Supplement

Contents

Code is available at: https://github.com/epiforecasts/model-structure-evaluation.

1 Model characteristics

Table 1: Model characteristics contributing to the European COVID-19 Forecast Hub, by method used, number of countries targeted, and number of forecasts contributed.

Model	Method	Country Targets	Case forecasts	Death forecasts
AMM-EpiInvert	Statistical	Multi-country	2788 (1.3%)	
CovidMetrics-epiBATS	Statistical	Single-country	343 (0.2%)	
DSMPG-bayes	Semi-mechanistic	Multi-country	760 (0.4%)	
EuroCOVIDhub-baseline	Statistical	Multi-country	13082 (6.3%)	13040 (6.3%)
FIAS_FZJ-Epi1Ger	Mechanistic	Single-country	264 (0.1%)	264 (0.1%)
GoeWroc-BaseBayes	Semi-mechanistic	Single-country	12 (0%)	
HZI-AgeExtendedSEIR	Mechanistic	Single-country	382 (0.2%)	382 (0.2%)
ICM-agentModel	Agent-based	Single-country	334 (0.2%)	334 (0.2%)
IEM_Health-CovidProject	Mechanistic	Multi-country	7710 (3.7%)	7708 (3.7%)
ILM-EKF	Semi-mechanistic	Multi-country	11998 (5.8%)	11961 (5.8%)
ITWW-county_repro	Semi-mechanistic	Single-country	650 (0.3%)	600 (0.3%)
Imperial-DeCa	Semi-mechanistic	Multi-country		571 (0.3%)
Imperial-RtI0	Semi-mechanistic	Multi-country		571 (0.3%)
Imperial-sbkp	Semi-mechanistic	Multi-country		571 (0.3%)
JBUD-HMXK	Mechanistic	Multi-country	1324 (0.6%)	1324 (0.6%)
KITmetricslab-bivar_branching	Statistical	Single-country	8 (0%)	
Karlen-pypm	Mechanistic	Multi-country	3208 (1.5%)	3186 (1.5%)
LANL-GrowthRate	Semi-mechanistic	Multi-country	3692 (1.8%)	3696 (1.8%)
LeipzigIMISE-SECIR	Mechanistic	Single-country	16 (0%)	16 (0%)
MIMUW-StochSEIR	Mechanistic	Single-country	76 (0%)	76 (0%)
MIT_CovidAnalytics-DELPHI	Mechanistic	Single-country	348 (0.2%)	500 (0.2%)
MOCOS-agent1	Agent-based	Single-country	386 (0.2%)	386 (0.2%)
MUNI-ARIMA	Statistical	Multi-country	10979 (5.3%)	11314 (5.4%)
MUNI-LaggedRegARIMA	Statistical	Multi-country		736 (0.4%)
MUNI-VAR	Statistical	Multi-country	976 (0.5%)	976 (0.5%)
MUNI_DMS-SEIAR	Mechanistic	Single-country	224 (0.1%)	200 (0.1%)
PL_GRedlarski-DistrictsSum	Mechanistic	Single-country	378 (0.2%)	
RobertWalraven-ESG	Statistical	Multi-country	9190 (4.4%)	10465 (5%)
SDSC_ISG-TrendModel	Statistical	Multi-country	1756 (0.8%)	1744 (0.8%)
UB-BSLCoV	Statistical	Single-country	96 (0%)	96 (0%)
UC3M-EpiGraph	Agent-based	Single-country	94 (0%)	
ULZF-SEIRC19SI	Mechanistic	Single-country	249 (0.1%)	249 (0.1%)
UMass-MechBayes	Mechanistic	Multi-country		5948 (2.9%)
UMass-SemiMech	Semi-mechanistic	Multi-country	1888 (0.9%)	1904 (0.9%)
UNED-PreCoV2	Statistical	Single-country	147 (0.1%)	147 (0.1%)
UNIPV-BayesINGARCHX	Statistical	Multi-country	426 (0.2%)	
USC-SIkJalpha	Mechanistic	Multi-country	12900 (6.2%)	12688 (6.1%)
UpgUmibUsi-MultiBayes	Semi-mechanistic	Single-country	99 (0%)	99 (0%)
bisop-seirfilter	Mechanistic	Single-country	32 (0%)	32 (0%)
bisop-seirfilterlite	Mechanistic	Multi-country	336 (0.2%)	336 (0.2%)
epiMOX-SUIHTER	Mechanistic	Single-country	134 (0.1%)	134 (0.1%)
epiforecasts-EpiExpert	Qualitative	Multi-country	945 (0.5%)	948 (0.5%)
epiforecasts-EpiExpert_Rt	Qualitative	Multi-country	404 (0.2%)	404 (0.2%)
epiforecasts-EpiExpert_direct	Qualitative	Multi-country	394 (0.2%)	392 (0.2%)
epiforecasts-EpiNow2	Semi-mechanistic	Multi-country	8843 (4.3%)	7721 (3.7%)
epiforecasts-weeklygrowth	Statistical	Multi-country	5971 (2.9%)	
itwm-dSEIR	Mechanistic	Single-country	406 (0.2%)	406 (0.2%)
prolix-euclidean	Semi-mechanistic	Multi-country	800 (0.4%)	800 (0.4%)

