

Sports Economics: Betting Market Efficiency & Player Labor Market

Dr. Sam Hoey
sportecon@ese.eur.nl

Today (and previous lecture)

Questions:

1. What forms of biases are there in sports betting?
2. How much do athletes earn and clubs spend on wages?

→ I will use some mentimeter quizzes throughout the lecture(s).

A stylized, handwritten-style logo of the word "Erasmus" in a dark blue or black color, located in the bottom right corner of the slide.

Readings

- Academic papers
 - Angelini, G., & De Angelis, L. (2019). Efficiency of online football betting markets. *International Journal of Forecasting*, 35(2), 712-721.
 - Winkelmann, D., Ötting, M., Deutscher, C., & Makarewicz, T. (2024). Are betting markets inefficient? Evidence from simulations and real data. *Journal of Sports Economics*, 25(1), 54-97.

Find them in the literature section in canvas!

A stylized, handwritten-style logo of the word "Erasmus" in a dark blue or black color, located in the bottom right corner of the slide.

Recap

Erasmus

What we have done so far - modelling

1. We have looked at the different methods for modelling match outcomes

- a) ELO Ratings (Arntzen & Hvattum, 2010)
- b) Past performance (Dixon and Coles, 1997)
- c) Player ratings (Holmes and McHale, 2024)
- d) Market values, wisdom of the crowds (Peeters, 2018)

→ All of them model differences in team strength in different ways

2. Turning Parameters and Hyperparameters

- a) Parameters: trained within model
- b) Hyperparameters: chosen by researcher/econometrician → tuned using optimization method
- c) Metrics: Brier Score, Accuracy (and betting performance)



What we have done so far – bookmaker odds

What are odds? (decimal format)

- Decimal odds o : your **total return per \$1 staked** if the outcome wins.
Example: $o = 2.40 \rightarrow$ win returns \$2.40 (profit \$1.40).

Implied probability

- $p_k^{book} = \frac{1}{odds_k}$ for outcome k (home/draw/away)
- For a fair market, $p_h^{book} + p_d^{book} + p_a^{book} = 1$

Bookmaker margin (overround)

- $R = \sum_k p_k^{book} = p_h^{book} + p_d^{book} + p_a^{book}$ (typically $R > 1$)
- $(R - 1)$ is the margin spread across outcomes (or the vig)

Converting odds (removing margin)

$$p_k^{fair} = \frac{p_k^{book}}{R}$$

Example:

Odds: Home **2.00**, Draw **3.50**, Away **3.70**

Implied probabilities: 0.500 (home) , 0.286 (draw) , 0.270 (away)

Overround: $R = 0.500 + 0.286 + 0.270 = 1.056$ ($\approx 5.6\%$ margin)

Fair probabilities:

$$p_{home}^{fair} = 0.500/1.056 \approx 0.474$$

$$p_{draw}^{fair} = 0.286/1.056 \approx 0.270$$

$$p_{away}^{fair} = 0.270/1.056 \approx 0.256$$

A stylized, handwritten-style logo for Erasmus, featuring a large, flowing 'E' followed by the word 'Erasmus' in a cursive script.

What we have done so far – Expected value (EV) betting

What is EV?

- With decimal odds o and your model probability p :
- $EV = p \cdot o - 1$ is the expected profit per \$1 staked.
 - $EV > 0 \Rightarrow$ positive edge (profitable in expectation)
 - $EV = 0 \Rightarrow$ break-even
 - $EV < 0 \Rightarrow$ negative edge
- Crucial: EV hinges on your p being the true probability

Selection rule (typical):

- Bet outcome k only if $EV_k \geq \tau$ (in principle no limits to τ , but typical is from 0% to 5%)
- You can even tune τ in your training/tuning sample to max profits.

Staking (in academic papers)

- Bet **unit stake/flat stake** on most likely outcome when EV threshold is met (default in many papers)
 - Max one bet per match
- Kelly criterion
 - Let $b = o - 1$ (o = bookmaker odds), p = model probability and $q = 1 - p$
 - Kelly fraction: $f^* = \frac{bp - q}{b}$
 - Modified application (Holmes & McHale, 2024): tune EV threshold \rightarrow if met, stake 1 unit $\times f^*$ (no bankroll)
 - Can place multiple bets per match.



Biases in Betting



Erasmus

Efficient Markets in Betting Odds

EMH form (Fama, 1970)	Finance definition	Betting-market analogue	What you'd expect if it holds	Simple test
Weak-form	Prices reflect all past returns/prices.	Odds already embed historical results, streaks, H2H, past odds.	No edge from using past scores/odds alone.	Show a model using only lagged outcomes/odds; out-of-sample ROI ≈ 0 .
Semi-strong	Prices reflect all public info.	Odds incorporate public news (injuries, lineups, weather, ratings, travel).	Public-data models don't beat the odds after margin.	Calibration regression (more later) Brier of odds \approx best achievable.
Strong-form	Prices reflect all info (even private).	Even inside info already in odds.	No edge for anyone, ever.	Not realistically testable; look for abnormal wins by "insiders" (rare/impractical).



Favorite-Longshot Bias (FLB)

FLB says markets/prices tend to be **too pessimistic about favorites** (team likely to win) and **too optimistic about longshots** (team unlikely to win) when you convert odds into probabilities.

- For **longshots**, the **implied probability** from the odds is often **higher than the true frequency** → they look better than they are.
- For **favorites**, the implied probability is **lower than the true frequency** → they look worse than they are.

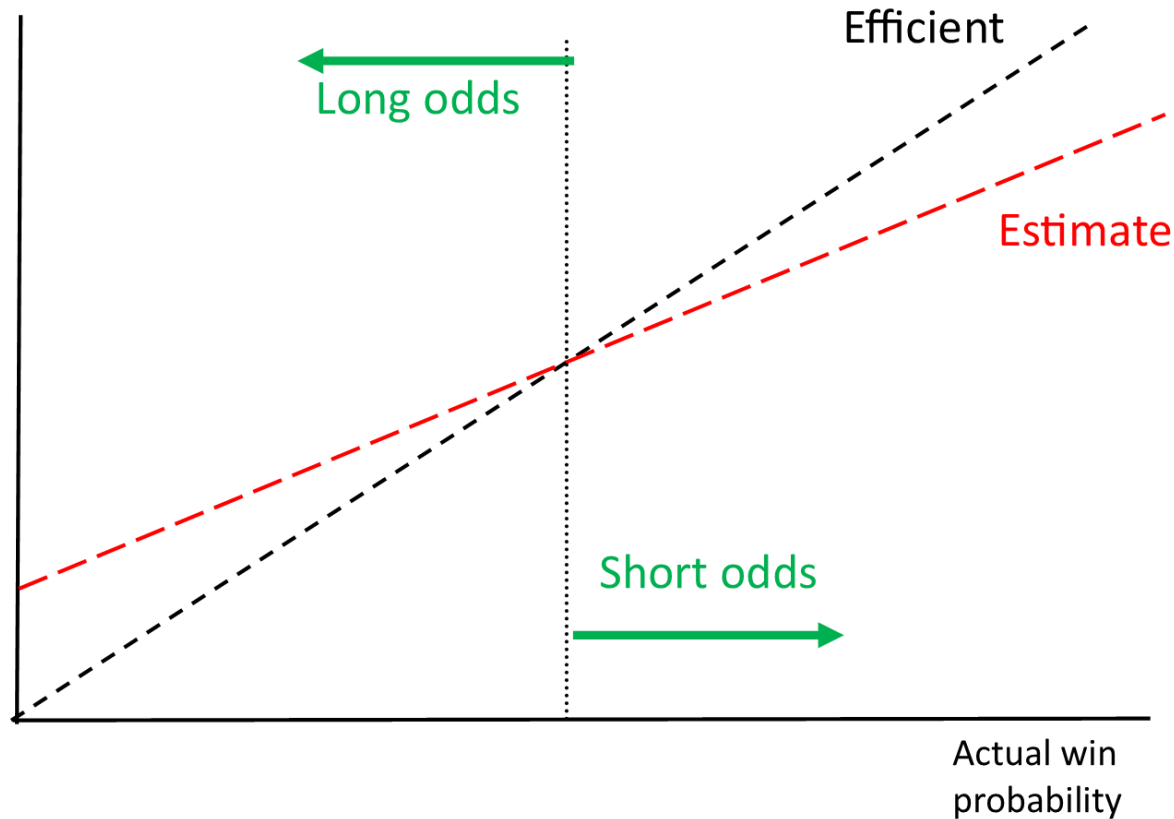
Explanations for FLB:

