

- Figure 1: Initial Model
- Figure 2: Histograms of All Variables
- Figure 3: Improved Residuals vs Fitted Graph
- Figure 4: Scatterplot Matrix of Data
- Figure 5: Improved Model
- Figure 6: Interaction Plot
- Figure 7: Interaction Model
- Figure 8: Removing Transfer Value Model
- Figure 9: Final Model
- Figure 10: CMF (Final Model)

Appendices

Zack Eisman
2024-04-19

Figure 1: Initial Model

```
##
## Call:
## lm(formula = potential ~ overall + age + value_eur + league_name,
##     data = fifa)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.2679 -1.5268 -0.1989  1.2459  9.7700
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.977e+01  5.888e-01  84.519  < 2e-16 ***
## overall        6.706e-01  1.078e-02  62.229  < 2e-16 ***
## age           -8.122e-01  1.253e-02 -64.844  < 2e-16 ***
## value_eur      3.636e-08  4.040e-09   9.000  < 2e-16 ***
## league_nameFrench Ligue 1 -1.151e+00  1.349e-01  -8.531  < 2e-16 ***
## league_nameGerman 1. Bundesliga -9.832e-01  1.360e-01  -7.228  6.26e-13 ***
## league_nameItalian Serie A  -7.195e-01  1.361e-01  -5.288  1.33e-07 ***
## league_nameSpain Primera Division -3.579e-02  1.314e-01  -0.272    0.785
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.299 on 2870 degrees of freedom
## Multiple R-squared:  0.8012, Adjusted R-squared:  0.8007
## F-statistic: 1652 on 7 and 2870 DF, p-value: < 2.2e-16
```

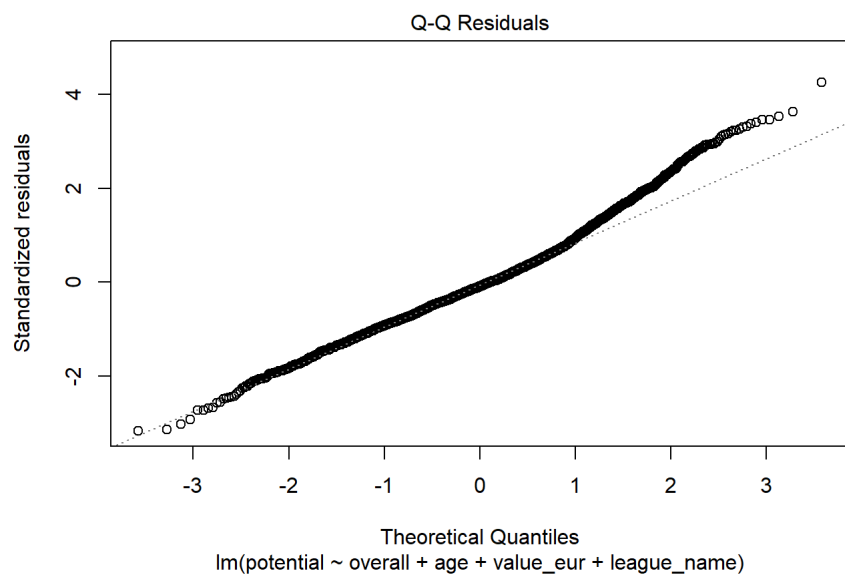
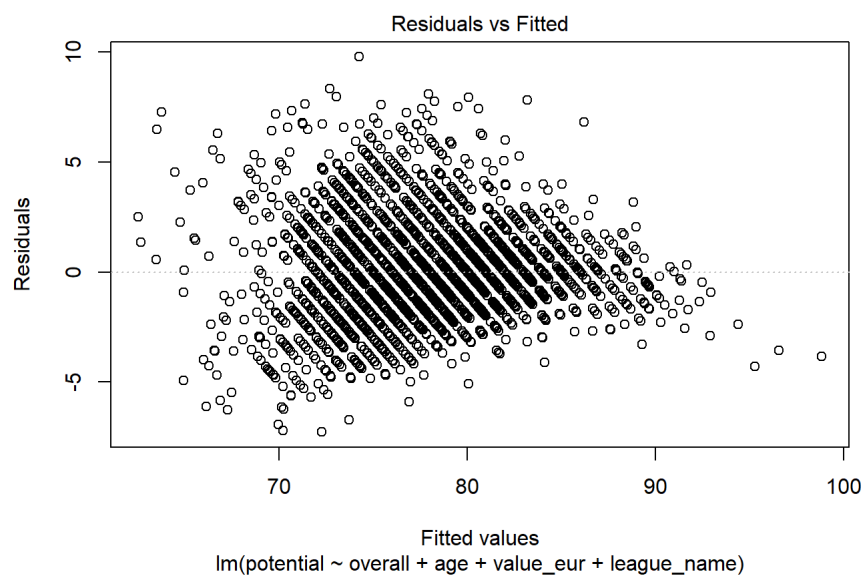


