BAYES BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

E-CREDA
Annual Conference 2024

Assessing the forecast performance of machine learning algorithms and econometric models in real estate

Sotiris Tsolacos & Yijun Wang
Bayes Business School

Background – use of ML in investment

- Rapid development of machine learning methods and ever rising use in investment analytics.

- Applications widespread in the investment universe
  - Stock-market - equity price forecasting
  - Cryptocurrencies
  - Exchange rates
  - And more …

- Deployment of ML algorithms in real estate pricing
  - Forecasting house prices (e.g. Sharma et al., (2024); Mora-Garcia et al., (2022)).
  - Valuations

- Fewer applications in commercial real estate.

Our research interest

- Assess the capacity of alternative ML algorithms for predicting commercial prices (yields, capital growth).
- We study whether the success of ML algorithms differs by sector.
- Econometric/time series forecasting models are also used to compare the forecasts with those obtained from ML methods.
- Also interested in gains from forecast combination (conventional models & ML algorithms).

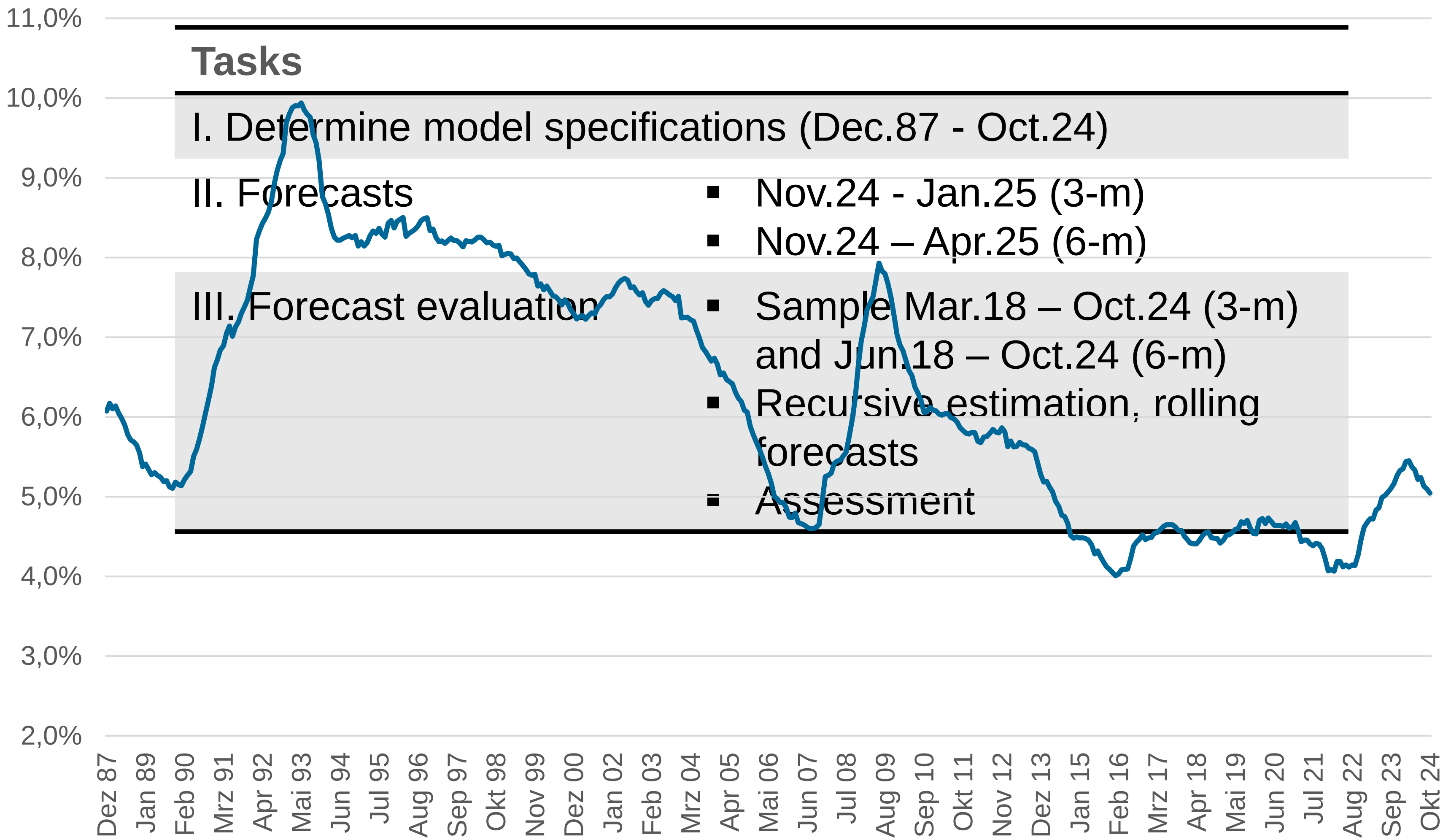
Background to forecasting

Cyclical influences

Long-term developments

Target series – UK office yields

(net initial yields)



Tasks

I. Determine model specifications (Dec.87 - Oct.24)

II. Forecasts
- Nov.24 - Jan.25 (3-m)
- Nov.24 – Apr.25 (6-m)

III. Forecast evaluation
- Sample Mar.18 – Oct.24 (3-m) and Jun.18 – Oct.24 (6-m)
- Recursive estimation, rolling forecasts
- Assessment

Source: MSCI

Methodologies

- ARIMA (Benchmark)
  - Uses the autocorrelated pattern of data along with the duration of shocks to predict; can include external information