1. **Bookmakers** want to avoid large payouts if underdog (longshot) wins → adjust odds for longshot → adjust odds for favorite to keep balanced book.
2. **Risk loving bettors:** many bettors enjoy “lottery tickets”. They accept low EV bets for the thrill of a big win
3. **Overconfidence:** bettors overestimate likelihood of unlikely event → bookmakers exploit this and adjust odds to increase profits
4. **Behavioral economics:**
 - a) People tend to overestimate small probabilities
 - b) High probabilities are treated as certain



Favorite-Longshot Bias (FLB)

Expected win probability based
on bookmaker odds



For low probabilities:

- Lower odds
- Higher implied probability

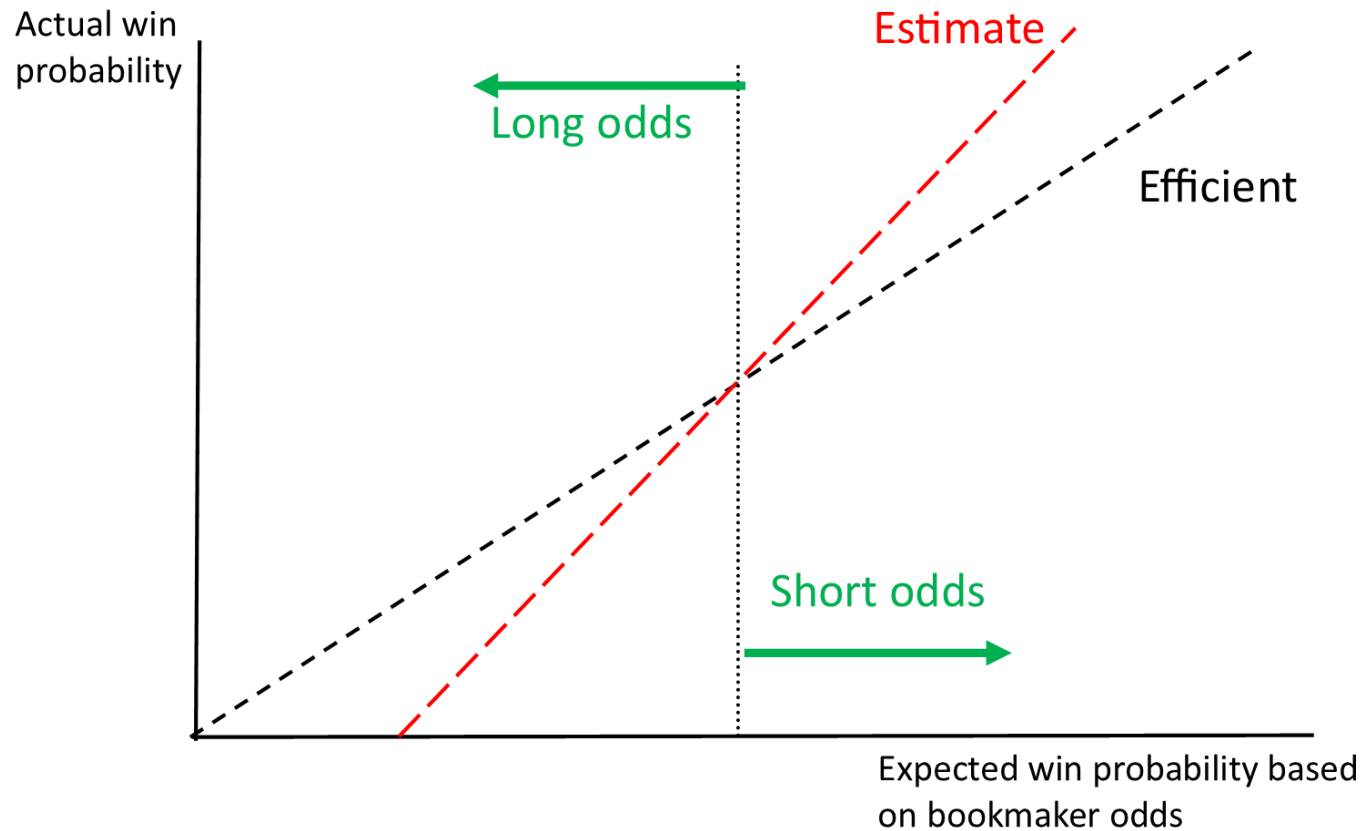
For high probabilities:

- Higher odds
- Lower implied probability

Long odds = low probability event
Short odds = high probability event

Ezra

Empirical Studies: Favorite-Longshot Bias (FLB)



- $Y_i = \alpha + \beta * p_i + v_i$
- Y_i = actual outcome;
- p_i = predicted outcome (bookmaker odds)
- Efficient market: $\alpha = 0$ & $\beta = 1$
- Longshot bias: $\alpha < 0$ & $\beta > 1$

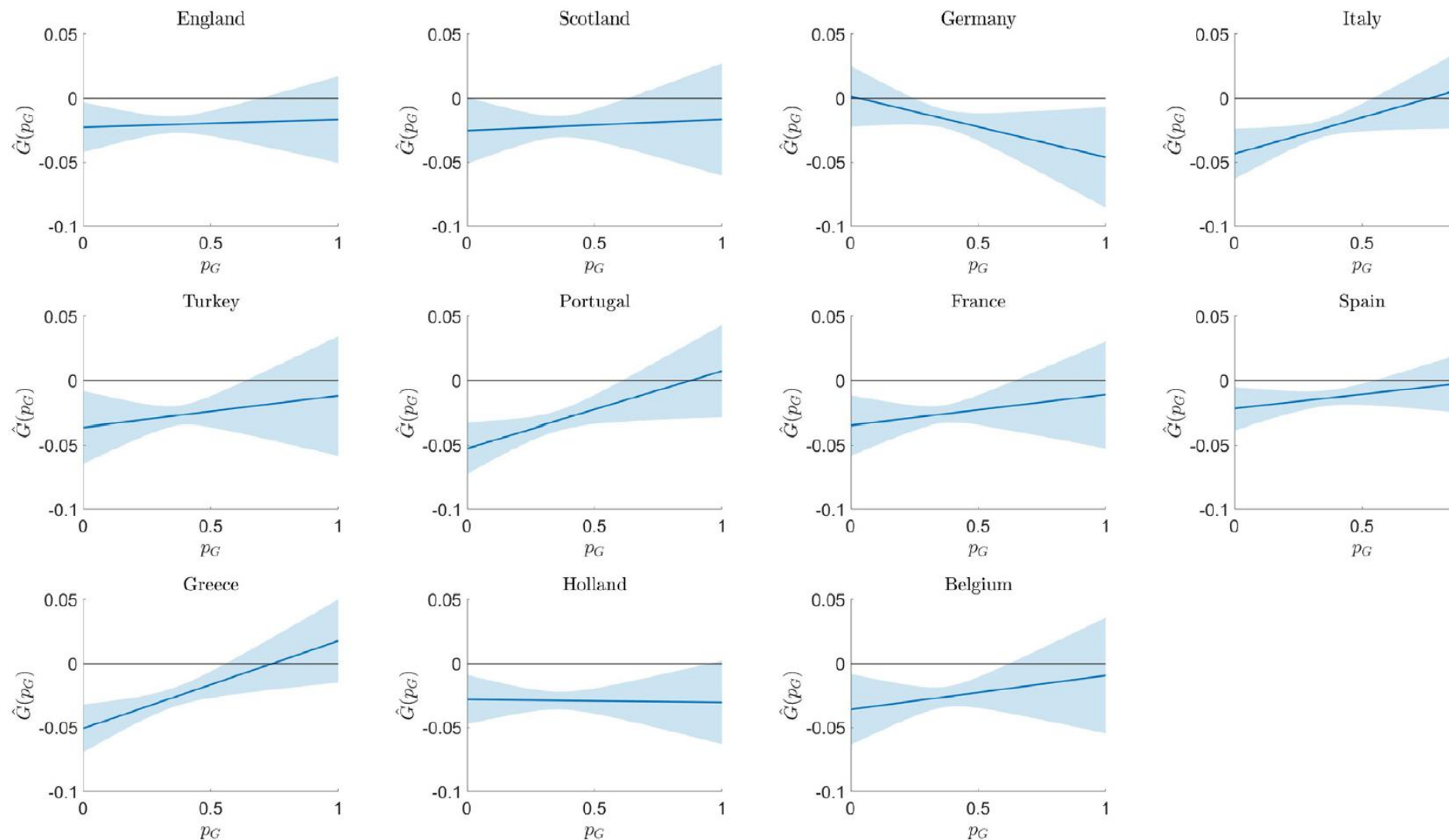
Ezra

Angelini & De Angelis (2019)

- Looks at European football online betting markets: 41 international bookmakers on 11 leagues over 11 years.
- As odds have limited predictive power for draws, they drop them from the sample and focus on home/away wins.
- Set-up:
 - y_i is an indicator for a particular match outcome i and is equal to 1 if the match outcome is realized (0 otherwise)
 - o_i are the decimal odds
 - $p_{Gi} = 1/o_i$ is the corresponding implied probability
 - No correction for bookmaker margin
- $G_i = y_i - p_{Gi} \rightarrow$ bookmaker forecast error for match outcome i
- Estimate the following model:
- $G_i = \alpha + \beta * p_{Gi} + v_i$
 - α captures bookmaker commission (expected to be < 0) and β evaluates market efficiency
 - $\beta = 0 \rightarrow$ market is unbiased
 - $\beta > 0 \rightarrow$ market characterized by favorite-longshot bias: forecast error increases with forecasted probability



Angelini & De Angelis (2019)



afm

Angelini & De Angelis (2019)

Main conclusions:

- α significantly smaller than 0 in all but one country (Germany)
- β significantly larger than 1 in three countries: Italy, Portugal & Greece → favorite-longshot bias
- Germany → reverse FLB
- All efficiency curves are below zero line, except for largest values of p_G for Italy, Portugal and Greece
- Confidence bands: no significant positive values can be achieved
- This implies bettors cannot systematically achieve positive returns.
- Bookmakers profit especially from longshots

Bottom line: Odds show a margin and some FLB-style tilts, but **no statistically robust way to earn positive returns.**

The **house especially profits on longshots**, where miscalibration and margin bite hardest.