Figure 2: Histograms of All Variables

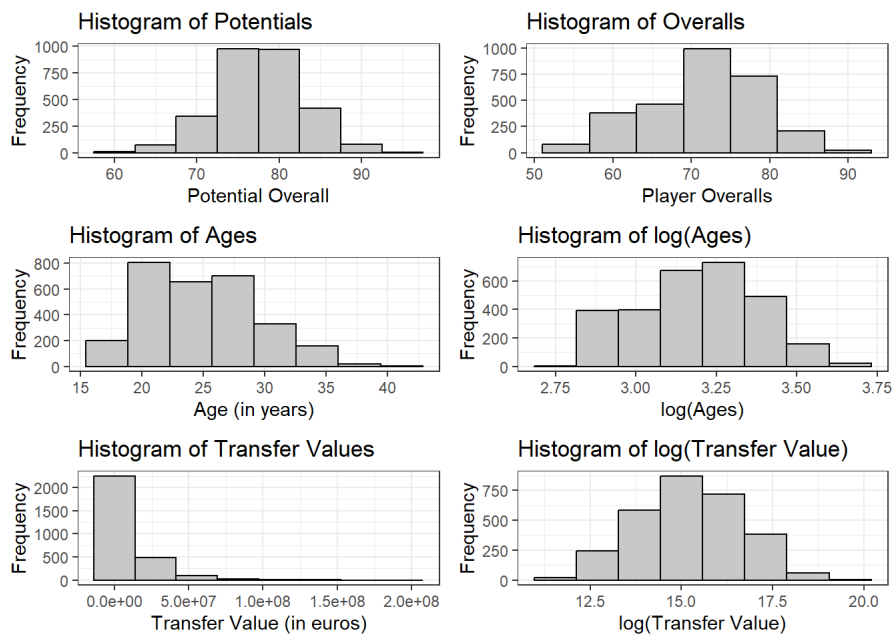


Figure 3: Improved Residuals vs Fitted Graph

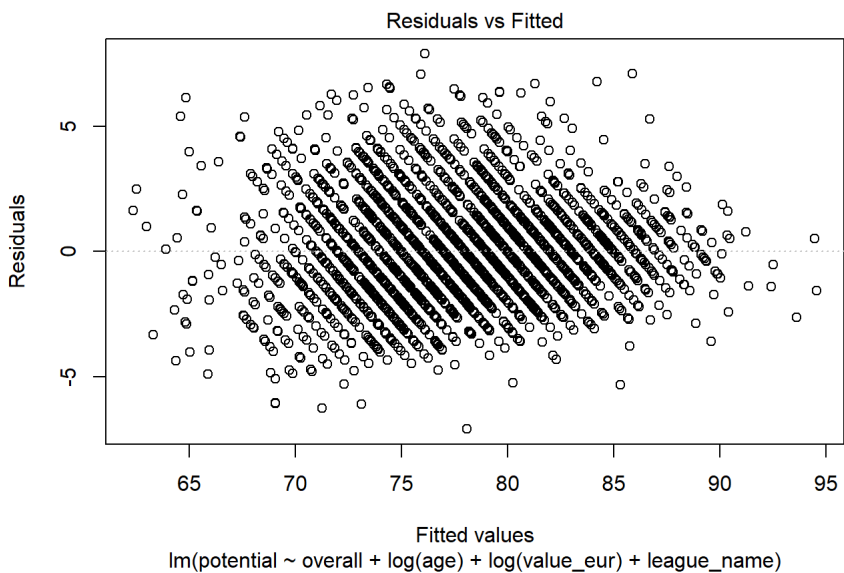


Figure 4: Scatterplot Matrix of Data

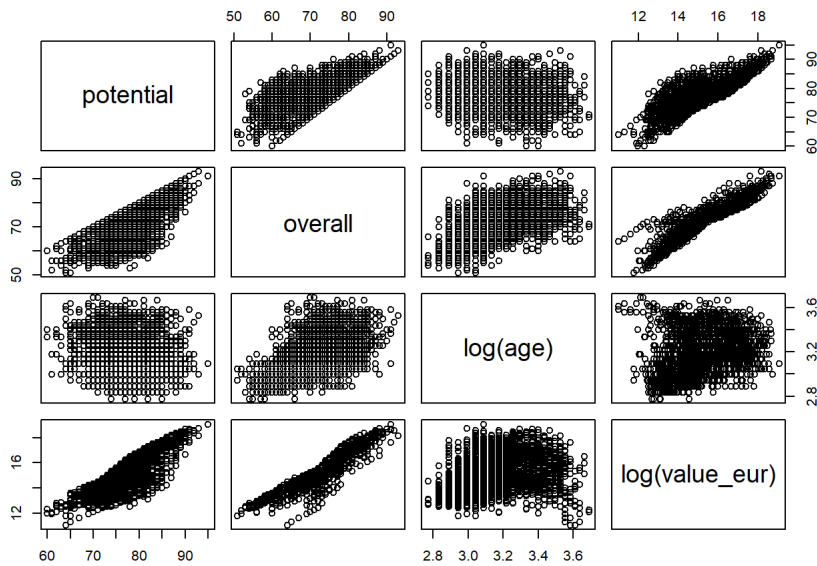


Figure 5: Improved Model

```
##
## Call:
## lm(formula = potential ~ overall + log(age) + log(value_eur) +
##   league_name, data = fifa)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.0971 -1.3923 -0.1708  1.1850  7.9034
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      96.96564    1.45137   66.810 < 2e-16 ***
## overall           0.80706    0.03208   25.155 < 2e-16 ***
## log(age)        -23.61149    0.50647  -46.620 < 2e-16 ***
## log(value_eur)   -0.09014    0.14175   -0.636  0.525
## league_nameFrench Ligue 1    -1.22346    0.12093  -10.117 < 2e-16 ***
## league_nameGerman 1. Bundesliga -1.00332    0.12171   -8.243 2.52e-16 ***
## league_nameItalian Serie A    -0.67378    0.12150   -5.546 3.20e-08 ***
## league_nameSpain Primera Division -0.07013    0.11723   -0.598  0.550
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.061 on 2870 degrees of freedom
## Multiple R-squared:  0.8403, Adjusted R-squared:  0.8399
## F-statistic: 2158 on 7 and 2870 DF, p-value: < 2.2e-16
```

Figure 6: Interaction Plot

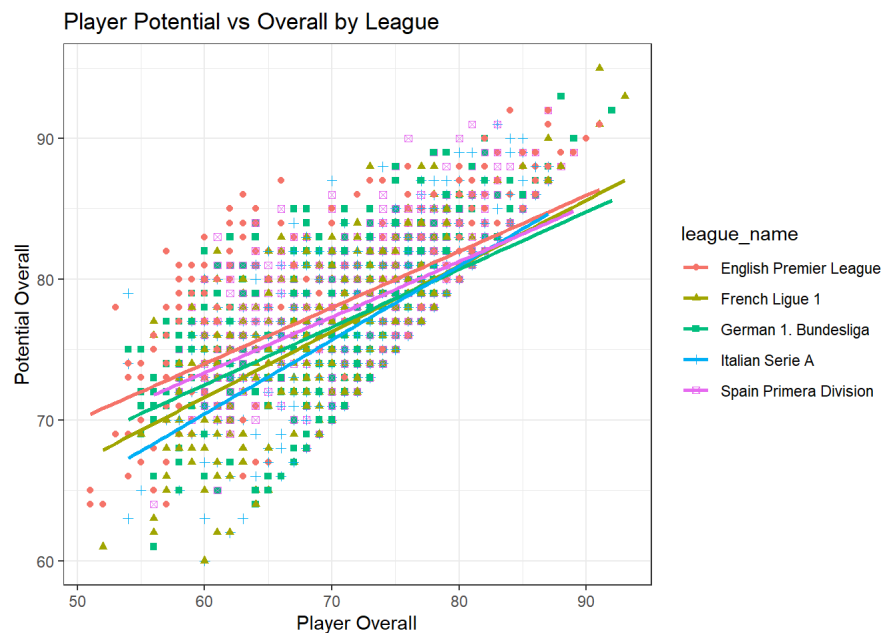
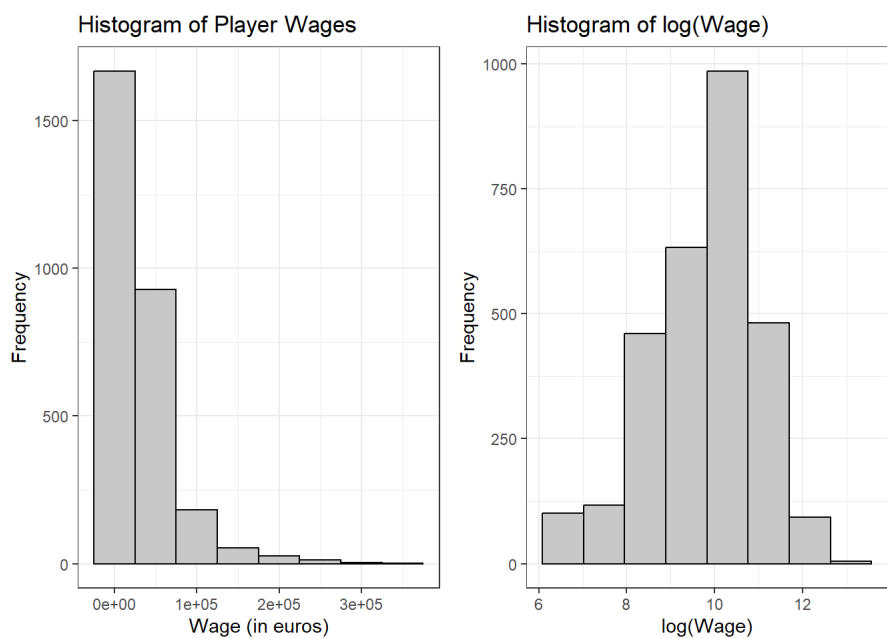