- Regression model
- Elastic net linear regression models
  - Regularised models; objective is prediction
- Random Forest Model: Two specifications:
  - Internal determinants:
    - 3-month forecasts: Lag 3 to Lag 15
    - 6-month forecasts: Lag 6 to Lag 18
  - All determinants:
    - 3-month forecasts: Uses lag 3 of all variables as features.
    - 6-month forecasts: Uses lag 6 of all variables as features.

Data set: About 30 variables are considered containing real economy, monetary, financial, survey and real estate data series.

A key task – forecast evaluation

- Forecast evaluation – what is our objective?
  - Bias, dispersion (risk), direction, other?
- Real world (dynamic)
- Rolling samples

The basic academic approach

	ARIMAs	Regression	Elastic Net	RF-Int	RF-Ext
3-month forecasts					
ME	-0.008	-0.05	-0.06	0.01	-0.39
MAE	**0.08**	0.12	0.13	0.14	0.43
MSE	**0.011**	0.025	0.026	0.036	0.305
U1	0.01	0.02	0.02	0.02	0.05
Dir. Fore	55.7%	**73.8%**	54.5%	59.7%	46.8%
6-month forecasts					
ME	-0.02	-0.07	-0.08	-0.04	-0.35
MAE	**0.13**	**0.13**	0.21	0.30	0.47
MSE	**0.026**	0.028	0.059	0.152	0.331
U1	0.02	0.02	0.03	0.04	0.06
Dir. Fore	54.4%	**71.4%**	46.8%	50.6%	51.9%

ME: Mean error; MAE: Mean absolute error; MSE: Mean squared error; U1: Theil's U1 statistic; DirF: Success in predicting the direction of the yield movement three and six months ahead correctly.

Forecasts

Forecast made at end of November 2024			
	Office yield (%) Oct-2024	3-M Forecast (%) Jan-2025 [Dec-2025]	6-M Forecast (%) Apr-2025 [Mar-2025]
	5.04		
ARIMA		4.88	4.79
Regression model		5.0	4.83* (4.75)**
Elastic Net		[5.38]	[5.51]
RF-internal		[5.61]	[6.34]
RF-external		[5.19]	[4.93]