Home bias and sentiment bias

What they are

- **Home bias:** Odds (or bettors) systematically **overvalue the home team** after controlling for fundamentals → home payoffs worse than they “should” be.
- **Sentiment bias:** Odds tilt toward **popular/attention-grabbing teams/players** (brand, media buzz), independent of win probability.

Why it can happen

- Crowd effects in stadiums, referee/psychology, travel fatigue (home advantage) **over-translated into prices.**
- Fans bet **for** their team; media hype shifts demand; books shade odds to balance books (limit liabilities).

Calibration regression:

$$y = \alpha + \beta * p_{book}^{fair} + \gamma * H + \delta * S + \epsilon$$

Testable predictions

- $\gamma = 0$ no differential odd pricing/bias for the home team
- $\delta = 0$ no differential odd pricing/bias for the fan favorite

A stylized, handwritten-style logo for Erasmus, featuring a large, flowing 'E' followed by the word 'Erasmus' in a cursive script.

Winkelmann et al. (2024)

Specification:

$$y_i = \beta_0 + \beta_1 * p_i + \beta_h * H_i + \beta_s * \Delta Attend_i + \epsilon_i$$

- y_i is an indicator for a win
- p_i represents the win probability based on bookmaker odds
- H_i is a dummy variable for a bet on the home team
- $\Delta Attend_i$ is the difference in mean attendance between the two opponents (during the current season) intended as an indicator for the sentiment of the gamblers.

A stylized, handwritten-style logo of the word "Erasmus" in a dark blue or black color, located in the bottom right corner of the slide.

Winkelmann et al. (2024)

Table 9. Estimation Results for the Regression Model Fitted to Data From the English Premier League.

	Dependent variable:														
	Won														
	All seasons	2005-2006	2006-2007	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016	2016-2017	2017-2018	2018-2019
Implied probability	5.004*** (0.222)	6.008*** (0.887)	4.847*** (0.917)	7.012*** (0.889)	5.177*** (0.887)	4.029*** (1.021)	2.726*** (0.931)	2.784*** (0.900)	5.159*** (0.926)	5.548*** (0.842)	5.217*** (0.818)	4.201*** (0.789)	6.028*** (0.823)	4.569*** (0.735)	5.155*** (0.680)
Home	0.136** (0.063)	0.122 (0.256)	0.305 (0.245)	-0.098 (0.245)	0.020 (0.246)	0.792*** (0.258)	0.681*** (0.252)	0.243 (0.249)	-0.020 (0.257)	-0.089 (0.241)	-0.041 (0.235)	-0.101 (0.221)	0.254 (0.229)	0.250 (0.227)	0.024 (0.233)
Diff Attend	0.002 (0.002)	-0.001 (0.008)	0.003 (0.007)	-0.002 (0.007)	0.006 (0.007)	0.013 (0.010)	0.011 (0.007)	0.018** (0.008)	0.008 (0.008)	-0.001 (0.007)	-0.002 (0.006)	-0.001 (0.006)	-0.004 (0.006)	0.003 (0.006)	0.005 (0.006)
Constant	-2.529*** (0.075)	-2.758*** (0.275)	-2.557*** (0.297)	-3.226*** (0.298)	-2.540*** (0.283)	-2.560*** (0.334)	-2.044*** (0.301)	-1.743*** (0.288)	-2.621*** (0.303)	-2.527*** (0.285)	-2.497*** (0.280)	-2.135*** (0.274)	-2.957*** (0.298)	-2.493*** (0.264)	-2.430*** (0.244)
Observations	10,640	760	760	760	760	760	760	760	760	760	760	760	760	760	760

Note: Values in parentheses are robust standard errors.

* $p < .1$; ** $p < .05$; *** $p < .01$.

We have:

1. Favorite-longshot bias (implied probability coef.)
2. Home bias in pooled regression (Home coef.)
3. No sentiment bias (Diff Attend coef.)

Ezra

In class exercise

Testing for FLB in the NHL data

Ezra

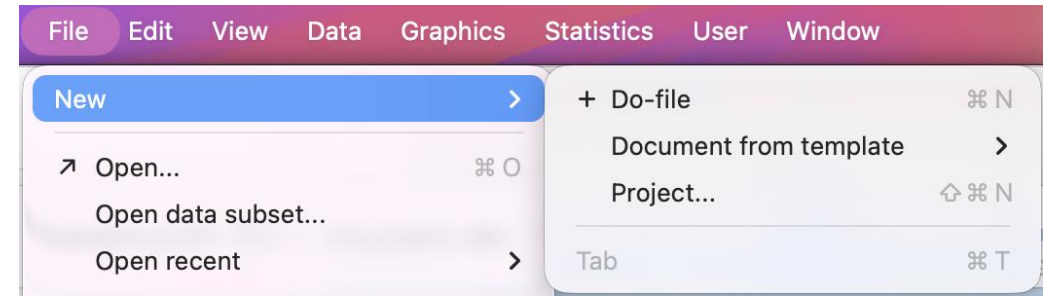
FLB Calibration Test (Favourite–Longshot Bias)

In-class mini-assignment • 10–12 minutes • Work in pairs

Steps:

1. Download the `oddslong.dta` data set (STATA specific)
2. Open a do file in STATA
3. Set your working directory to the folder where you stored the data using `cd "PATH/TO/YOUR/FOLDER"`
4. Run the calibration regression `reg y x, vce(cluster gameld)` → choose an appropriate X and y
5. Perform a statistical test on regression output to determine whether FLB occurs in data using `test (condition 1) (condition2)`

More hints in README file



Opening do-file

Ezra

Join at menti.com | use code 5200 0513

Mentimeter



SH



I find that there is FLB in the NHL data



Menti

New presentation



What can we use to predict match outcomes?

