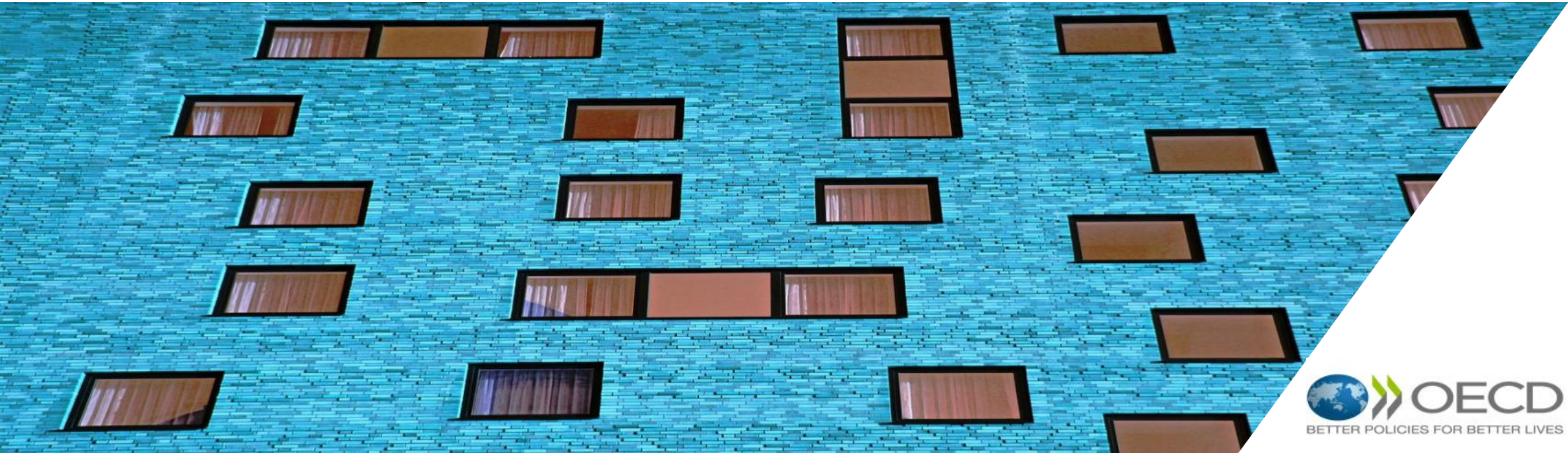
Assessing Economic Risks Around forecasts:

A mixture of fined-tuned BERT and economic experts

Manuel Betin, Thomas Chalaux, Marnix Dek, David Turner

BDF/OECD Innovation Lab conference, 13-06-2024

Leveraging Natural Language Processing (NLP) to answer economic questions



OECD

BETTER POLICIES FOR BETTER LIVES

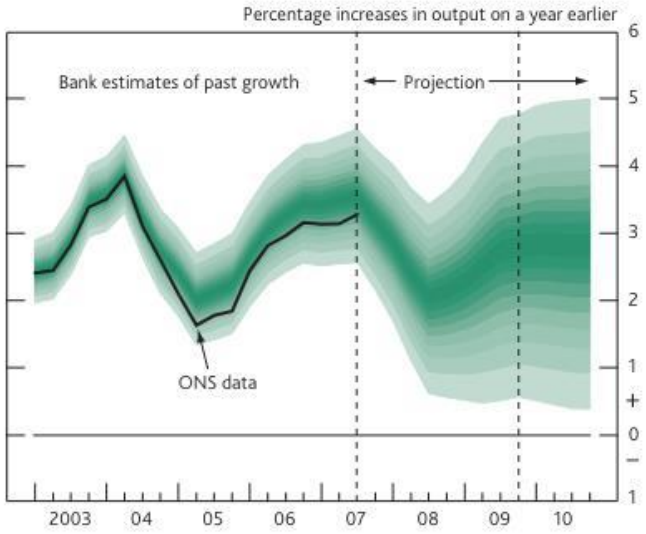
Outline

Motivation and Research Question

1. A Simple framework of Risks Around Forecasts (RAF)
2. Description of the Corpus
3. Text Estimation
4. Macroeconomic Estimation
5. Results and conclusions

Motivation

Chart 1 GDP projection based on market interest rate expectations



Erroneous symmetric fan charts

> *« … the MPC's adjustments to both the width and skew of the fan chart appear to have little or* ***no explicit grounding in data*** *or quantitative analysis … »*
> *Bernanke (2024)*

Accurate downside risk assessment

> ***The balance of risks around this central case is on the downside****. First, the global financial system remains vulnerable [...]. [...] availability of credit may tighten by more than assumed in the central projection [...]. [...] any significant decline in asset prices could materially weaken household and corporate balance sheets [....]*

Source: BoE Forecasts and Macroeconomic Risk Assessment (2007)