Figure 7: Interaction Model

```
##
## Call:
## lm(formula = potential ~ overall + log(age) + log(value_eur) +
##     league_name + league_name * overall, data = fifa)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.7157 -1.3544 -0.1709  1.1555  7.7473
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    101.21730    1.56820   64.544 < 2e-16
## overall         0.76525    0.03243   23.599 < 2e-16
## log(age)       -23.78700    0.50321  -47.270 < 2e-16
## log(value_eur)  -0.13400    0.14067   -0.953  0.34089
## league_nameFrench Ligue 1    -8.26456    1.13351   -7.291 3.96e-13
## league_nameGerman 1. Bundesliga -3.18442    1.11581   -2.854 0.00435
## league_nameItalian Serie A    -8.26944    1.20022   -6.890 6.84e-12
## league_nameSpain Primera Division -5.85858    1.13245   -5.173 2.46e-07
## overall:league_nameFrench Ligue 1    0.09844    0.01583    6.217 5.81e-10
## overall:league_nameGerman 1. Bundesliga  0.02979    0.01552    1.920 0.05499
## overall:league_nameItalian Serie A    0.10486    0.01646    6.370 2.19e-10
## overall:league_nameSpain Primera Division 0.07977    0.01546    5.159 2.65e-07
##
## (Intercept)          ***
## overall              ***
## log(age)             ***
## log(value_eur)
## league_nameFrench Ligue 1      ***
## league_nameGerman 1. Bundesliga **
## league_nameItalian Serie A      ***
## league_nameSpain Primera Division ***
## overall:league_nameFrench Ligue 1 ***
## overall:league_nameGerman 1. Bundesliga .
## overall:league_nameItalian Serie A ***
## overall:league_nameSpain Primera Division ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.038 on 2866 degrees of freedom
## Multiple R-squared:  0.844, Adjusted R-squared:  0.8434
## F-statistic: 1410 on 11 and 2866 DF, p-value: < 2.2e-16
```