past results
toto xg
injuries
home advantage

I find that there is FLB in the NHL data



Erasmus

Athlete Earnings and Team Wagebills

Earnings

Individual athlete earnings



+ 5 soccer
players

Erasmus

Individual athlete earnings

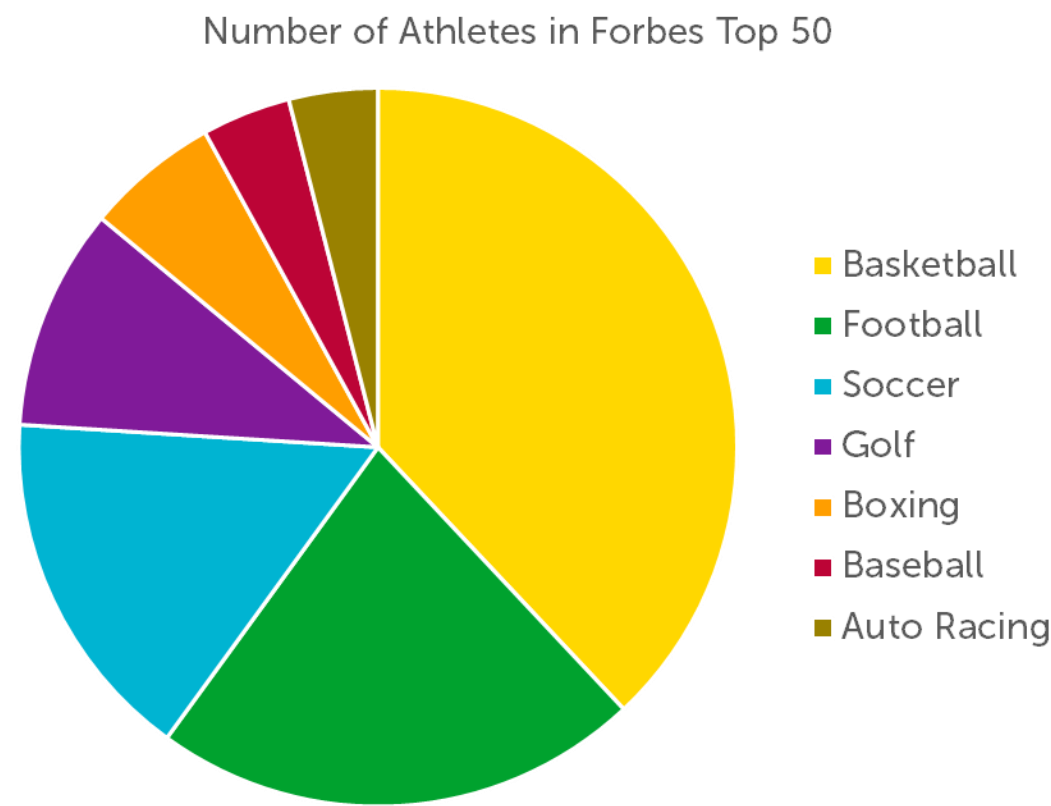
Rank	Name	Sport	Total earnings (\$ Million)	On-the-field earnings (\$ million)	Off-the-field earnings (\$ million)
1	Cristiano Ronaldo	Soccer	260.0	200.0	60.0
2	Jon Rahm	Golf	218.0	198.0	20.0
3	Lionel Messi	Soccer	135.0	65.0	70.0
4	LeBron James	Basketball	128.2	48.2	80.0
5	Giannis Antetokounmpo	Basketball	111.0	46.0	65.0
6	Kylian Mbappé	Soccer	110.0	90.0	20.0
7	Neymar	Soccer	108.0	80.0	28.0
8	Karim Benzema	Soccer	106.0	100.0	6.0
9	Stephen Curry	Basketball	102.0	52.0	50.0
10	Lamar Jackson	Football	100.5	98.5	2.0

Source: Forbes (2024)

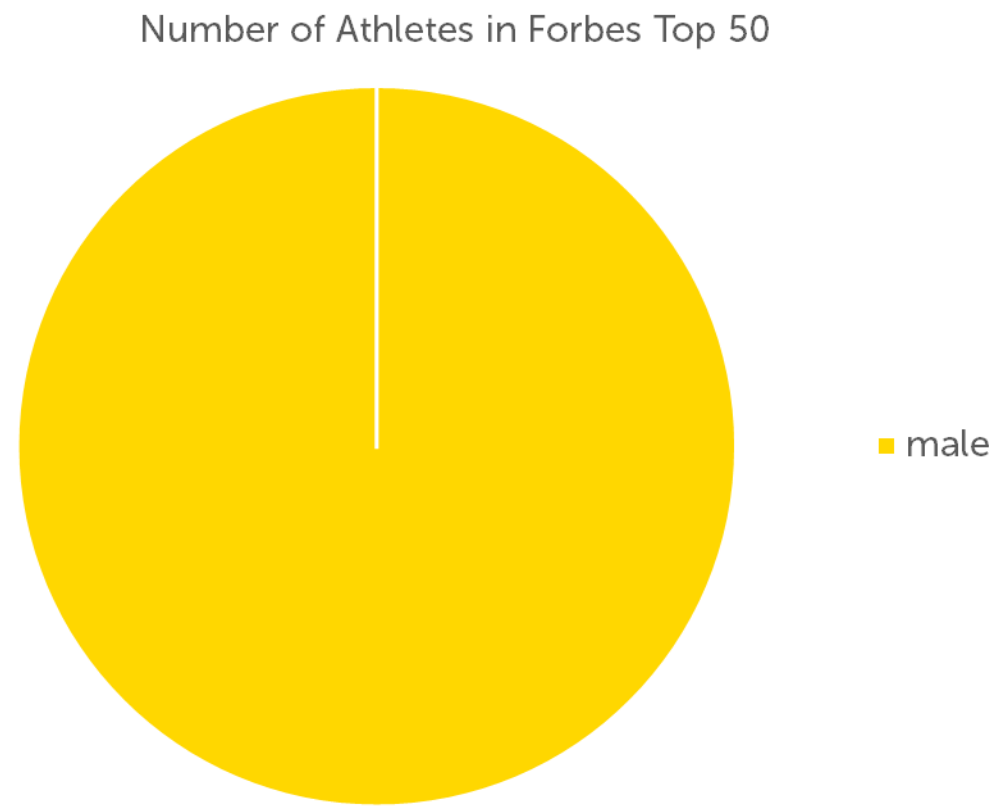
A stylized, handwritten-style logo for Erasmus, featuring a large, flowing 'E' followed by the word 'Erasmus' in a cursive script.