Research questions

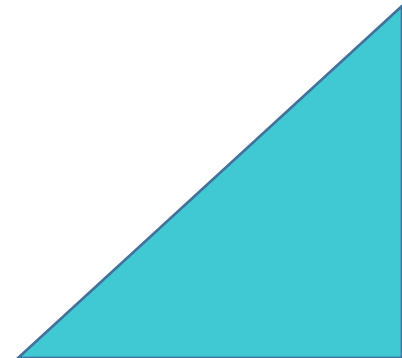
1. *Can **Risks Around Forecast** be estimated using the Economic Outlook of the OECD?*

2. *Are Risks Around Forecast **leading indicators** of future GDP growth?*

3. *Are Risks Around Forecast correlated with GDP growth **Forecast Errors**?*

4. *Can Risks Around Forecast be used to provide **Risk Adjusted GDP growth forecasts scenarios**?*

1. A Simple Risk Framework



A Simple Framework to measure Risk Around Forecast (RAF)

Z_{ct} = **Total** Risk around forecast of country c at time t

G_t = **Global** risk around forecast at time t

(1) $Z_{ct} = \alpha_c * G_t + I_{ct}$

I_t = **Idiosyncratic** risk around forecast in country c at time t

α_c = Exposure to global risk of country c at time t

G_t = Possible future macroeconomic developments in the **GLOBAL** that could affect negatively (positively) the growth forecast

I_{ct} = Possible future **SPECIFIC** macroeconomic developments in the country that could affect negatively (positively) the growth forecast

Estimating Risks Around Forecasts (RAF) from text

Use the text of the Economic Outlook published by the OECD to estimate $\boldsymbol{G_t}$ and $\boldsymbol{I_{ct}}$

(2) $G_t = \mathrm{f}(LLM(GA_t))$

where

***f(.)** is a weighted average of the risk score of each risk sentence*

***LLM(.)** is a Transformer based text classifiers*

GA= General Assessment of the Economic Outlook published at time t

(3) $I_{ct} = \mathrm{g}(LLM(CN_{ct}))$

where

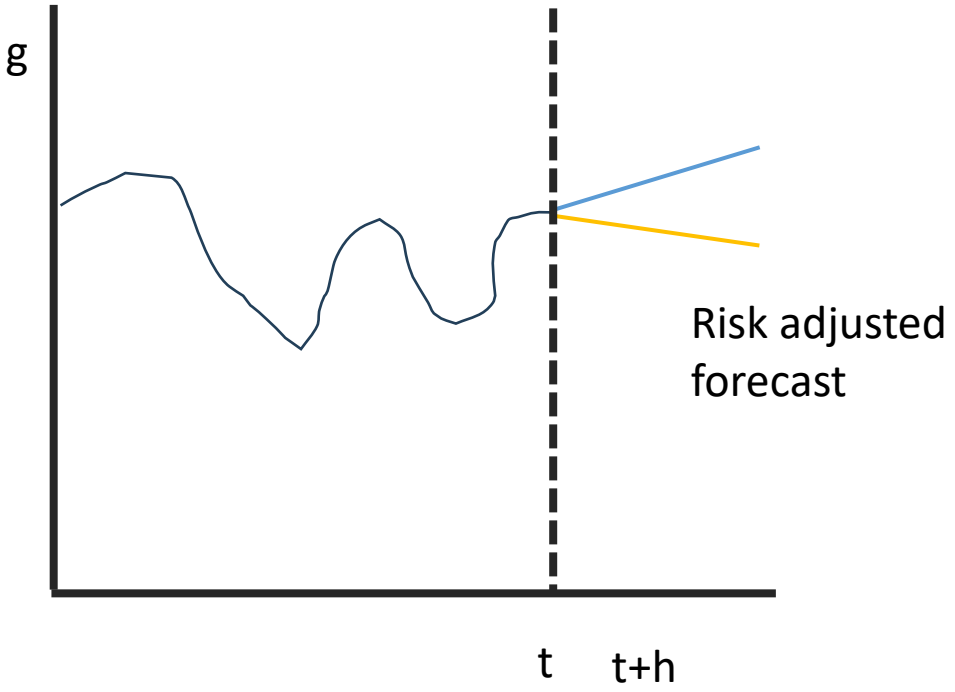
***g(.)** is a weighted average of the risk score of each risk sentence*

***LLM(.)** is a Transformer based text classifier*

CN= Country Note of country c in the Economic Outlook published at time t

A Simple Framework of Risk-Adjusted Growth Forecasts

(4) $\hat{g}_{ct+1}^{adj} = \hat{g}_{ct+1} + \beta_c * Z_{ct}$



Risk Adjusted Forecast

Baseline Forecast

Risk Adjustment Factor

Text Based Risk Score

A Simple Framework of Risk-Adjusted Growth Forecasts

$$\hat{g}_{ct+1}^{adj} = \hat{g}_{ct+1} + \beta_c * Z_{ct}$$

(4)+ (1)

$$\hat{g}_{ct+1}^{adj} = \hat{g}_{ct+1} + \beta_c * (\alpha_c * G_t + I_{ct})$$

(4)+ (1) + (2) + (3)

$$\hat{g}_{ct+1}^{adj} = \hat{g}_{ct+1} + \beta_c * (\alpha_c * f(LLM(GA_t)) + \beta_c\, g(LLM(CN_{ct}))$$

$$\hat{g}_{ct+1}^{adj} = \hat{g}_{ct+1} + \beta_c \alpha_c * f(LLM(GA_t)) + \beta_c\, g(LLM(CN_{ct}))$$