2 Distribution of model scores

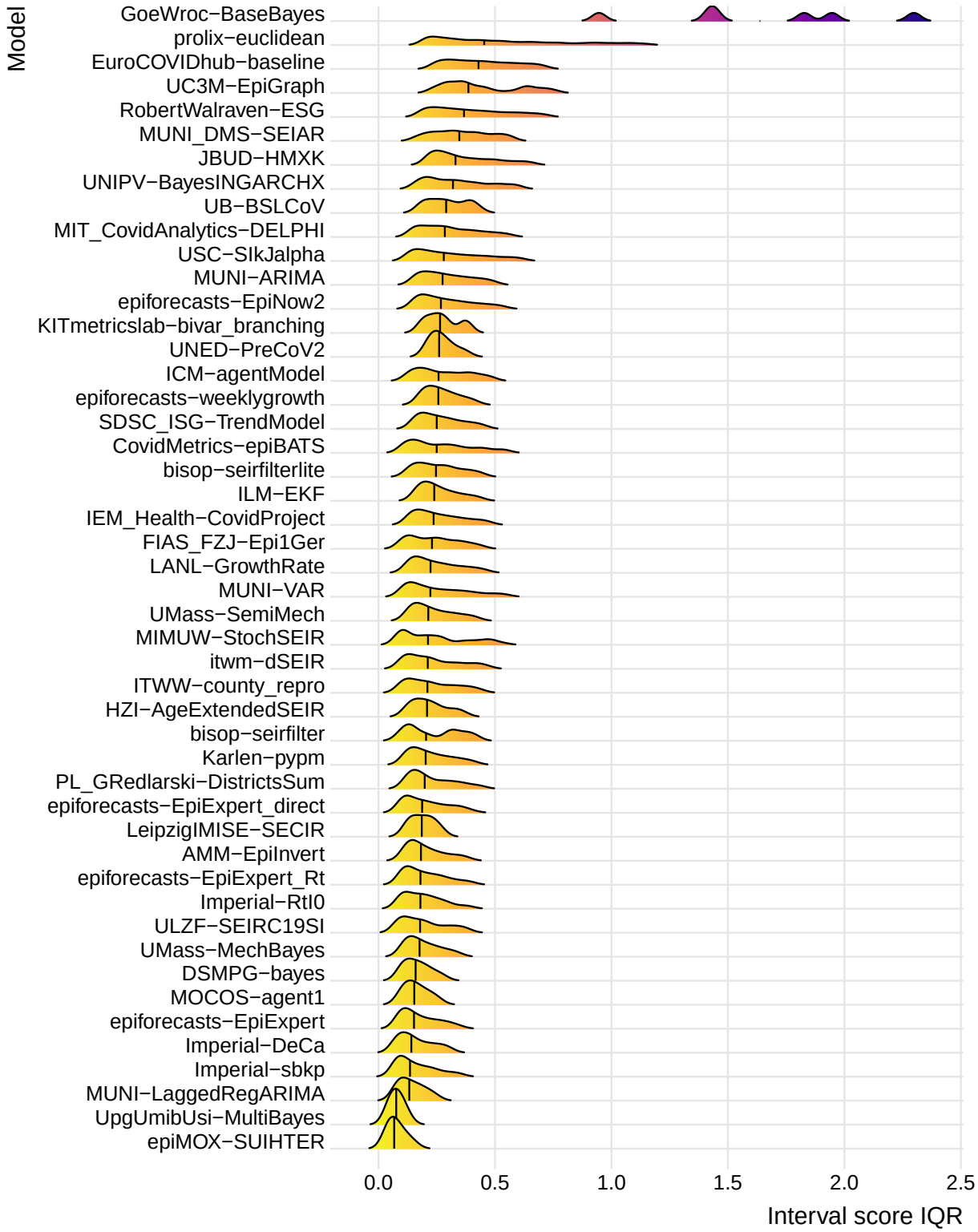


Figure 1: Distribution of forecast scores for each contributing model. Each distribution shows the interquartile range and median (vertical line) of interval scores across forecasts made by each model, with lower interval score indicating better predictive accuracy. Each model forecast for a different combination of targets, with some models contributing very few forecasts, meaning that forecast scores are not directly comparable.

