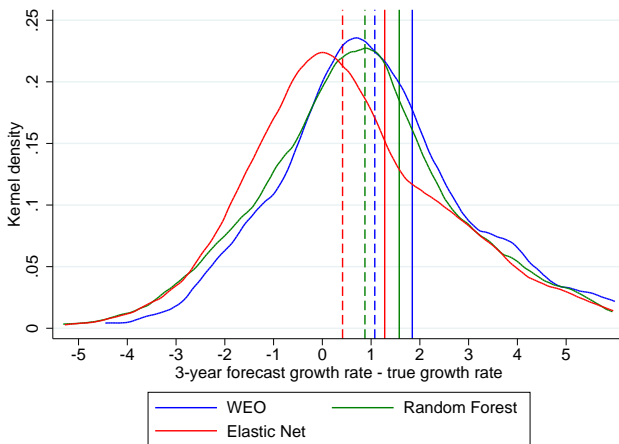
Vertical lines: (a) Solid blue: **mean** WEO error rate; (b) dashed blue: **median** WEO error rate; (c) solid green: **mean** RF error rate; (d) dashed green: **median** RF error rate; (e) solid red: **mean** EN error rate; (f) dashed red: **median** EN error rate.

## 3-year Error Rate: WEO, Random Forest, and Elastic Net



Vertical lines: (a) Solid blue: **mean** WEO error rate; (b) dashed blue: **median** WEO error rate; (c) solid green: **mean** RF error rate; (d) dashed green: **median** RF error rate; (e) solid red: **mean** EN error rate; (f) dashed red: **median** EN error rate.

## 5-year Error Rate: WEO, Random Forest, and Elastic Net

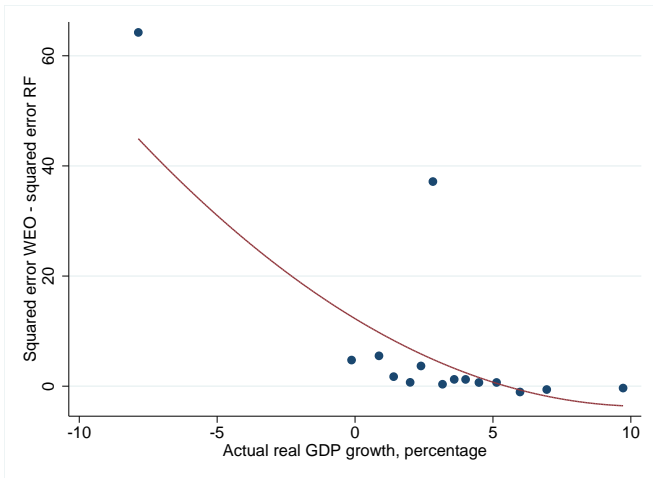


Vertical lines: (a) Solid blue: **mean** WEO error rate; (b) dashed blue: **median** WEO error rate; (c) solid green: **mean** RF error rate; (d) dashed green: **median** RF error rate; (e) solid red: **mean** EN error rate; (f) dashed red: **median** EN error rate.

## Performance in low growth years (1)

Random Forest outperforms in periods of low growth

► Extra

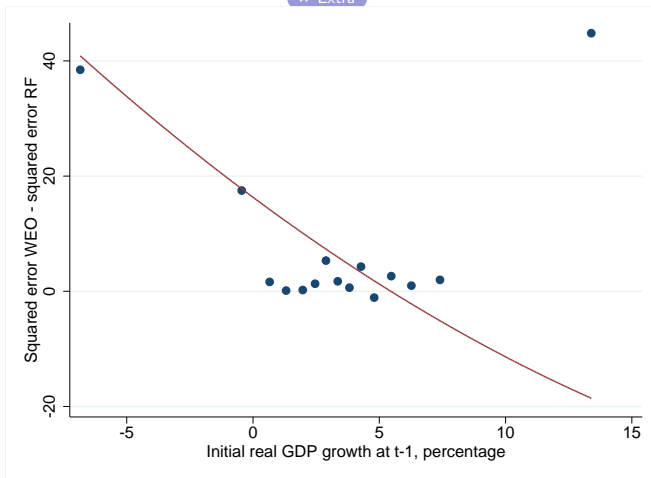


Binned scatter plot; y-variable represents mean value of 15 equally-sized bins.

## Performance in low growth years (2)

Random Forest outperforms in *initially* slower growing countries

» Extra

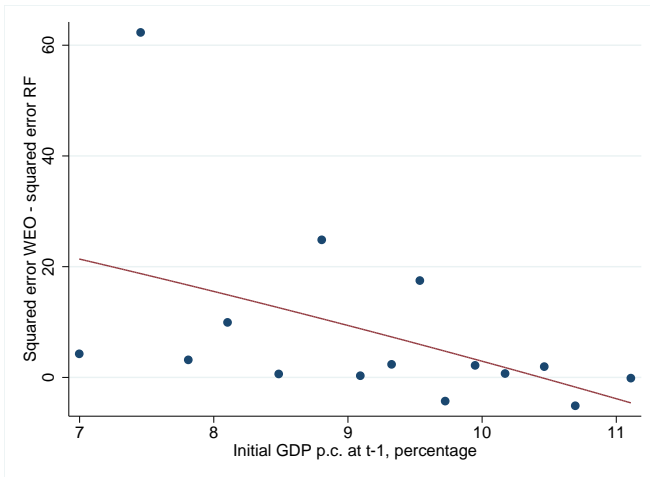


Binned scatter plot; y-variable represents mean value of 15 equally-sized bins.