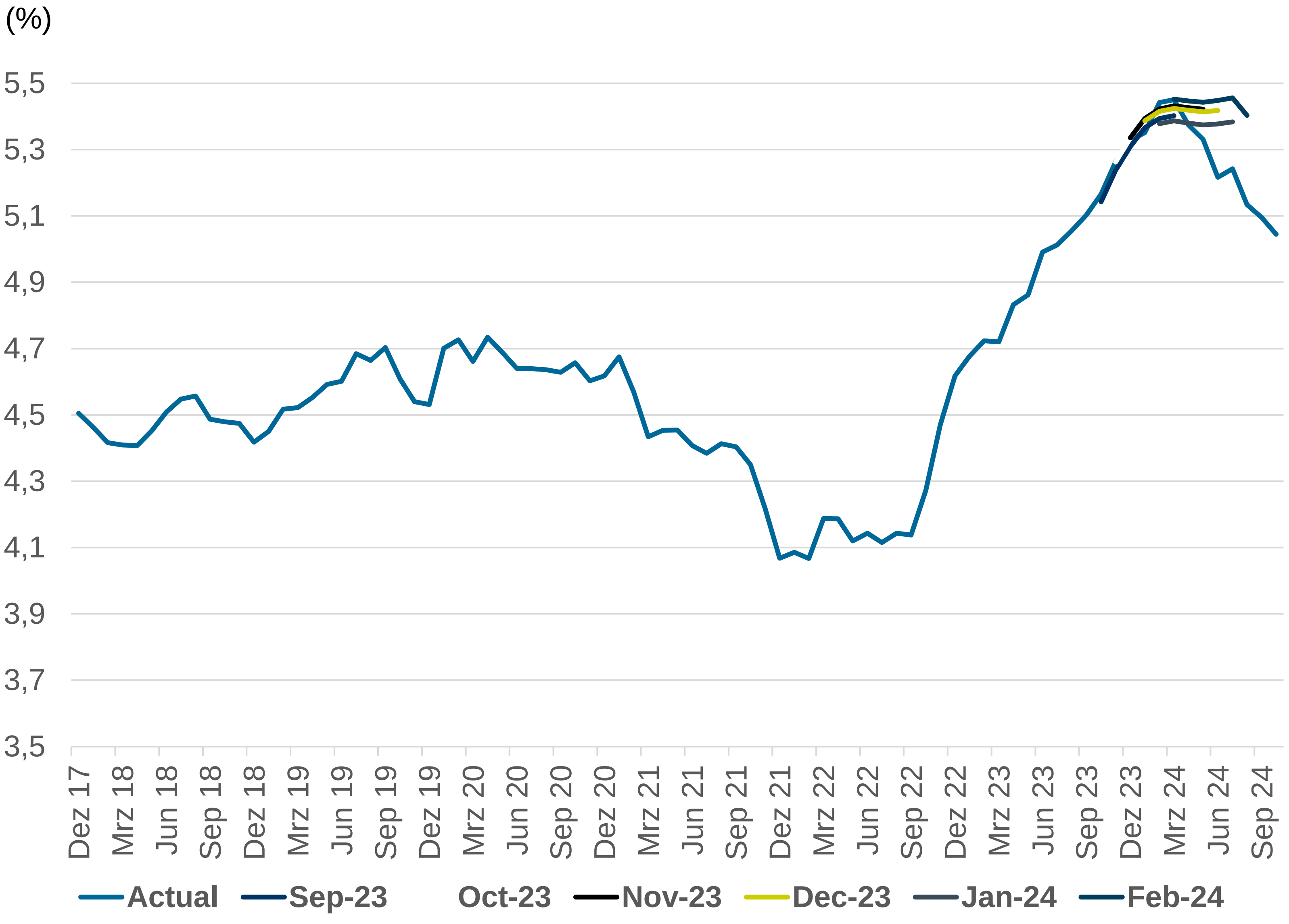
Notes:
- * Model excludes two employment variables
- ** Feb-2024 forecast (full model)