3 Trend identification

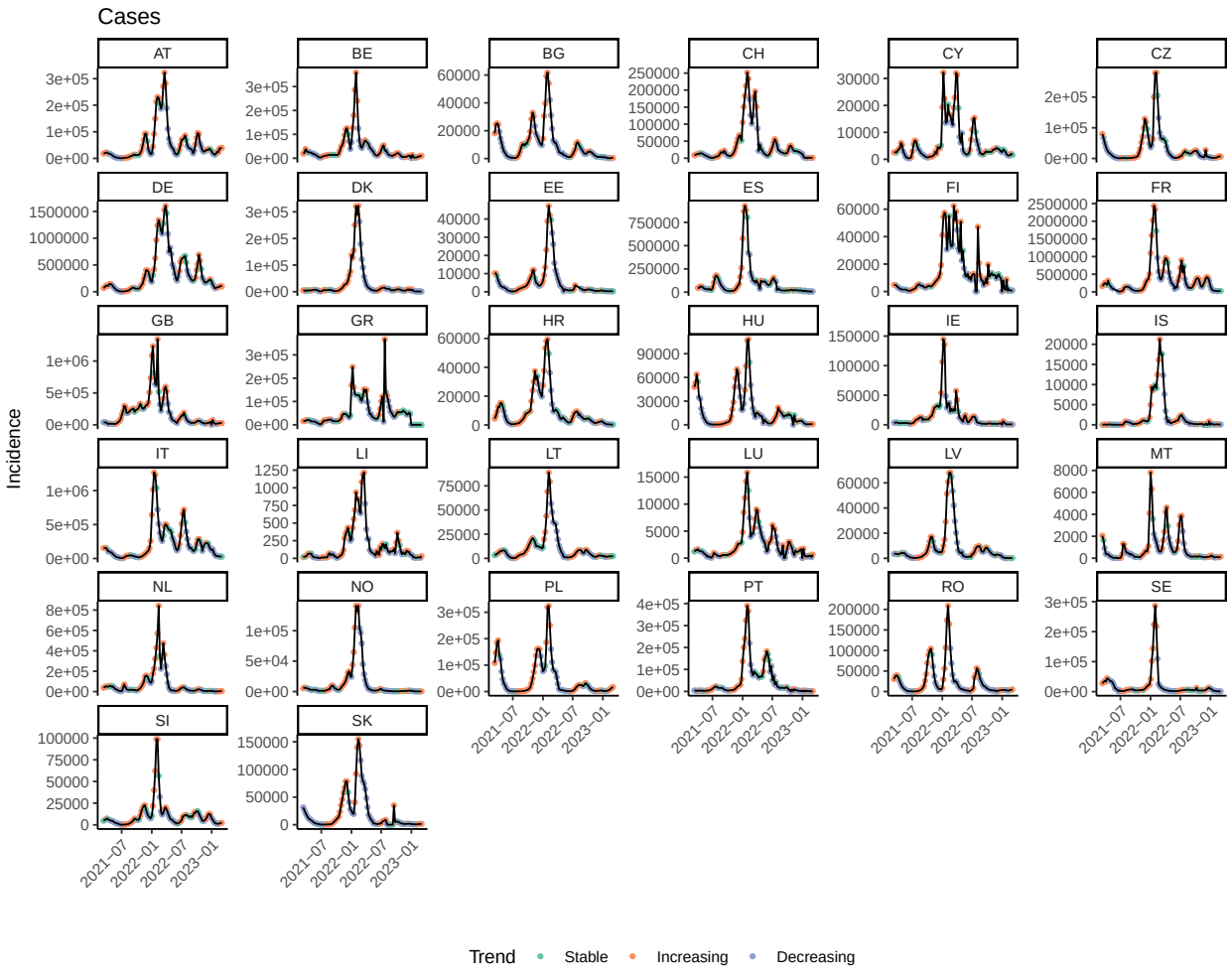


Figure 2: Trends (cases)