## Performance and GDP p.c.

Random Forest outperform in *initially* poorer countries

► Extra



Binned scatter plot; y-variable represents mean value of 15 equally-sized bins.

## Regression framework

### Robust in regression framework (Random Forest)

[▶ Extra](#)

	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Real GDP growth	-5.654** (2.231)	-5.982*** (2.278)	-5.994*** (2.169)	-2.902* (1.473)
Log GDP p.c., t-1		-9.704* (4.956)	-6.482* (3.607)	-0.994 (2.009)
Africa			18.64 (13.53)	15.48** (7.408)
Asia			10.06* (5.874)	5.738 (4.806)
Europe			1.458 (6.527)	-2.093 (3.428)
Latin America			-5.213 (6.313)	0.653 (2.652)
Year FE	Y	Y	Y	Y
Continent FE	N	N	Y	Y
Observations	741	741	741	2176
Adjusted $R^2$	0.101	0.118	0.122	0.045

Standard errors in parentheses, clustered at country-level

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Mean squared error difference between WEO and RF  $\approx 8.0$ . Column (4) runs full regression on 2004-2016 sample.

## Combining ML and WEO Predictions

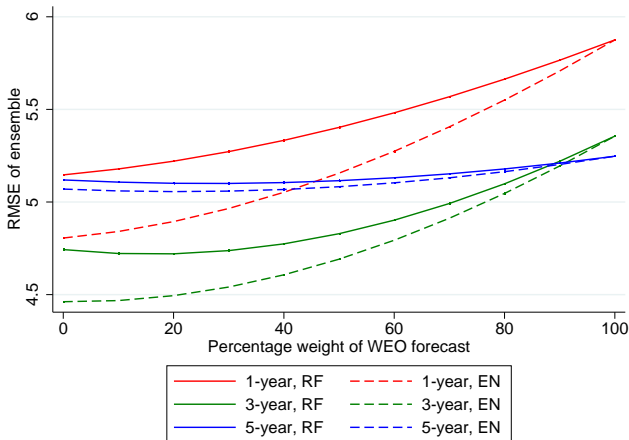
- ML models outperform in most settings
- What about a combination of ML models and WEO?  
→ complementarity between expert and ML expertise
- Combine by taking convex combination (ensemble) predicted growth rates
  - e.g.  $y_i^{convex} = 0.5 \cdot y_i^{RF} + 0.5 \cdot y_i^{WEO}$



## Ensemble of WEO and RF forecasts

WEO adds only little to ensemble in terms of RMSE

► Extra

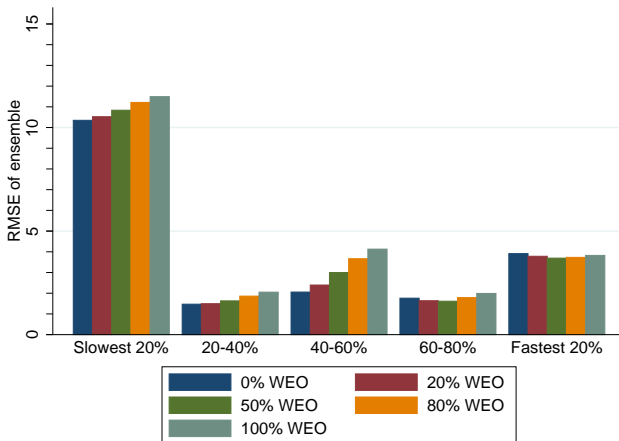




# 1-year forecasting gains: actual growth

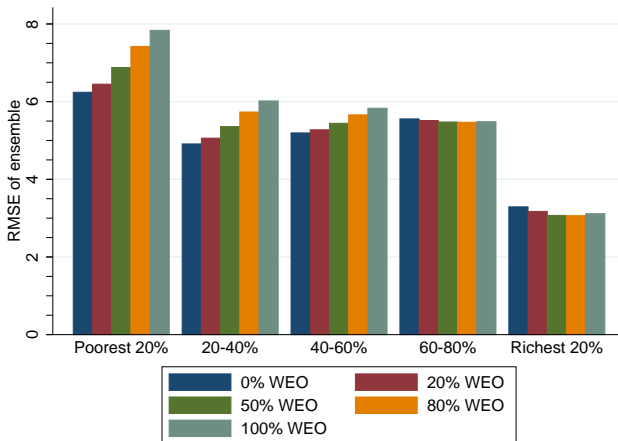
Ensemble slightly improves RMSE for two fastest growing quintiles