Individual athlete earnings

Sports in top 50

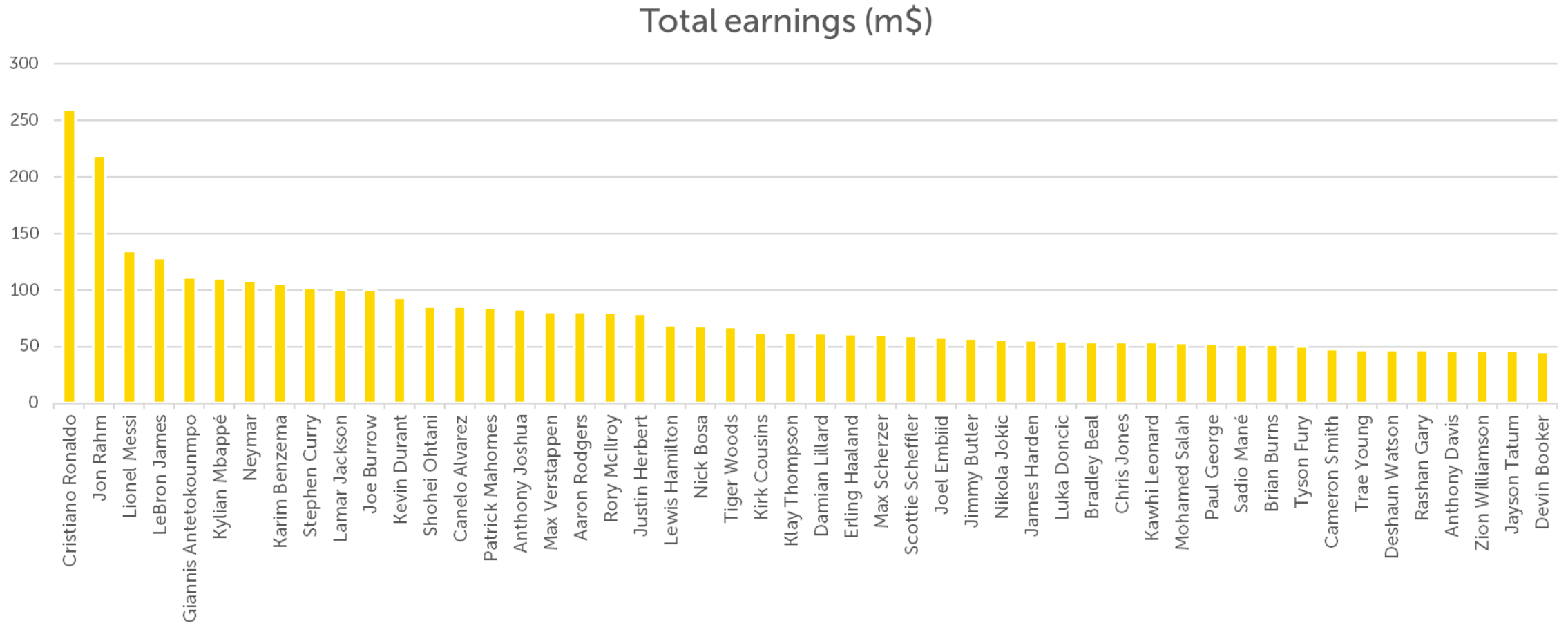


Gender in top 50



Ezra

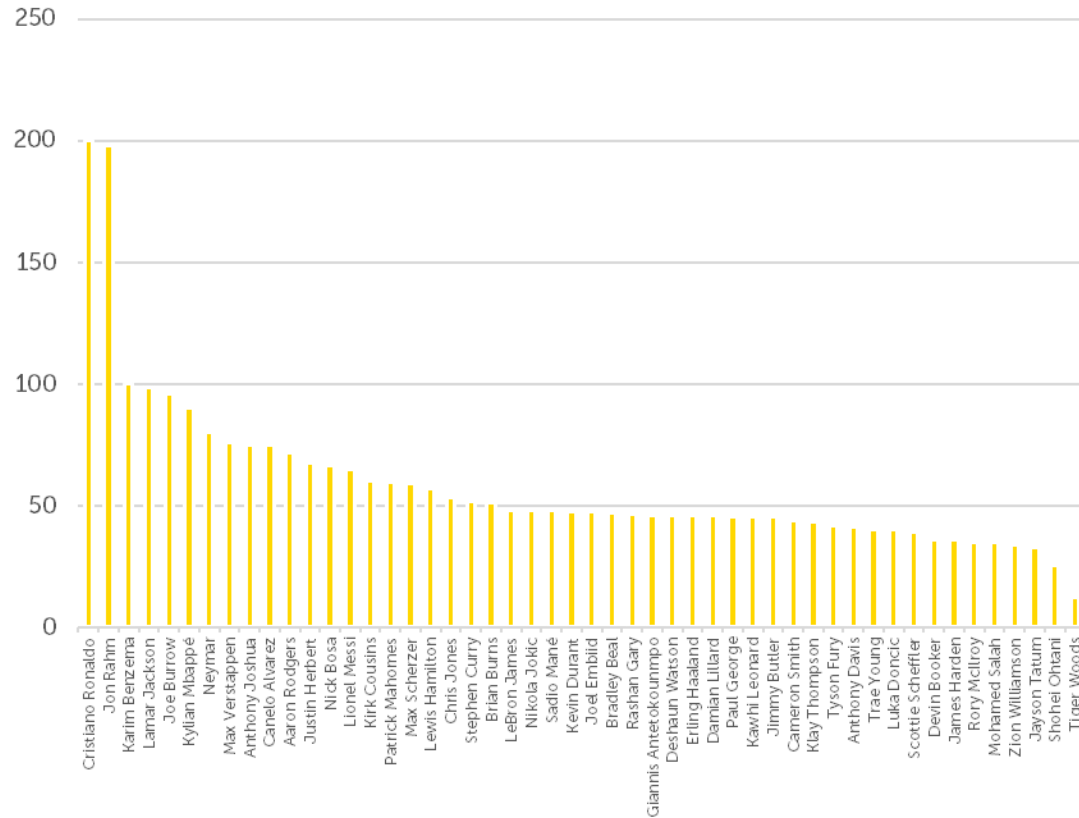
Individual athlete earnings



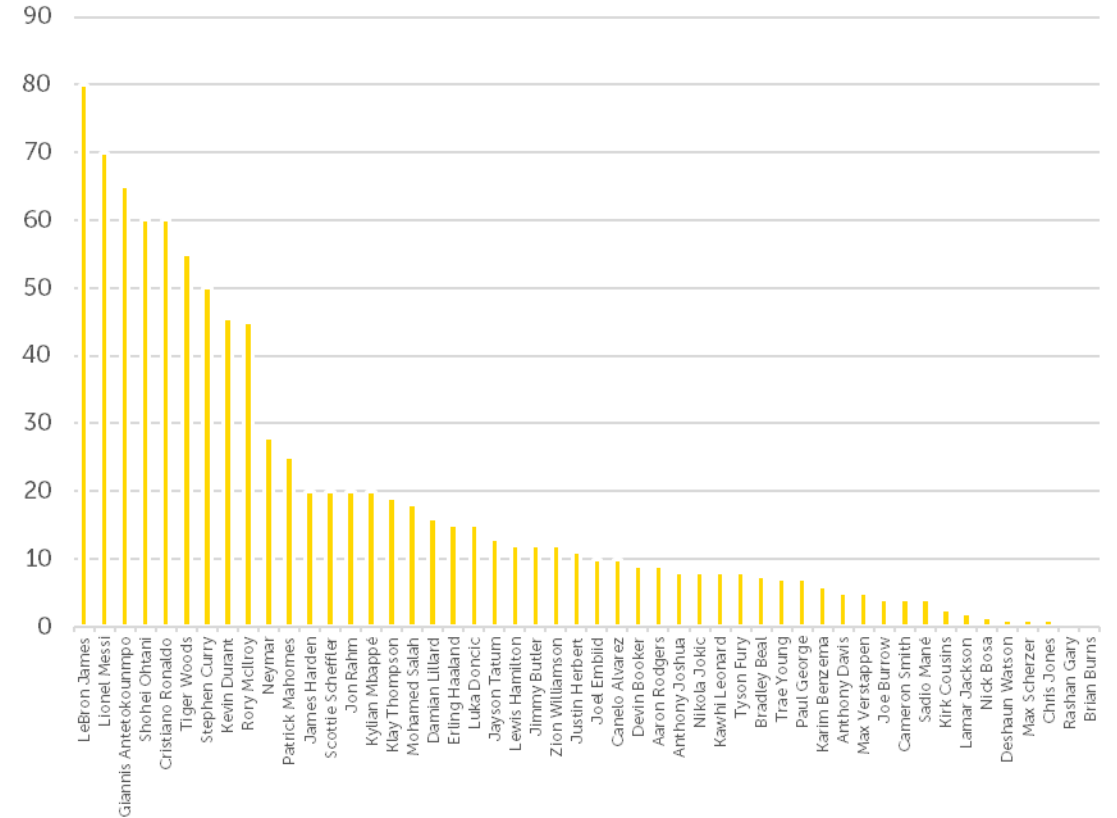
Ezra

Individual athlete earnings

Salary/prize money (\$m)



Endorsements (\$m)

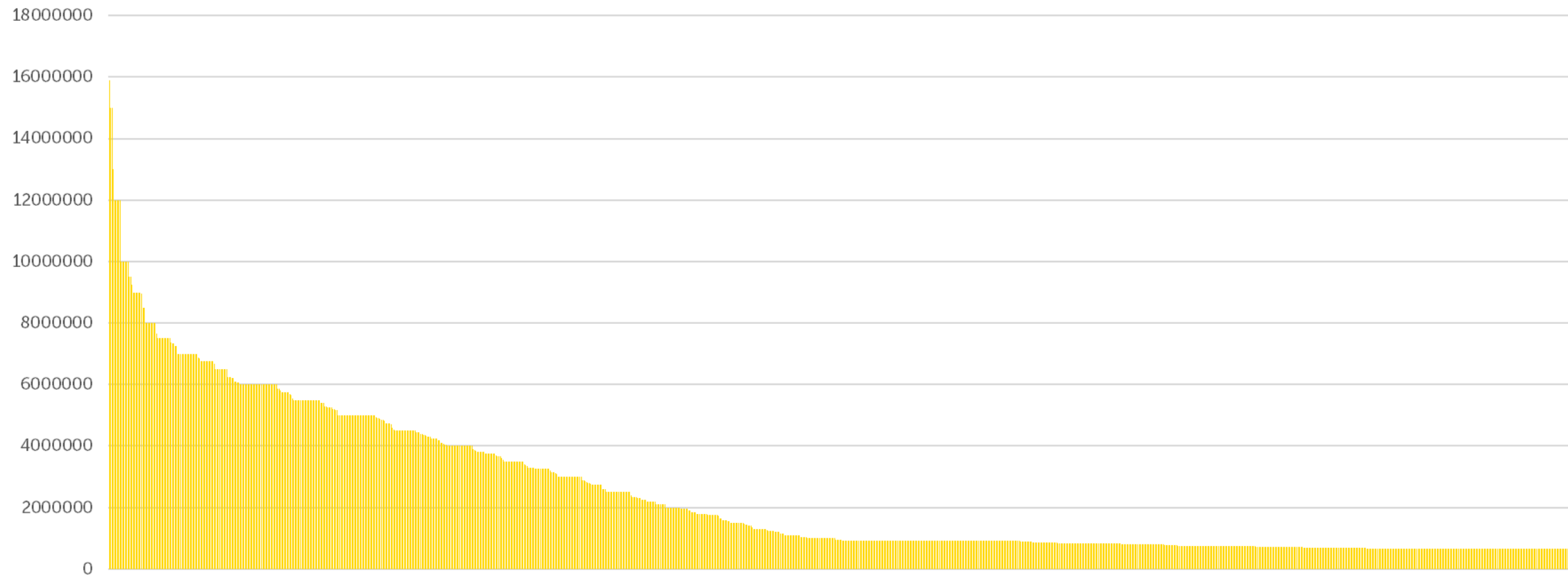


Correlation coefficient: 0.05

Erasmus

Individual athlete earnings

NHL Salaries 2018 (US \$)