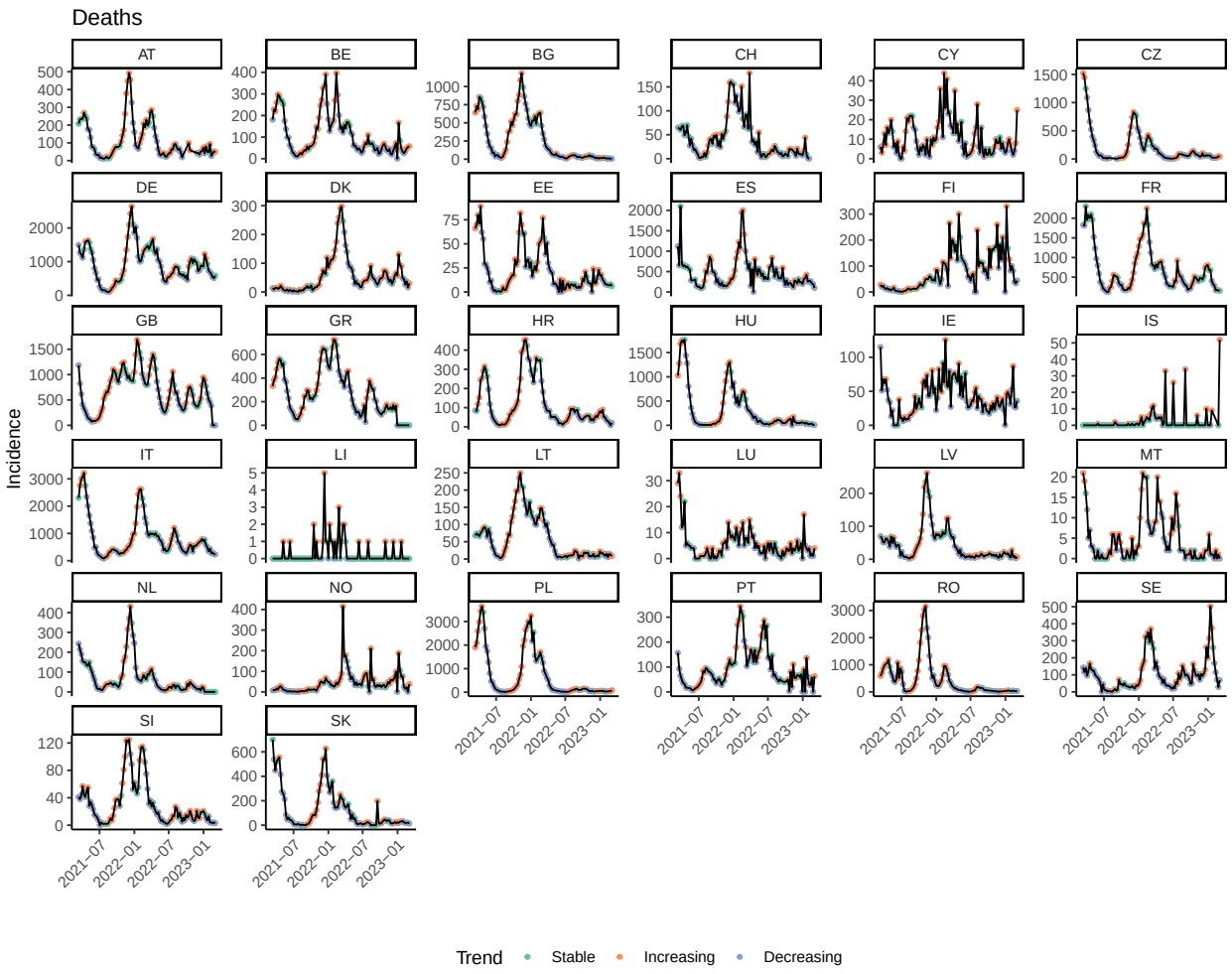


Figure 3: Trends (deaths)