» Extra



# 1-year forecasting gains: initial GDP p.c.

Ensemble slightly improves RMSE for two richest quintiles



Motivation  
○○○○○

ML Models  
○○○○○○○○○○

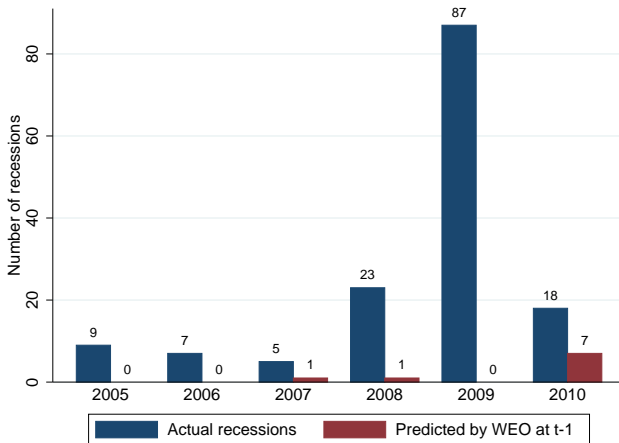
Method  
○○

Results: Growth  
○○○○○○○○○○○○○○○○○○○○

**Results: Recessions**  
●○○○○○○

Conclusion  
○

## WEO forecasts miss most recessions



## Formalizing the Asymmetric Loss Function

- WEO: many false negatives  $\longleftrightarrow$  few false positives
- Recessions occur only  $\approx 14\%$  of years  
→ need asymmetric loss function to call *any* recessions
- ...at cost of false positives ('can't miss if you don't shoot')
- ML models can help make this trade-off explicit
  - use random forest classification model with variable loss function
  - recession:  $y_i = 1$  if  $\Delta y_{n,t} < 0$

## A Loss Function Framework (Elkan, 2001)

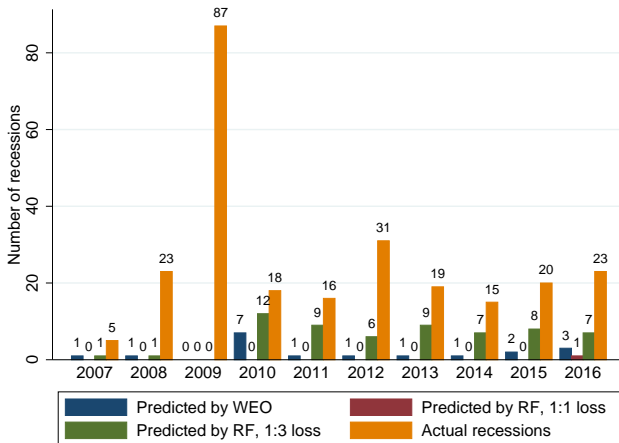
	Actual growth	Actual recession
Predict growth	0	$c_{fn}$
Predict recession	$c_{fp}$	0

- Symmetric loss function:  $c_{fp} = c_{fn}$ 
  - forecasts noisy + recessions infrequent  $\rightarrow$  predict growth
- Threshold of  $p = P(y_i = 1|X_i)$  satisfies:
  - $$\underbrace{p^* \cdot c_{fn}}_{\text{exp. cost false positive}} = \underbrace{(1 - p^*) \cdot c_{fp}}_{\text{exp. cost false negative}} \rightarrow p^* = \frac{c_{fp}}{c_{fn} + c_{fp}}$$
- Intuition:
  - high cost of recession  $\rightarrow$  certainty  $\rightarrow$  accept false negatives



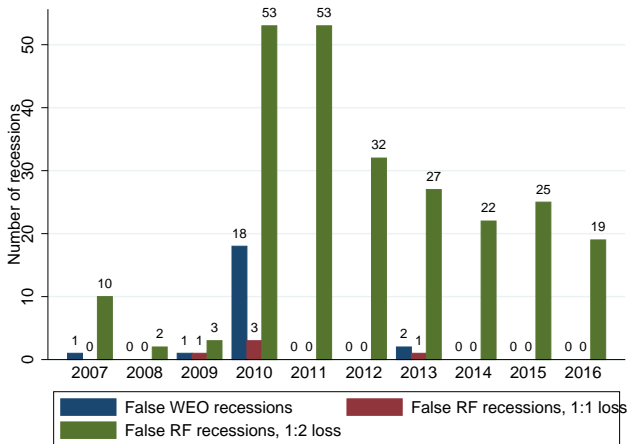
## Recessions: 1 Year Ahead (2007-2016)

Random Forest predicts more recessions with 1:3 loss function...