Key take-aways

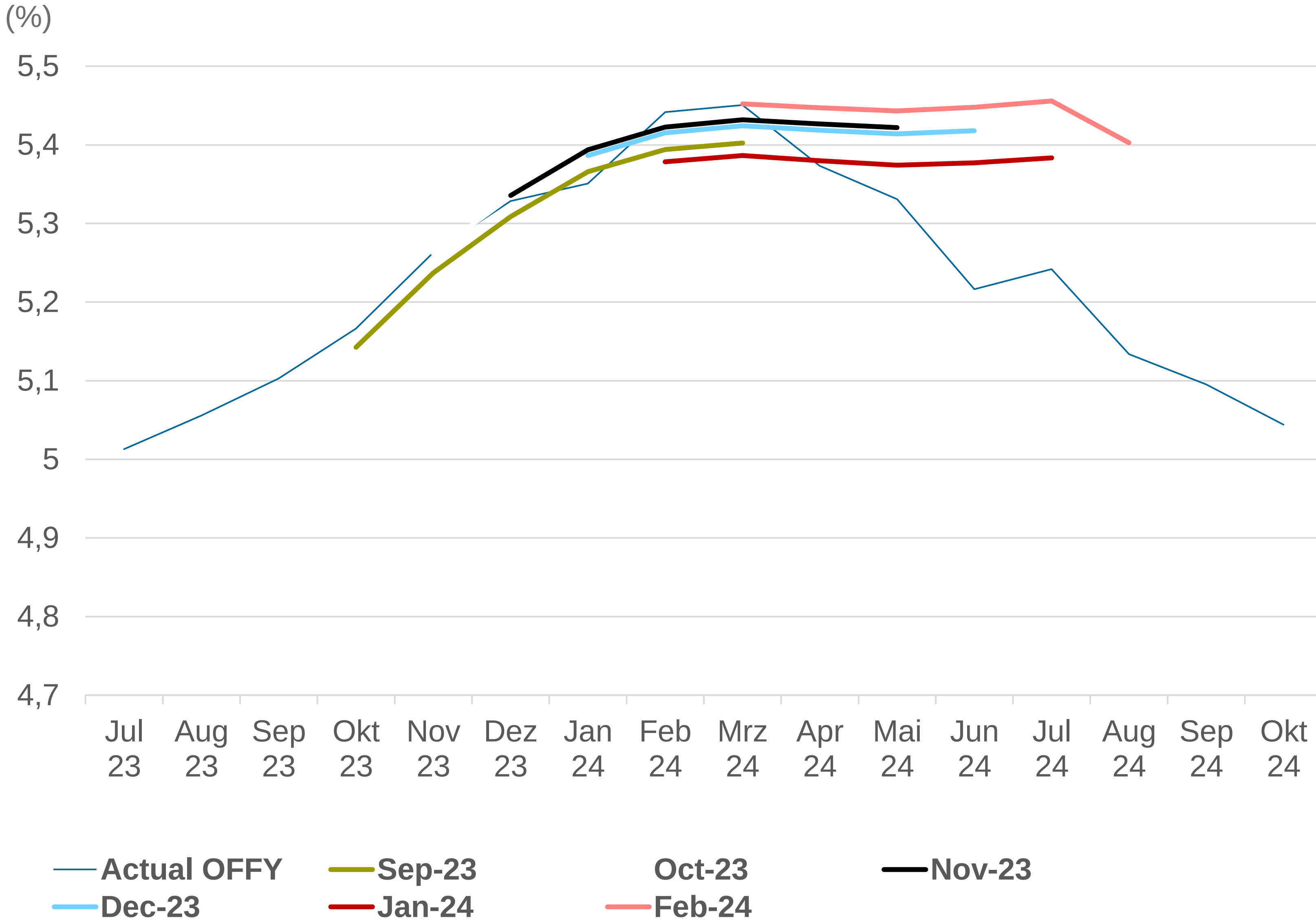
- Traditional techniques perform well
- Sample big enough to train ML algorithms?
- Must target a specific forecast objective

Turning point forecast

(6-month ahead, based on regression model)



Peak predicted, slow to indicate yield decline



Thank you