Figure 8: Removing Transfer Value Model

```
##
## Call:
## lm(formula = potential ~ overall + log(age) + log(wage_eur) +
##     league_name + league_name * overall, data = fifa)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.4878 -1.3351 -0.1419  1.1502  7.4933
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      99.14938      0.86295 114.896 < 2e-16
## overall           0.84548      0.01476  57.281 < 2e-16
## log(age)        -23.02183      0.25774 -89.321 < 2e-16
## log(wage_eur)    -0.80616      0.08092  -9.963 < 2e-16
## league_nameFrench Ligue 1    -9.94844      1.12737  -8.825 < 2e-16
## league_nameGerman 1. Bundesliga -4.47340      1.10476  -4.049 5.28e-05
## league_nameItalian Serie A   -8.75873      1.18039  -7.420 1.53e-13
## league_nameSpain Primera Division -6.80522      1.11575  -6.099 1.21e-09
## overall:league_nameFrench Ligue 1  0.11535      0.01566   7.366 2.29e-13
## overall:league_nameGerman 1. Bundesliga  0.04244      0.01531   2.771 0.00562
## overall:league_nameItalian Serie A    0.10728      0.01618   6.632 3.95e-11
## overall:league_nameSpain Primera Division 0.08576      0.01518   5.649 1.78e-08
##
## (Intercept)          ***
## overall              ***
## log(age)             ***
## log(wage_eur)        ***
## league_nameFrench Ligue 1 ***
## league_nameGerman 1. Bundesliga ***
## league_nameItalian Serie A ***
## league_nameSpain Primera Division ***
## overall:league_nameFrench Ligue 1 ***
## overall:league_nameGerman 1. Bundesliga **
## overall:league_nameItalian Serie A ***
## overall:league_nameSpain Primera Division ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.004 on 2866 degrees of freedom
## Multiple R-squared:  0.8492, Adjusted R-squared:  0.8486
## F-statistic: 1467 on 11 and 2866 DF, p-value: < 2.2e-16
```



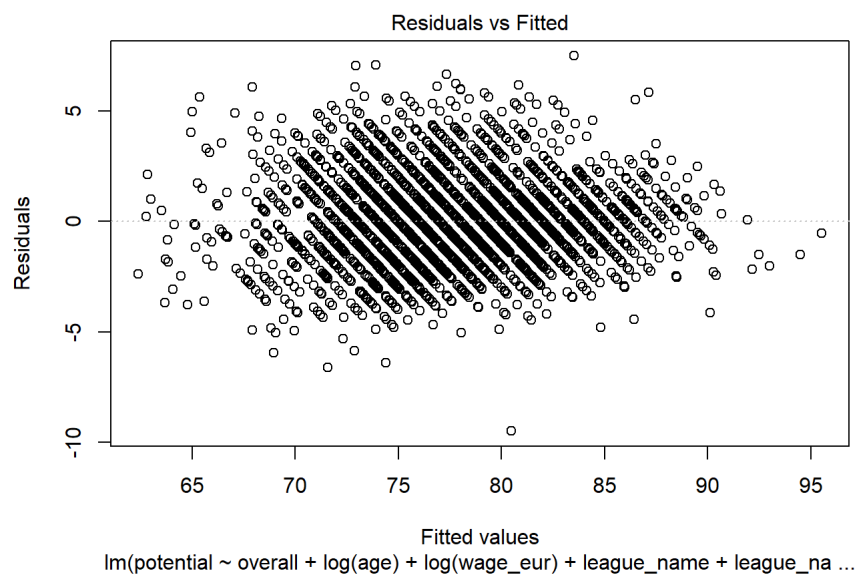
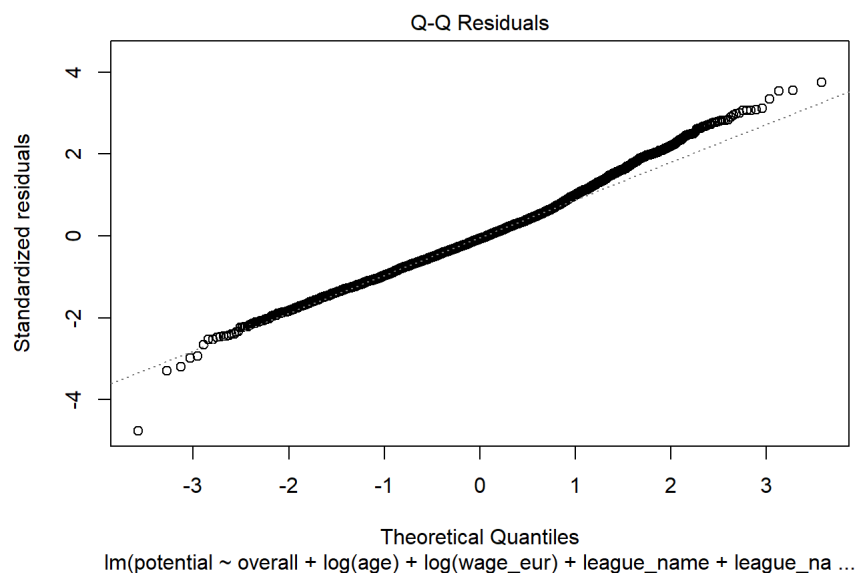
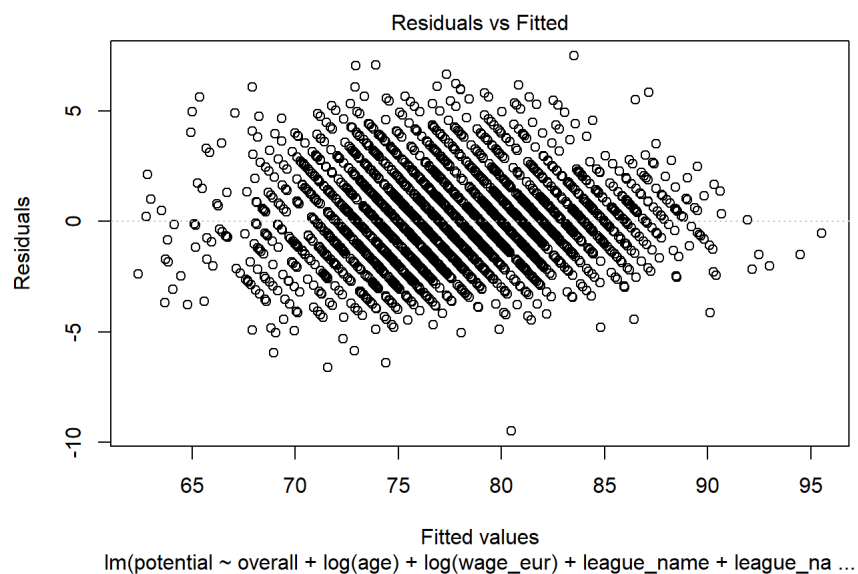