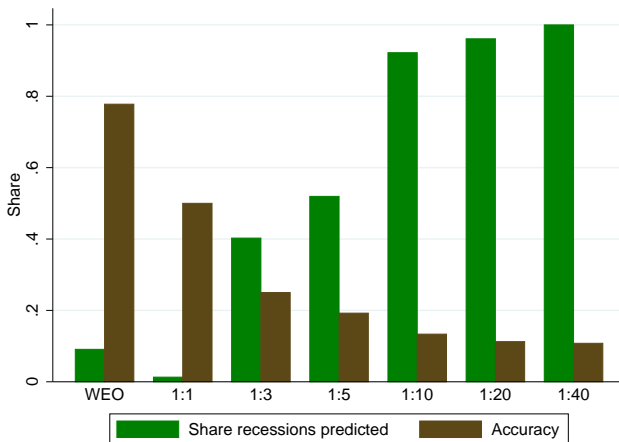


## Recessions: False Positives (2007-2016)

...but Random Forest with 1:3 loss function also predicts more false positives



## Success Rate and Accuracy of Different Loss Functions (2007-2016)



## Conclusion

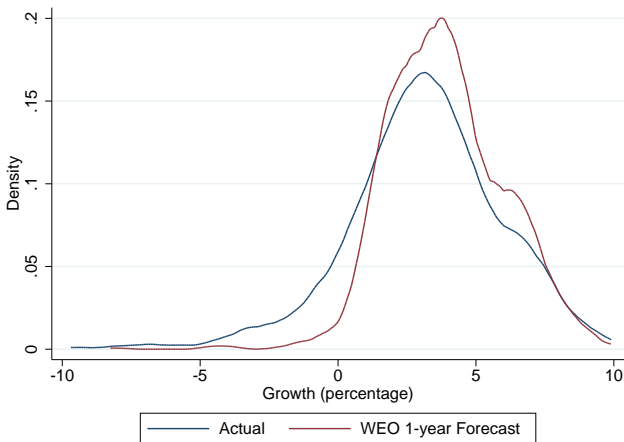
- Results suggest ML methods can help improve macroeconomic forecasts
- Although not goal of this paper, fine-tuning forecasts could improve prediction further
  - data: country-specific, global macroeconomic data, sentiment surveys, etc.
- However, ML methods are not a panacea:
  - substantial prediction errors remain
  - implementation is still researcher-dependent
- Additional applications
  - macroeconomic aggregates (current account, inflation, etc.)
  - discrete events (e.g. sudden stops)

## ◀ Motivation

- use ML models and additional data to improve 1-q and 1-y predictions for 7 high- and middle-income countries

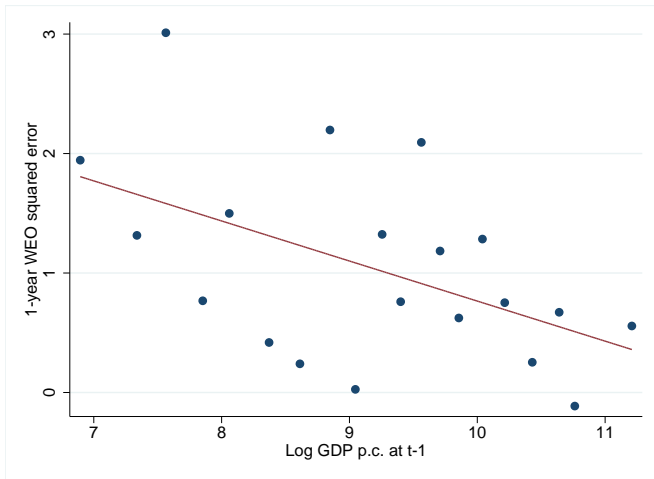
## ◀ Motivation

## WEO is over-optimistic in its 1-year forecasts



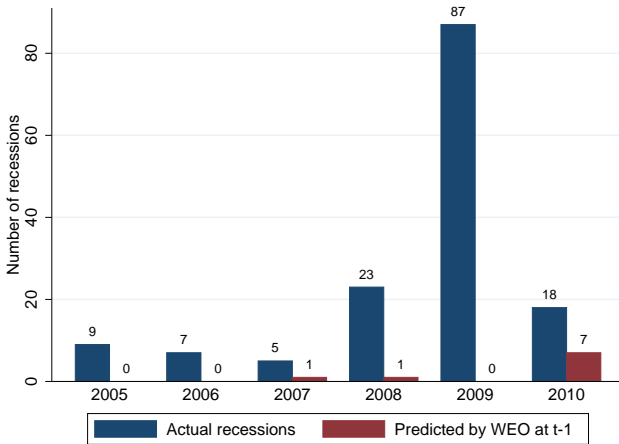
## ◀ Motivation

## WEO forecast errors are higher for low-income countries



## ◀ Motivation

## WEO forecasts miss most recessions





## ◀ Method

List of variables used from WEO, 2015:

- **GDP measures:** real GDP growth (+ p.c.), GDP at PPP (+ p.c.), GDP at int. (+ p.c.), nominal GDP in USD (+ p.c.), nominal GDP in LCU (+ p.c.), output gap, investment, exports, imports
- **Prices:** GDP deflator, PPP deflator, inflation
- **Government:** government balance, government expenditure, government net debt, government revenue,
- **Other:** current account, employment, labor, population, savings, unemployment



## ◀ Results: Growth



## ◀ Results: Growth



## ◀ Results: Growth



## ◀ Results: Growth



## Results: Growth, Summary 1-year, 2013-2017

Absolute error (EN, RF, 8 %)

	WEO	Elastic Net	Decision Tree	Random Forest
<b>AME</b>	2.5	2.3	3.8	2.3
<b>AMedE</b>	1.3	1.3	2.3	1.3
<b>SD</b>	5.3	4.2	6.7	4.6
<b>A-Max.</b>	63	60	123	57

Absolute error (EN 15%, RF 7%)

	WEO	Elastic Net	Decision Tree	Random Forest
<b>AME</b>	2.7	2.3	4.2	2.5
<b>AMedE</b>	1.5	1.4	2.6	1.5
<b>SD</b>	4.6	3.8	6.1	4.1
<b>A-Max.</b>	46	41	64	41



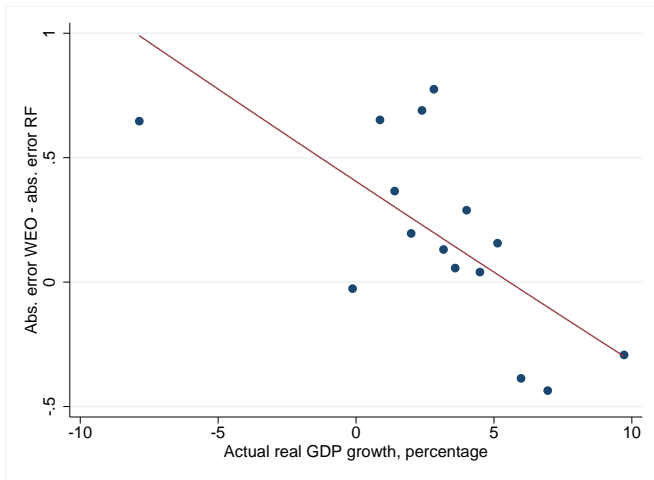
## Results: Growth, Summary 5-year, 2013-2017