Risk Adjusted Forecast

Baseline Forecast

Global Risk from the General Assessment

Idiosyncratic Risk from Country Notes

Country + Global **Quantitative** (Model based) expertise

Global **Qualitative** expertise

Country specific **Qualitative** expertise

A Simple Framework of Risk-Adjusted Growth Forecasts

$$\hat{g}_{ct+1}^{adj} = \hat{g}_{ct+1} + \beta_c * Z_{ct}$$

$$\hat{g}_{ct+1}^{adj} = \hat{g}_{ct+1} + \beta_c * (\alpha_c * G_t + I_{ct})$$

$$\hat{g}_{ct+1}^{adj} = \hat{g}_{ct+1} + \beta_c * (\alpha_c * f(LLM(GA_t)) + \beta_c\, g(LLM(CN_{ct}))$$

$$\hat{g}_{ct+1}^{adj} = \hat{g}_{ct+1} + \beta_c \alpha_c * f(LLM(GA_t)) + \beta_c g(LLM(CN_{ct}))$$

G_t I_{ct}

Text estimation of ƒ(LLM(.)) and g(LLM(.))

A Simple Framework of Risk-Adjusted Growth Forecasts

$$\hat{g}_{ct+1}^{adj} = \hat{g}_{ct+1} + \beta_c * Z_{ct}$$

$$\hat{g}_{ct+1}^{adj} = \hat{g}_{ct+1} + \beta_c * (\alpha_c * G_t + I_{ct})$$

$$\hat{g}_{ct+1}^{adj} = \hat{g}_{ct+1} + \beta_c * (\alpha_c * f(LLM(GA_t)) + \beta_c\, g(LLM(CN_{ct}))$$

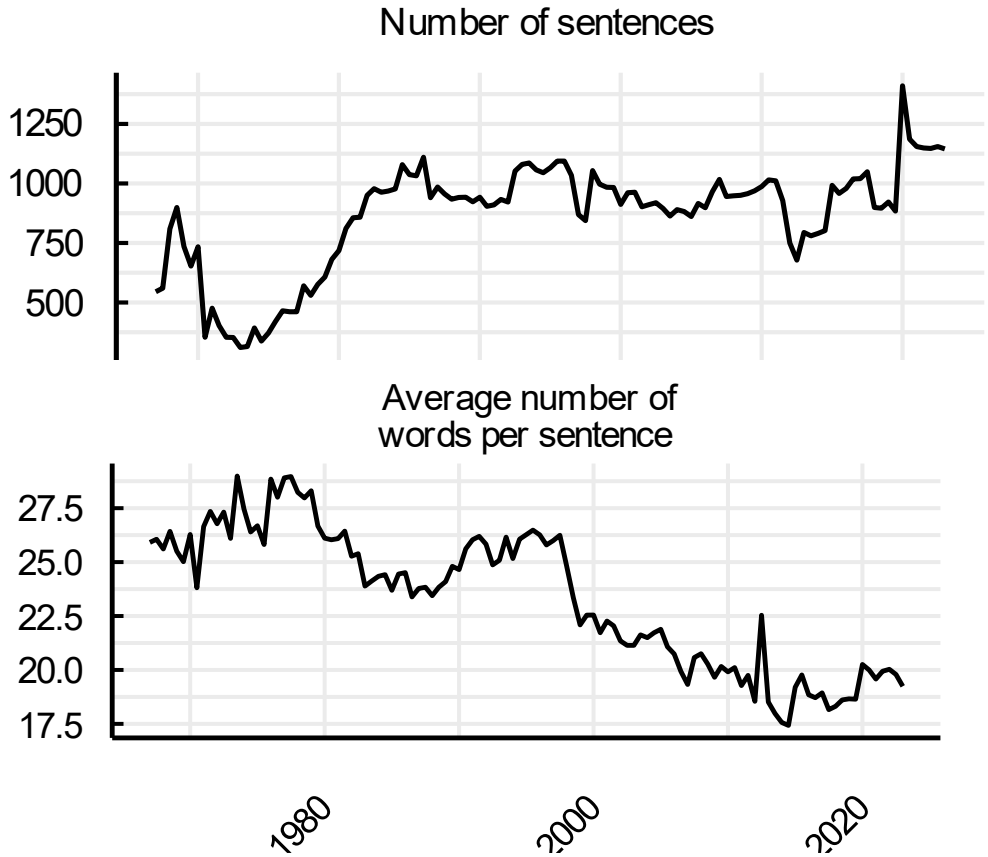
$$\hat{g}_{ct+1}^{adj} = \hat{g}_{ct+1} + \beta_c \alpha_c * f(LLM(GA_t)) + \beta_c\, g(LLM(CN_{ct}))$$