Ezra

Individual athlete earnings

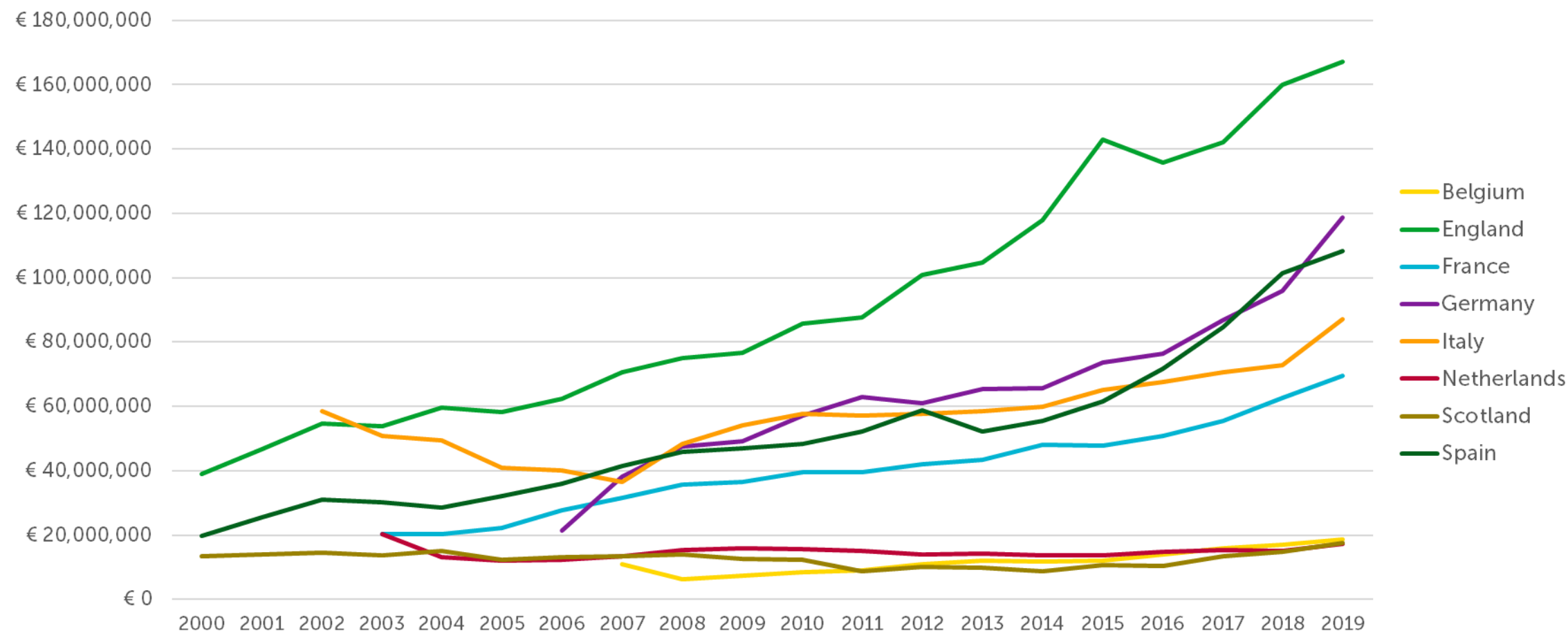
Take-away:

- Top earners:
 - Top 10 worldwide > \$100m annually
 - Top 50 worldwide > \$40m annually
- Extremely skewed:
 - Endorsements skewness >>> Base pay skewness
 - Concentrated in 5-10 sports
 - Gender difference = huge
 - Even within 1 league (e.g. NHL, MLB)

⇒ How does this look at team level?

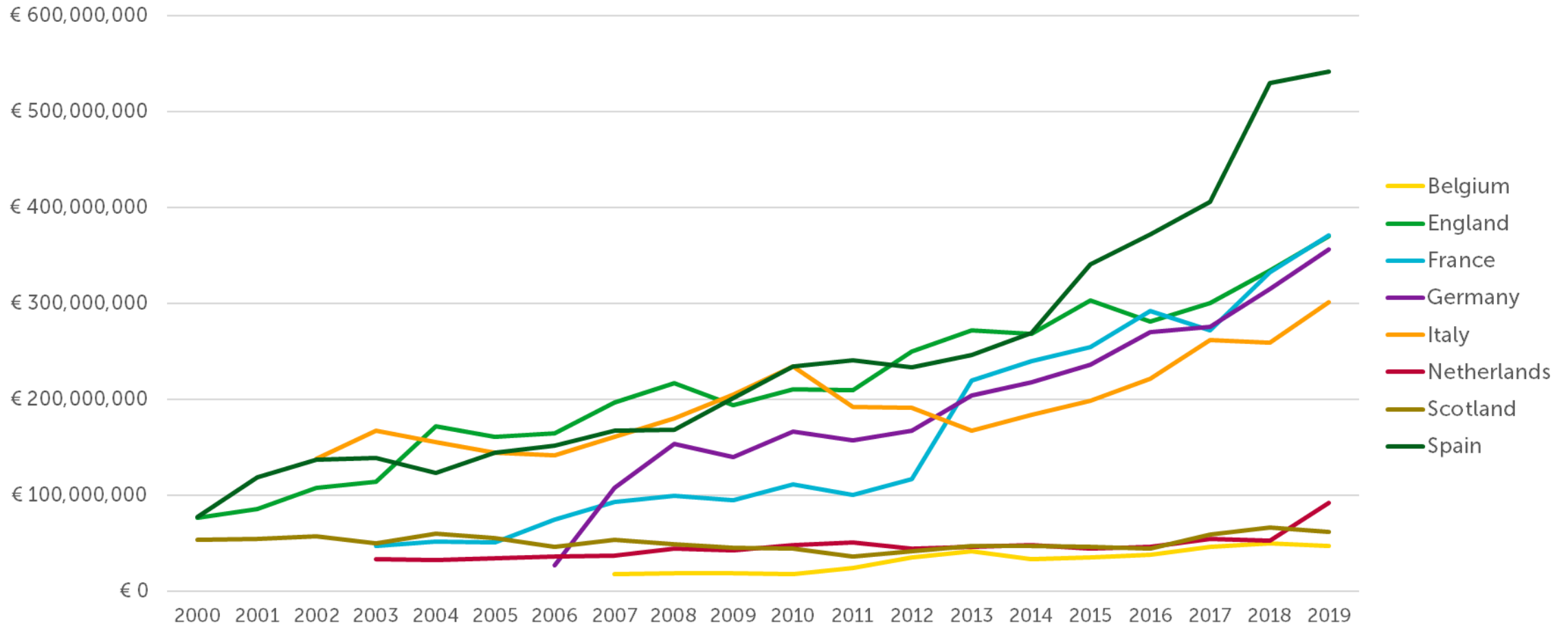
A stylized, handwritten-style logo for Erasmus, featuring a large, flowing 'E' followed by the word 'Erasmus' in a cursive script.