Absolute error (EN, RF, 7%)

	WEO	Elastic Net	Decision Tree	Random Forest
<b>AME</b>	2.7	2.5	3.7	2.5
<b>AMedE</b>	1.7	1.5	2.4	1.5
<b>SD</b>	4.5	4.4	5.0	4.4
<b>A-Max.</b>	61	58	58	59

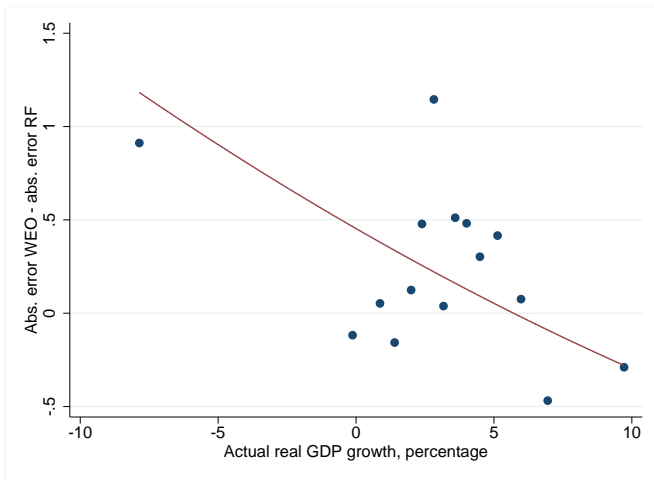
Random Forest outperform in periods of low growth: absolute error

◀ Results: Growth



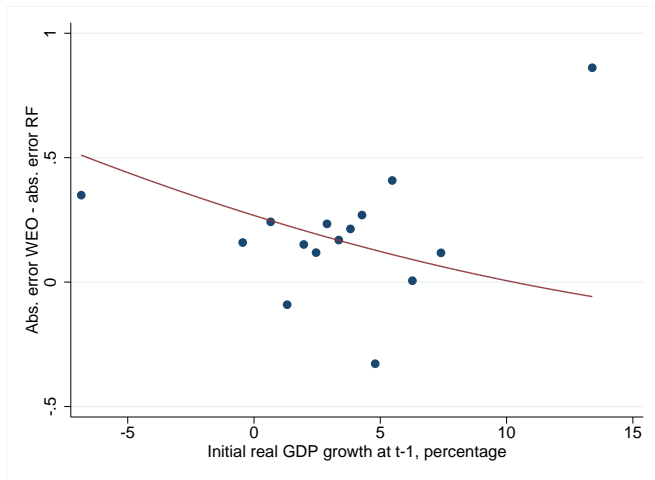
Elastic Net outperform in periods of low growth: absolute error

◀ Results: Growth

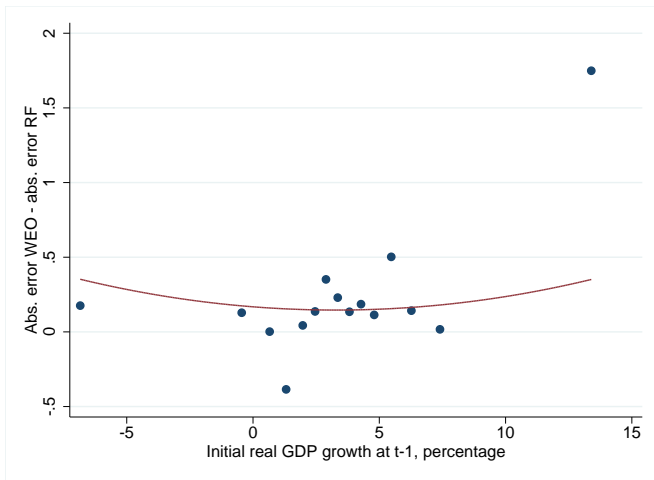




Random Forest outperforms in *initially* slower growing countries:  
absolute error ◀ Results: Growth



Elastic Net outperforms in *initially* slower growing countries:  
absolute error ◀ Results: Growth