Figure 9: Final Model

```
##
## Call:
## lm(formula = potential ~ overall + log(age) + log(wage_eur) +
##     league_name + league_name * overall, data = fifa)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.4878 -1.3351 -0.1419  1.1502  7.4933
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    99.14938    0.86295  114.896 < 2e-16
## overall         0.84548    0.01476   57.281 < 2e-16
## log(age)       -23.02183    0.25774  -89.321 < 2e-16
## log(wage_eur)  -0.80616    0.08092   -9.963 < 2e-16
## league_nameFrench Ligue 1    -9.94844    1.12737   -8.825 < 2e-16
## league_nameGerman 1. Bundesliga -4.47340    1.10476   -4.049 5.28e-05
## league_nameItalian Serie A   -8.75873    1.18039   -7.420 1.53e-13
## league_nameSpain Primera Division -6.80522    1.11575   -6.099 1.21e-09
## overall:league_nameFrench Ligue 1    0.11535    0.01566    7.366 2.29e-13
## overall:league_nameGerman 1. Bundesliga  0.04244    0.01531    2.771 0.00562
## overall:league_nameItalian Serie A    0.10728    0.01618    6.632 3.95e-11
## overall:league_nameSpain Primera Division 0.08576    0.01518    5.649 1.78e-08
##
## (Intercept)          ***
## overall              ***
## log(age)             ***
## log(wage_eur)        ***
## league_nameFrench Ligue 1 ***
## league_nameGerman 1. Bundesliga ***
## league_nameItalian Serie A ***
## league_nameSpain Primera Division ***
## overall:league_nameFrench Ligue 1 ***
## overall:league_nameGerman 1. Bundesliga **
## overall:league_nameItalian Serie A ***
## overall:league_nameSpain Primera Division ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.004 on 2866 degrees of freedom
## Multiple R-squared:  0.8492, Adjusted R-squared:  0.8486
## F-statistic: 1467 on 11 and 2866 DF, p-value: < 2.2e-16
```



##	overall	log(age)	log(wage_eur)
##	5.750872	1.595194	5.509427

Figure 10: CMF (Final Model)

$$E(Y_{potential}|X) = \beta_0 + \beta_1 X_{overall} + \beta_2 \log(X_{age}) + \beta_3 \log(X_{wage}) + \beta_4 I_{Ligue1} + \beta_5 I_{Bundesliga} + \beta_6 I_{SerieA} + \beta_7 I_{LaLiga} + \beta_8 X_{overall} I_{Ligue1} + \beta_9 X_{overall} I_{Bundesliga} + \beta_{10} X_{overall}$$