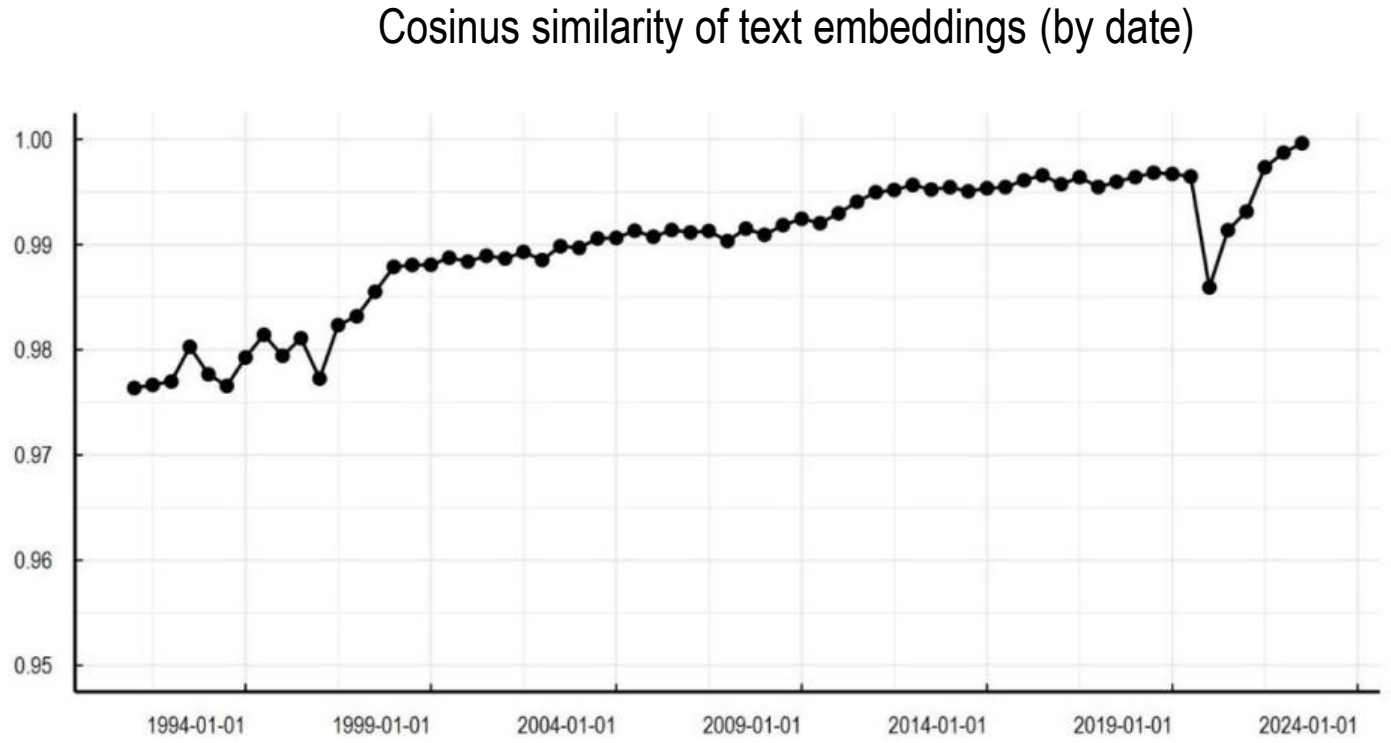
β_G^g β_I^g

Macroeconomic estimation

2. Description of corpus

Summary statistics of the text of the Economic Outlook (General Assessment (GA) + Country Notes (CN))



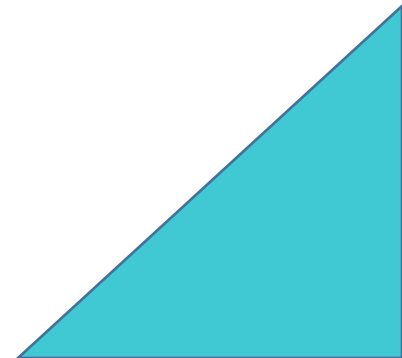


The format of the EO has changed overtime:

- Shorter sentences
- Less sentences per page (more figures)
- Less rich vocabulary

But the **text similarity** across period and countries is particularly high suggesting a homogeneous corpus across time and countries