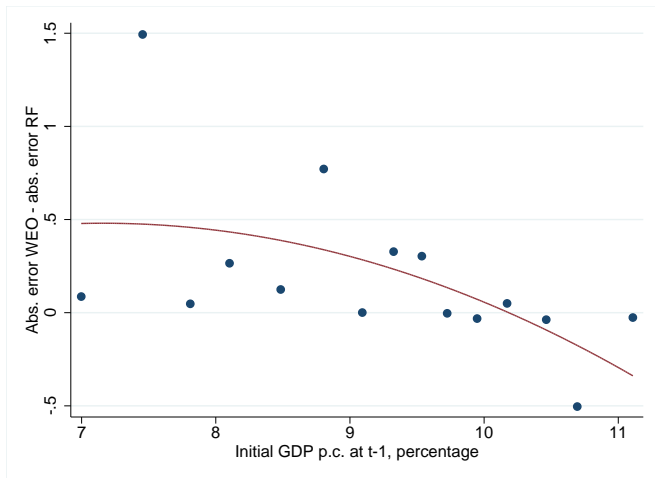


◀ Results: Growth

## Performance and GDP p.c.

Random Forest outperform in *initially* poorer countries: absolute error

◀ Results: Growth



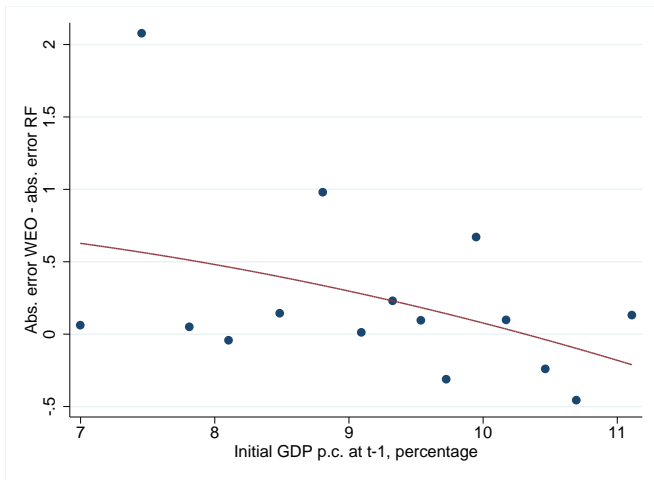
Notes



## Performance and GDP p.c.

Elastic Net outperform in *initially* poorer countries: squared error

◀ Results: Growth

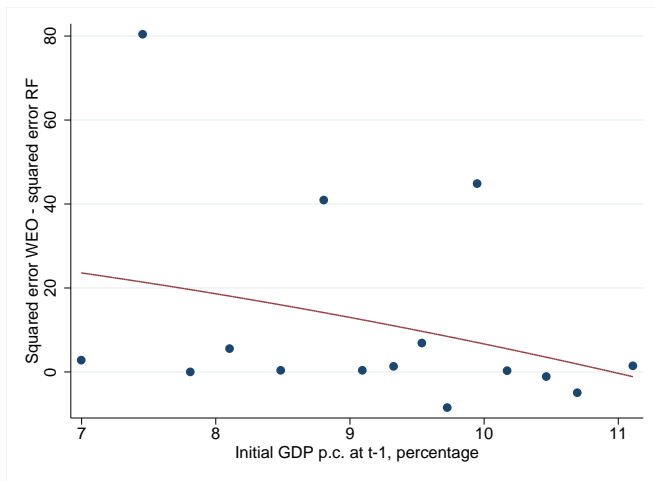


Notes

## Performance and GDP p.c.

Elastic Net outperform in *initially* poorer countries: squared error

◀ Results: Growth



Notes

# Regression framework

## Absolute error, Random Forest

[← Results: Growth](#)

	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Real GDP growth	-0.0760*** (0.0270)	-0.0843*** (0.0281)	-0.0862*** (0.0271)	-0.0660** (0.0284)
Log GDP p.c., t-1		-0.246** (0.107)	-0.217** (0.107)	-0.107* (0.0646)
Africa			0.375 (0.336)	0.340* (0.181)
Asia			0.414* (0.230)	0.157 (0.152)
Europe			0.155 (0.230)	0.133 (0.160)
Latin America			0.0633 (0.251)	0.0981 (0.141)
Year FE	Y	Y	Y	Y
Continent FE	N	N	Y	Y
Observations	741	741	741	2176
Adjusted $R^2$	0.031	0.050	0.050	0.032

Standard errors in parentheses, clustered at country-level

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## Regression framework

Absolute error, Elastic Net

← Results: Growth

	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Real GDP growth	-0.105 (0.0695)	-0.114 (0.0710)	-0.114* (0.0664)	-0.0675 (0.0453)
Log GDP p.c., t-1		-0.267* (0.148)	-0.0973 (0.148)	-0.0146 (0.0725)
Africa			0.567 (0.484)	0.529** (0.252)
Asia			0.169 (0.270)	0.136 (0.167)
Europe			-0.334 (0.292)	-0.0428 (0.150)
Latin America			-0.428 (0.276)	0.0254 (0.140)
Year FE	Y	Y	Y	Y
Continent FE	N	N	Y	Y
Observations	741	741	741	2176
Adjusted $R^2$	0.032	0.044	0.053	0.026

Standard errors in parentheses, clustered at country-level

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

# Regression framework

## Squared error, Elastic Net

[← Results: Growth](#)

	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Real GDP growth	-9.010* (5.412)	-9.392* (5.500)	-9.305* (5.272)	-4.244 (2.991)
Log GDP p.c., t-1		-11.29* (6.358)	-2.574 (6.597)	0.416 (2.255)
Africa			34.78 (22.84)	23.18* (13.28)
Asia			6.673 (10.48)	8.099 (8.030)
Europe			-9.510 (12.53)	-4.450 (4.745)
Latin America			-13.80 (10.43)	-0.176 (3.476)
Year FE	Y	Y	Y	Y
Continent FE	N	N	Y	Y
Observations	741	741	741	2176
Adjusted $R^2$	0.103	0.112	0.121	0.047