Average wage cost in division 1



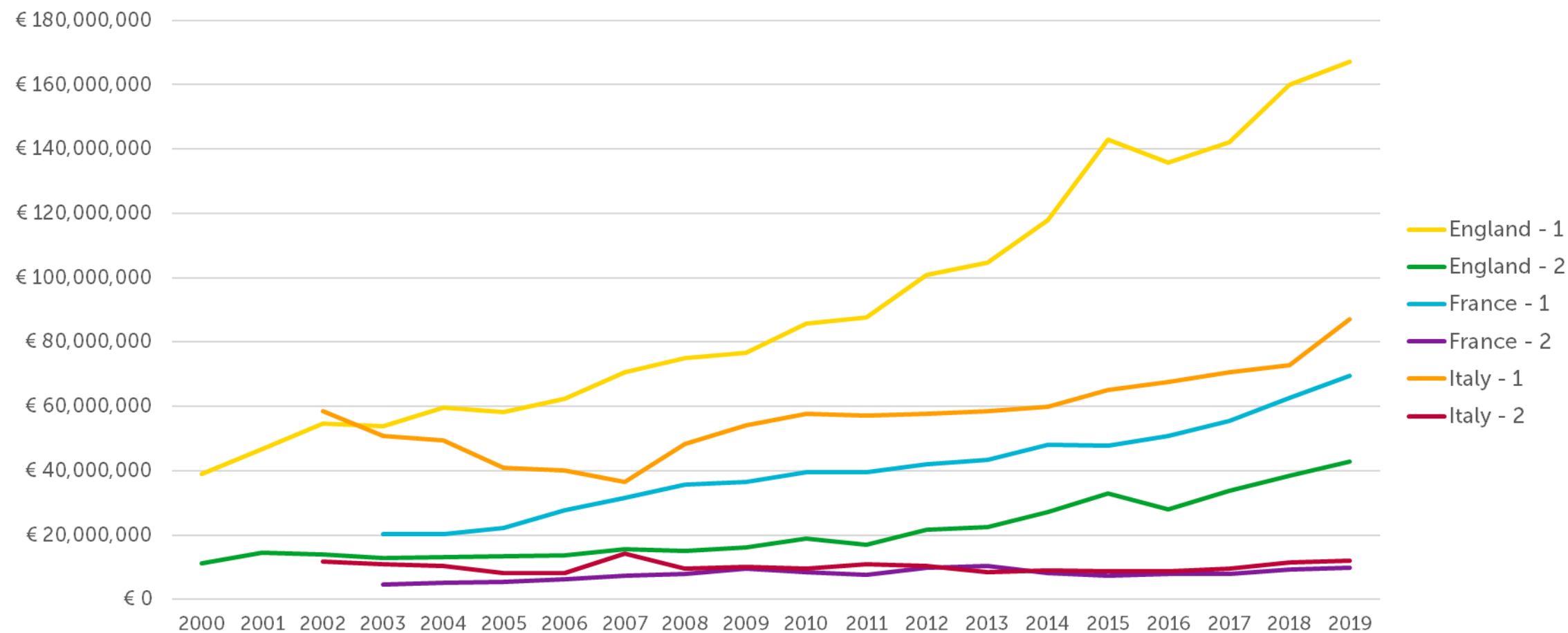
Erasmus

Maximum wage cost in division 1



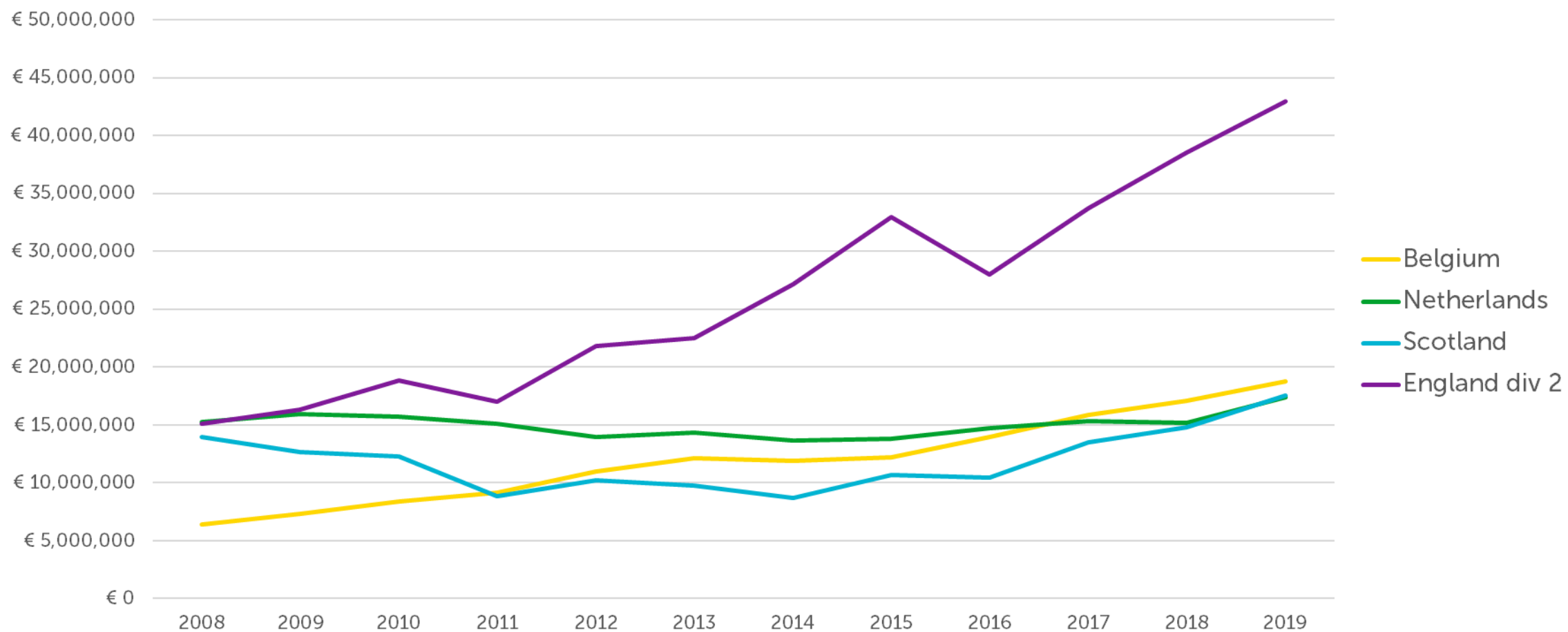
Erasmus

Average wage cost across divisions



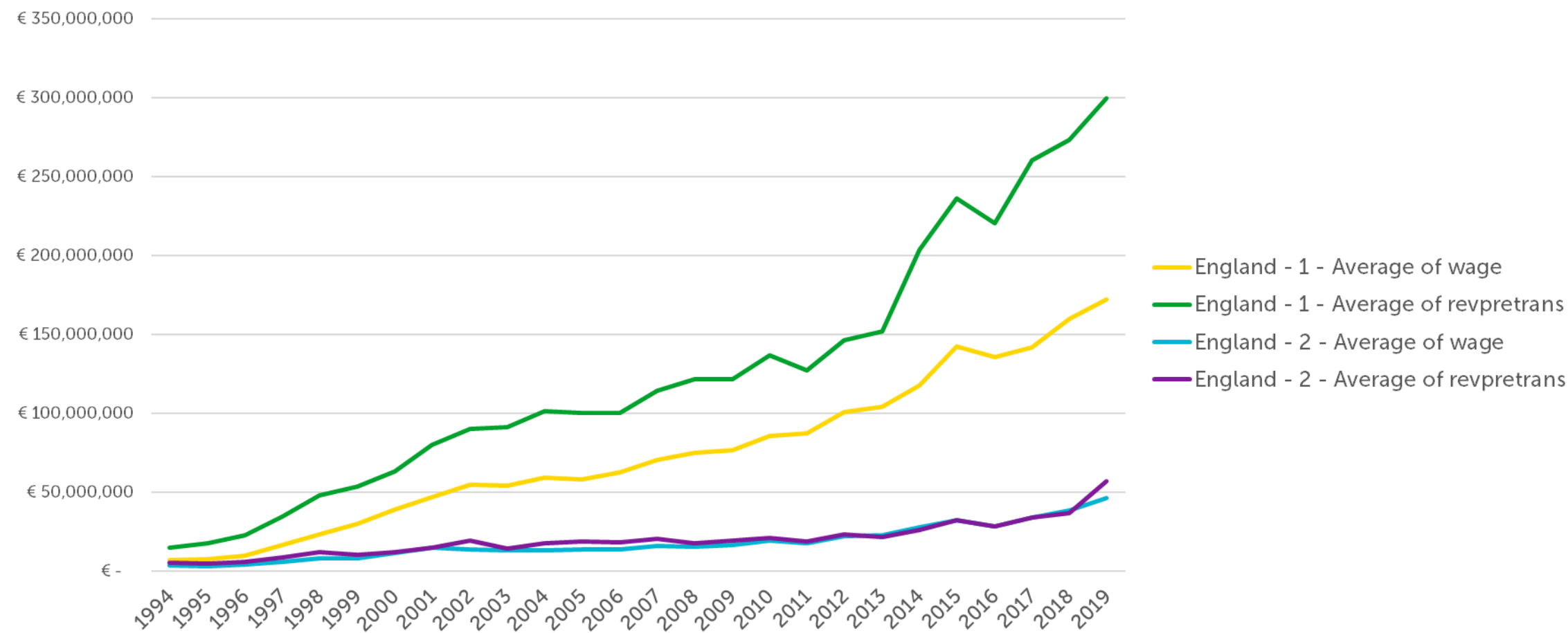
Erasmus

Average wage cost middle income leagues



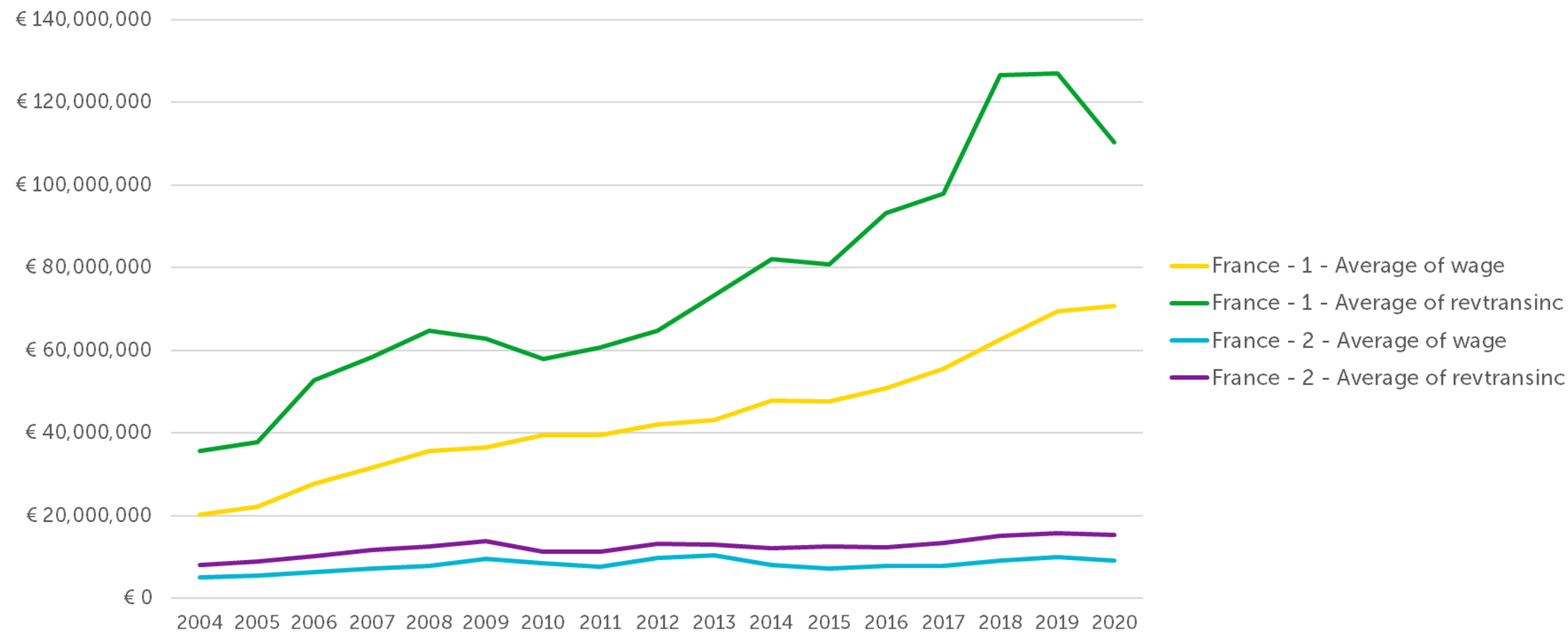
Erasmus

Average wage cost vs. revenues



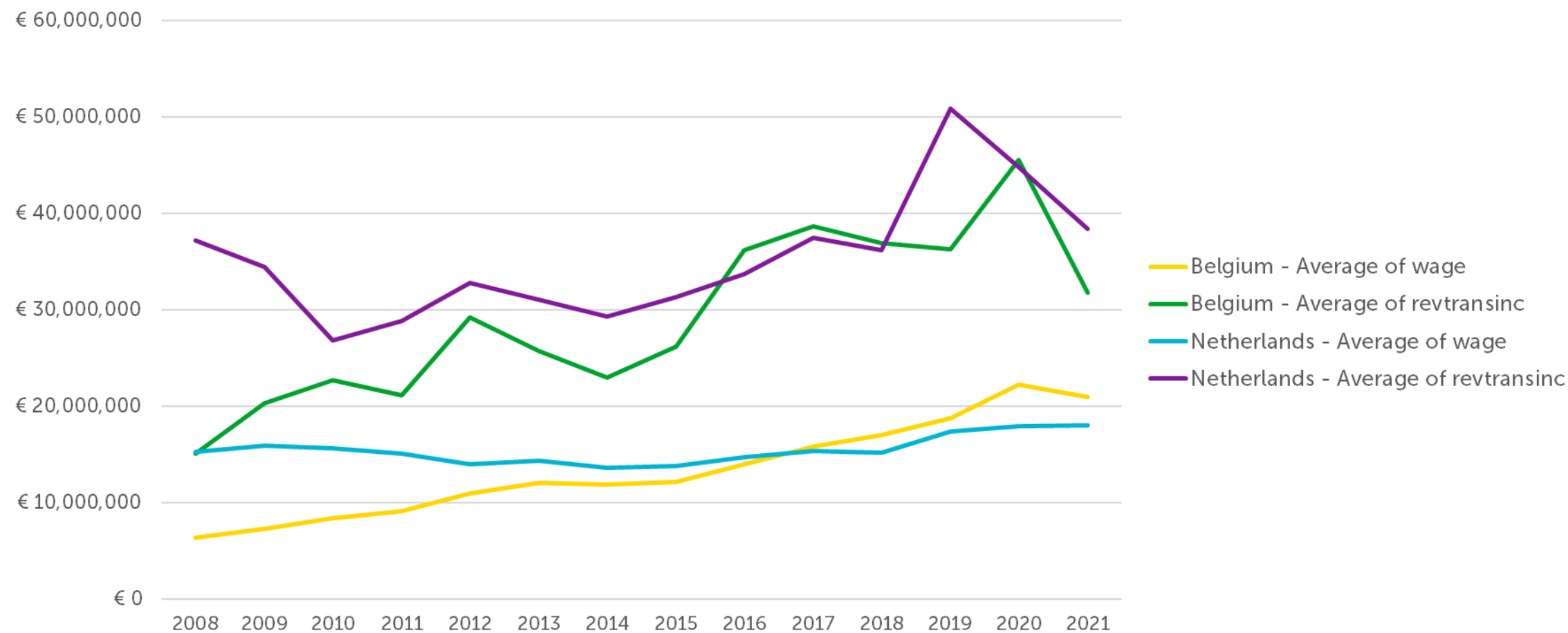
Erasmus

Average wage cost vs. revenues



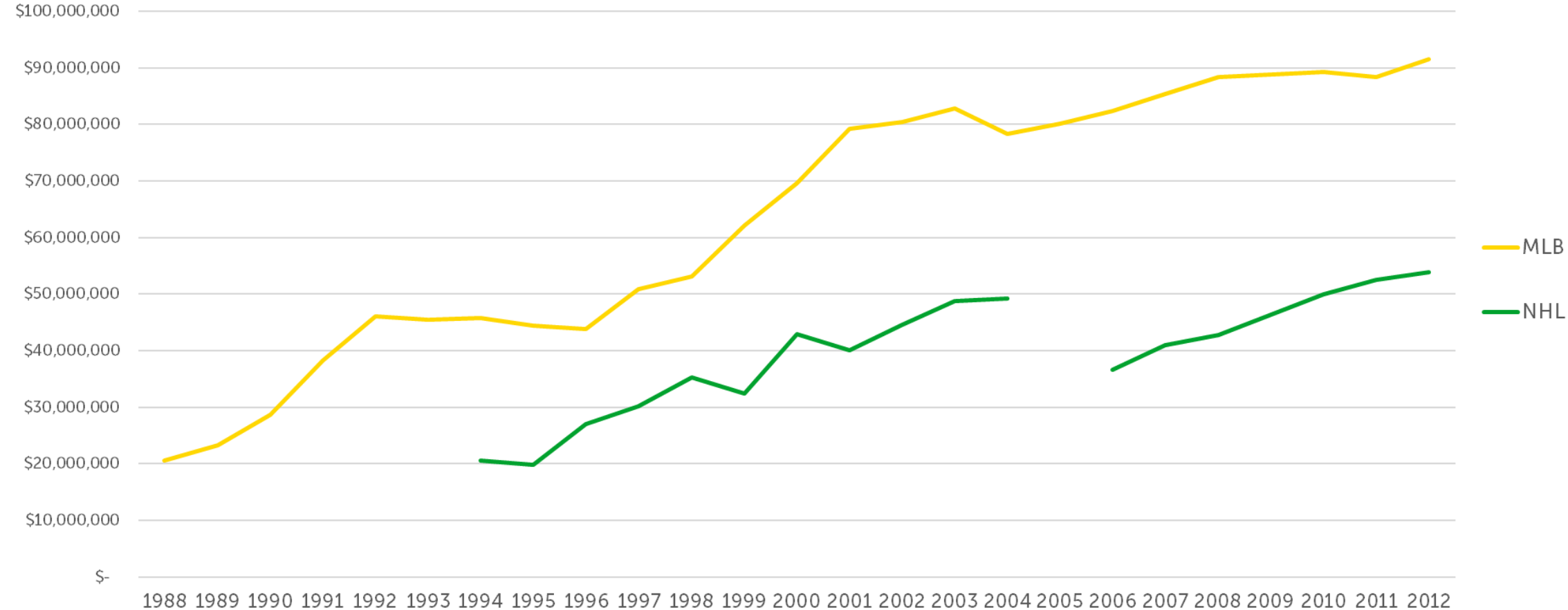
Erasmus

Average wage cost vs. revenues



Erasmus

Average club wage cost in Major Leagues



Erasmus

Club wage spending

Take-away:

- European football:
 - Strong increase in top clubs/leagues
 - Growing inequality between clubs/leagues
 - Follows evolution of club revenues
- US sports:
 - MLB/NHL modest increase in wages in past decade(s)
 - Different trend in revenues
 - Major leagues have regulated labor markets
- But: What is the difference with transfer spending? (teaser for Thursday)

A stylized, handwritten-style logo for Erasmus, featuring a large, flowing 'E' followed by the word 'Erasmus' in a cursive script.