2. Text Estimation



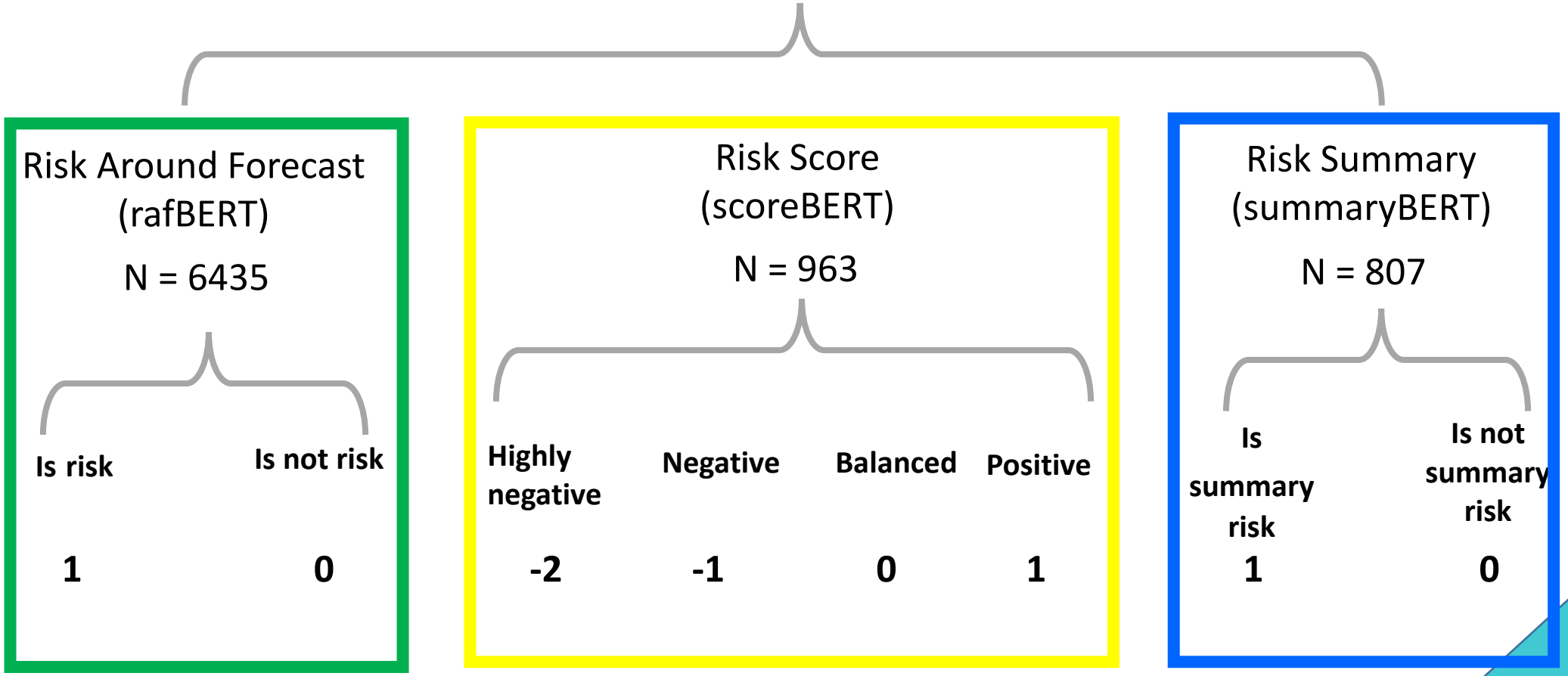
LLM(.) From text to Data

1. Compile/Create labeled **training datasets** (in-house)
2. **Fine-tune** existing foundation model for **text classification** (*Huggingface* + *Python* + *Transformer*)
3. **Predict** on each sentence
4. **Evaluate** the accuracy of the model (out-of-sample accuracy and benchmark against expert judgment)
5. **Aggregate** individual **risk sentences** into total **risk text : f(.) and g(.)**

Text classification

BERT base uncased

Finetuned text classifiers on each sentence of the EO



Validation out-sample

$$Accuracy = \frac{True\ positive\ (TP) + True\ negative\ (TN)}{sample\ size\ (N)}$$