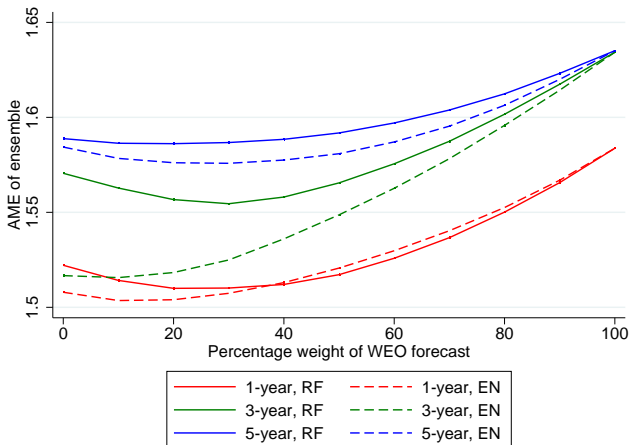
Standard errors in parentheses, clustered at country-level

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## Ensemble of WEO and RF forecasts

Mean absolute error

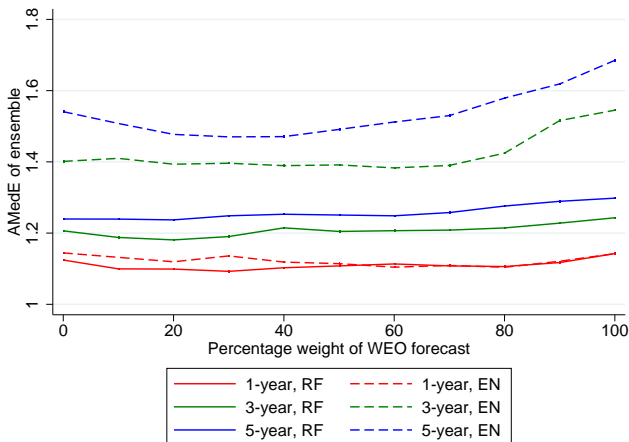
◀ Results: Growth



# Ensemble of WEO and RF forecasts

Median absolute error

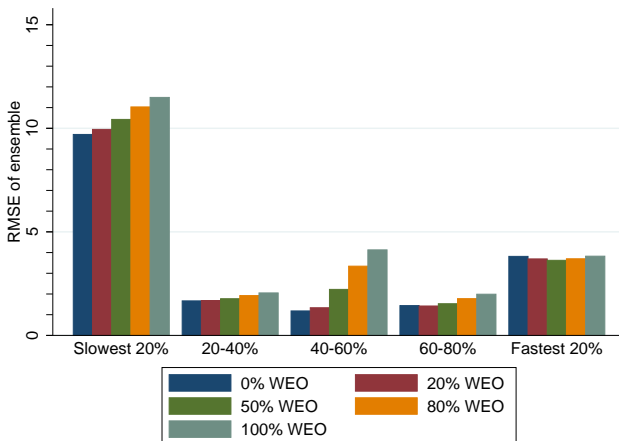
◀ Results: Growth



# 1-year forecasting gains: actual growth

Elastic Net, RMSE

◀ Results: Growth

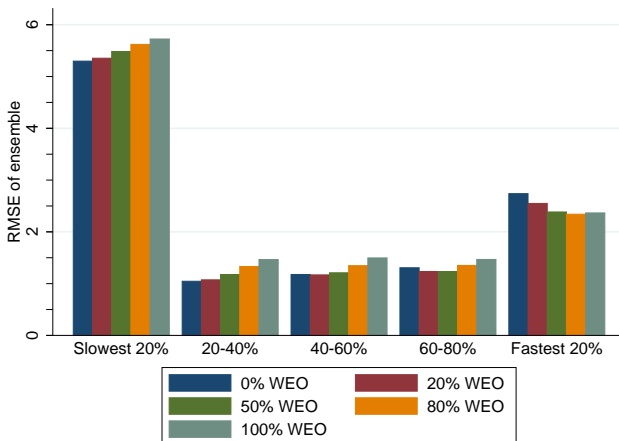




# 1-year forecasting gains: actual growth

Random Forest, Mean Absolute Error

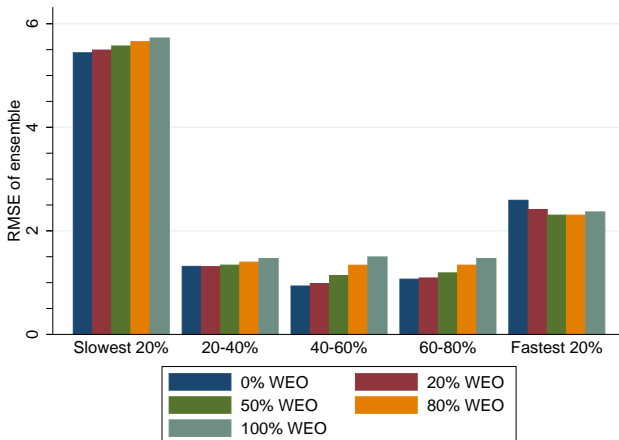
◀ Results: Growth



# 1-year forecasting gains: actual growth

Elastic Net, Mean Absolute Error