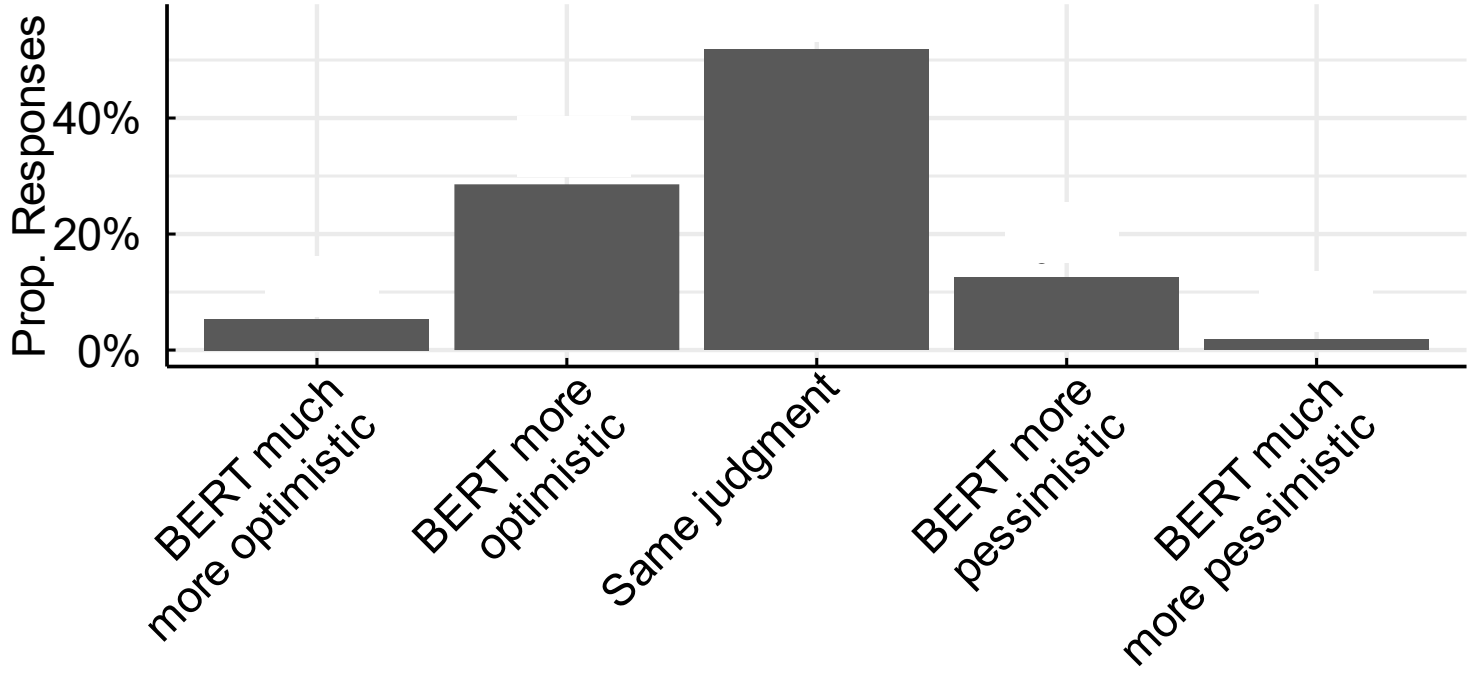
Table.3. Evaluation Metrics on test sample

Metric	rafBERT	scoreBERT
Loss	0.2282	0.6755
Accuracy	**0.9627**	**0.7526**

Out-of-sample prediction relatively satisfactory but could be improve with using **more capable LLMs** models or **increase the sample size for finetuning** …

… Although the scoring of risk does not make a complete **consensus among human experts**

Comparison BERT vs economic experts:
Confusion matrix



Findings from a survey on 50 experts of the Economic Department that scored selected risks paragraphes

- *BERT tend to be slightly more **optimistic** than experts but results depend on the final aggregation function (from sentence to paragraph)*
- ***Disagreement among experts** is high for positive risks and between very negative and negative.*
- *The **type of risk** (i.e financial) or country (EMEs) matters for the judgment of experts*

Estimated RAF sentences and score in Country Notes

Figure .X. Proportion of RAF Sentences in Country Notes

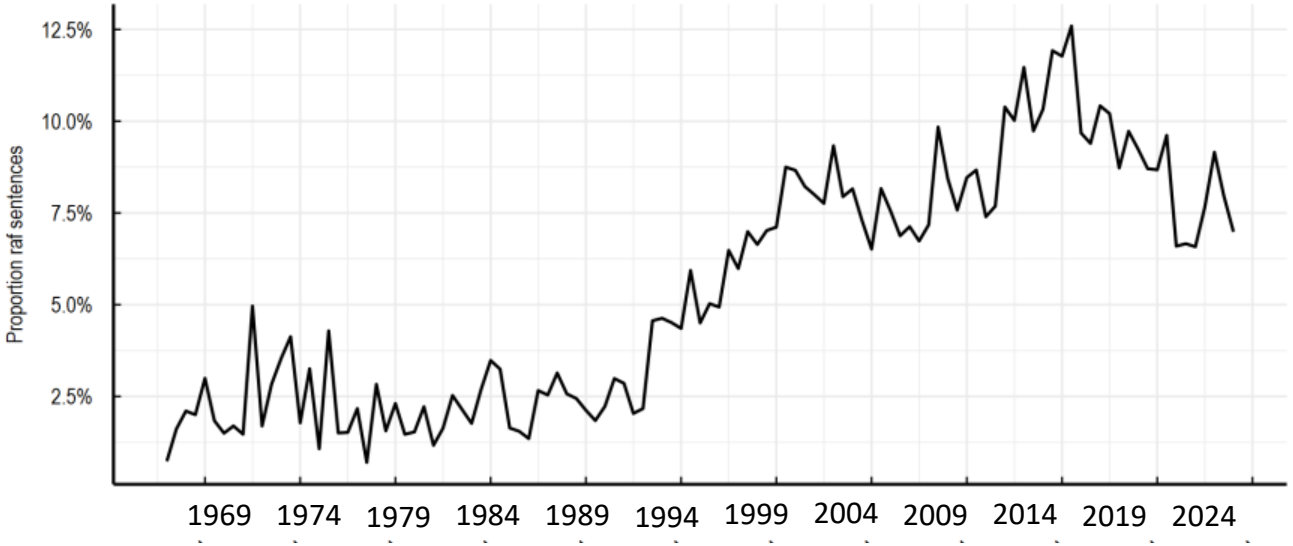
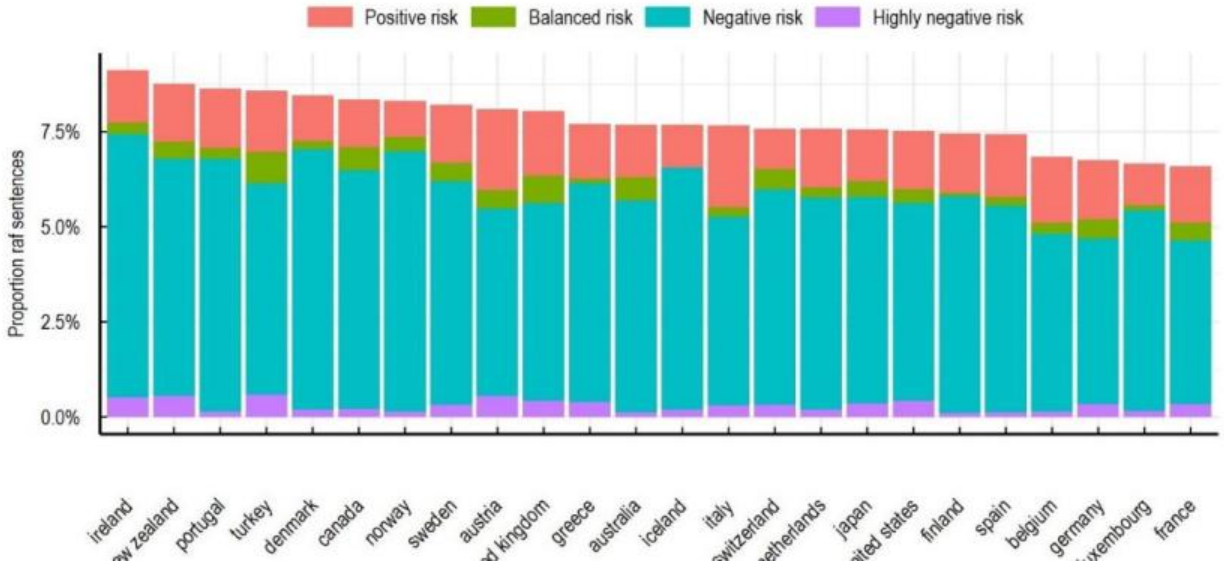


Figure .X. Proportion of RAF Sentences by Country and Score



The proportion of Country Notes allocated to risks has increased overtime (~ **5 sentences per CN**)

On average risks are mainly to the **downside** but cover both positive and negative risks.

Calibrating the aggregation function: g(.) and f(.)
from RAF sentence to RAF paragraphs

$I_t = \boldsymbol{g}(LLM(CN))$ ⇨ *Weighted average of the risk score of each risk sentence*

$G_t = \boldsymbol{f}(LLM(GA))$ ⇨ *Average of the risk score of each risk sentence*

4. Macroeconomic estimation

Specifications

(2) $g_{ct} = \propto_1 g_{ct-1} + \beta_G^g G_{t-1} + \boldsymbol{\beta_I^g} I_{ct-1} + \theta_{1c} + \varepsilon_{1ct}$ ➡ Is RAF index one semester ahead and **Real GDP growth** yoy correlated?

(4) $R_{ct} = \propto_2 g_{ct-1} + \beta_G^R G_{t-1} + \boldsymbol{\beta_I^R} I_{ct-1} + \theta_{2c} + \varepsilon_{2ct}$ ➡ Is RAF index one semester ahead and **Soft Recessions** correlated?

(6) $C_{ct} = \propto_3 g_{ct-1} + \beta_G^C G_{t-1} + \boldsymbol{\beta_I^C} I_{ct-1} + \theta_{3c} + \varepsilon_{3ct}$ ➡ Is RAF index one semester ahead and **Severe Recession** correlated?

$\boldsymbol{g_{ct}}$ = *Real gdp growth rate*
$\boldsymbol{R_{ct}}$ = *Dummy equal to 1 if GDP growth (yoy) between -2 and 0 %*
$\boldsymbol{C_{ct}}$ = *Dummy equal to 1 if GDP growth (yoy) below -2 %*

$\boldsymbol{G_t}$ = *Global Risk Around Forecast index*
$\boldsymbol{I_{ct}}$ = *Idiosyncratic Risk Around Forecast index*
$\boldsymbol{\delta_t}$ = *Time fixed effect*
$\boldsymbol{\theta_c}$ = *Country fixed effect*

$\forall\, i \in \{G, I\}$

$\boldsymbol{\beta_i^g} > 0$ ➡ Downside risk correlated to **economic slowdown**

$\boldsymbol{\beta_i^R} < 0$ ➡ Downside risk correlated with higher probablity of **soft recession**

$\boldsymbol{\beta_i^C} < 0$ ➡ Downside risk correlated with higher probability of **severe recession**

Estimation:

Is RAF correlated with the business cycle?