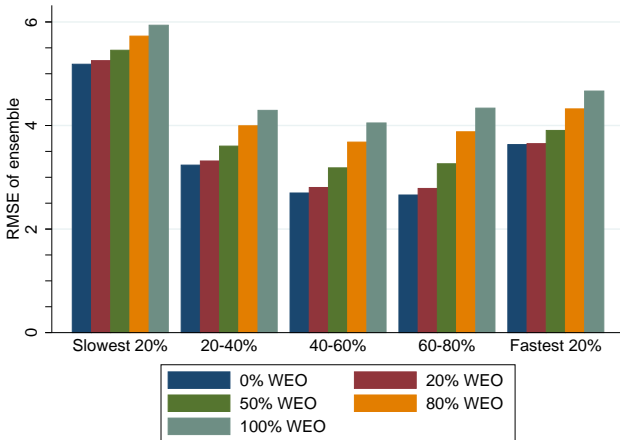
◀ Results: Growth





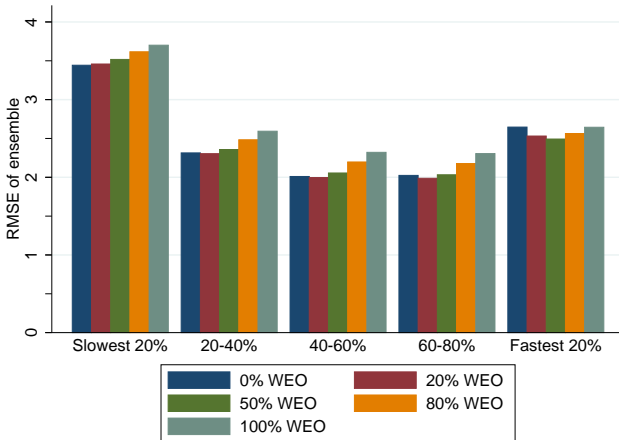
## Elastic Net, RMSE, 3-year

◀ Results: Growth



## Random Forest, MAE, 3-year

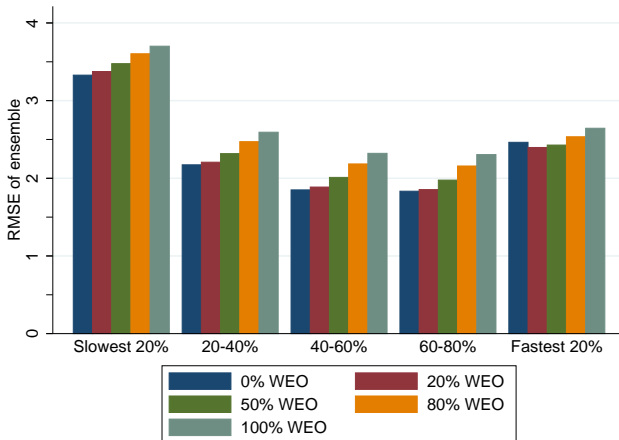
◀ Results: Growth



## Forecasting gains: actual growth

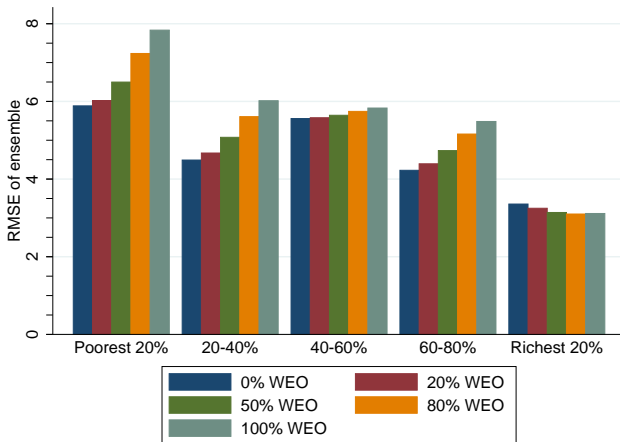
Elastic Net, Mean Absolute Error, 3-year

◀ Results: Growth



## Elastic Net, RMSE

◀ Results: Growth





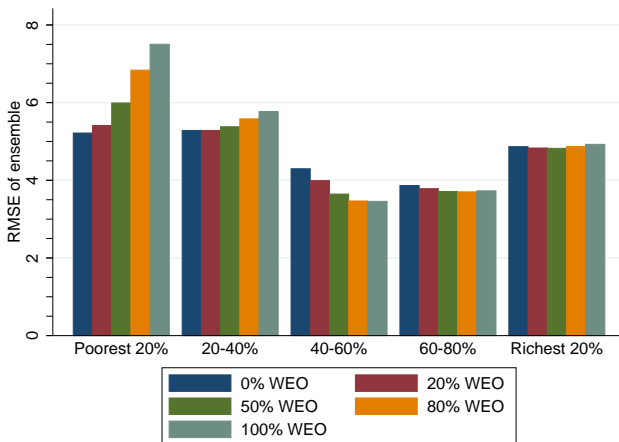




## Forecasting gains: initial GDP p.c.

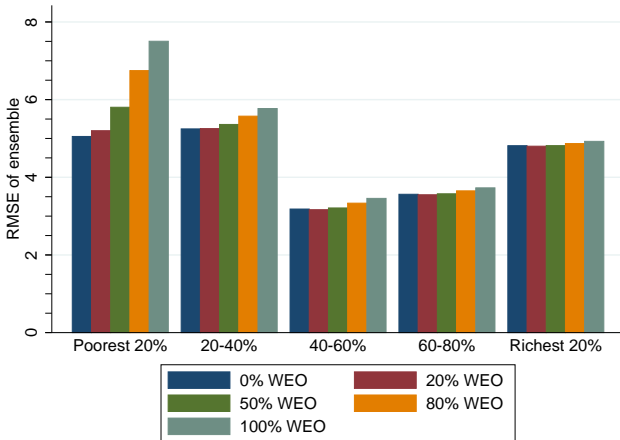
Random Forest, RMSE, 3-year

◀ Results: Growth



## Elastic Net, RMSE, 3-year

◀ Results: Growth



◀ Results: Growth