	Dependent variable:					
	GDPV yoy		Soft crisis (1 if -2<g<0)		Severe crisis (1 if g< -2)	
	(1)	(2)	(3)	(4)	(5)	(6)
lag1_GDPV_yoy	0.616***	0.575***	-0.048***	-0.050***	-0.032***	-0.033***
	(0.015)	(0.016)	(0.002)	(0.002)	(0.001)	(0.002)
Lag1_score	0.264**	0.342***	-0.045***	-0.034**	-0.040***	-0.031***
	(0.115)	(0.118)	(0.014)	(0.014)	(0.011)	(0.011)
Lag1_score_GA	0.543***	0.565***	-0.038***	-0.035***	-0.014	-0.008
	(0.106)	(0.107)	(0.013)	(0.013)	(0.010)	(0.010)
Constant	1.519***		0.206***		0.120***	
	(0.118)		(0.014)		(0.011)	
Time FE	No	Yes	No	Yes	No	Yes
Country FE	No	Yes	No	Yes	No	Yes
Observations	2,455	2,455	2,455	2,455	2,455	2,455
R2	0.406	0.360	0.221	0.223	0.169	0.174
Adjusted R2	0.406	0.348	0.220	0.208	0.168	0.158
F Statistic	559.158*** (df = 3; 2451)	452.370*** (df = 3; 2408)	231.115*** (df = 3; 2451)	229.894*** (df = 3; 2408)	166.668*** (df = 3; 2451)	169.094*** (df = 3; 2408)
Note:						*p<0.1; **p<0.05; ***p<0.01

β_I^g β_G^g β_I^R β_G^R β_I^c β_G^c

Both global and Idiosyncratic risks are correlated with e**conomic slowdowns** in next semester

Both global and Idiosyncratic risks are correlated with **SOFT recession** in next semester

Only **Idiosyncratic** risks are correlated with **SEVERE** recessions in next semester

Estimation:
Can RAF index reduce forecast errors ?

Step 1: $g_{ct} = \gamma * \widehat{g_{ct}} + \delta_t + \theta_c + \varepsilon_{ct}$

Step 2: $\varepsilon_{ct} = \beta_G^{\varepsilon} G_{t-1} + \boldsymbol{\beta_I^{\varepsilon}} I_{ct-1} + \mu_{ct}$

g_{ct} = *Realized real GDP growth (yoy)*
$\widehat{g_{ct}}$ = *1 year ahead forecast*
ε_{ct}= forecast error

	Dependent variable:	
	ε_{ct}	
	(1)	(2)
Lag1_score	0.422***	0.477***
	(0.132)	(0.134)
Lag1_score_GA	0.780***	0.786***
	(0.116)	(0.116)
Constant	0.293**	
	(0.127)	
Time FE	No	Yes
Country FE	No	Yes
Observations	1,935	1,935
R2	0.030	0.032
Adjusted R2	0.029	0.013
F Statistic	29.768*** (df = 2; 1932)	31.659*** (df = 2; 1897)
Note:		*p<0.1; **p<0.05; ***p<0.01

$\boldsymbol{\beta_I^{\varepsilon}}$ β_G^{ε}

Significantly **correlated** but **very modest** explanatory power

An example of *Risk Adjusted Forecasts*

Estimated ⟶ $\beta_I^g = 0.34$ $\beta_G^g = 0.56$

GDP Growth forecast (yoy) = **1.5%**

OECD Economic Outlook EO115

General Assessment (GA)

Text

Counyty notes (CN)

Text

RAF algorithm

G_t

I_{ct}

Positive Risk ⟶ $g_{ct}^{adj} = 1.5\% + 0.34 * (1) + 0.56 * (1)$

$g_{ct}^{adj} = 2.4\,\%$

Balanced Risk ⟶ $g_{ct}^{adj} = 1.5\% + 0.34 * (0) + 0.56 * (0)$

$g_{ct}^{adj} = 1.5\,\%$

Downside Risk ⟶ $g_{ct}^{adj} = 1.5\% + 0.34 * (-1) + 0.56 * (-1)$

$g_{ct}^{adj} = 0.59\,\%$

Severe Downside Risk ⟶ $g_{ct}^{adj} = 1.5\% + 0.34 * (-2) + 0.56 * (-2)$

$g_{ct}^{adj} = -0.3\,\%$

Conclusion

1. Economic Publications are very rich text material for economic analysis

2. Recent AI developments has democratised advanced NLP tools for classification/extraction/search etc..

3. Combining expert knowledge and LLM capabilities can help ground a scenario-based forecast approach.

4. The switching cost to a more advanced model is straightforward but improvements probably will not come from just having a more powerful LLM but from adjusting for **economic relevant preferences/nuances**