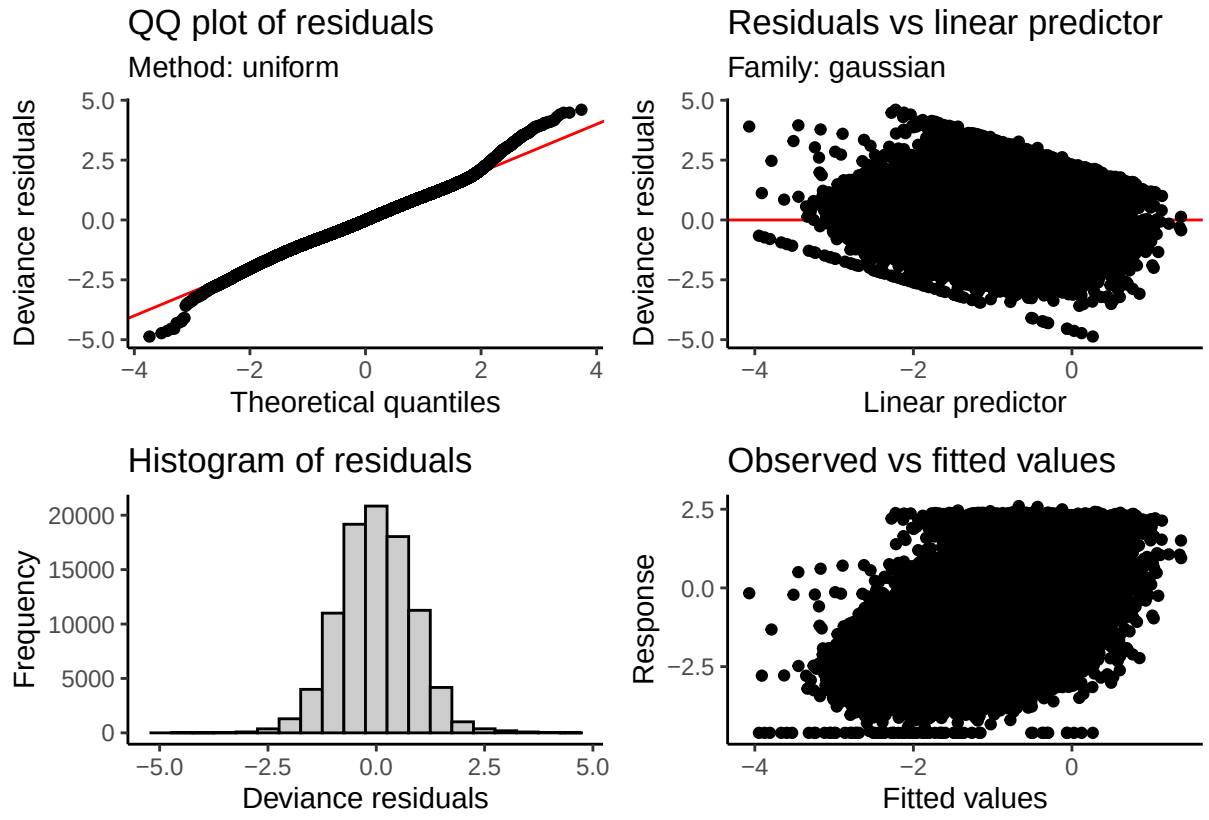
4 Model fitting

4.1 Model formula

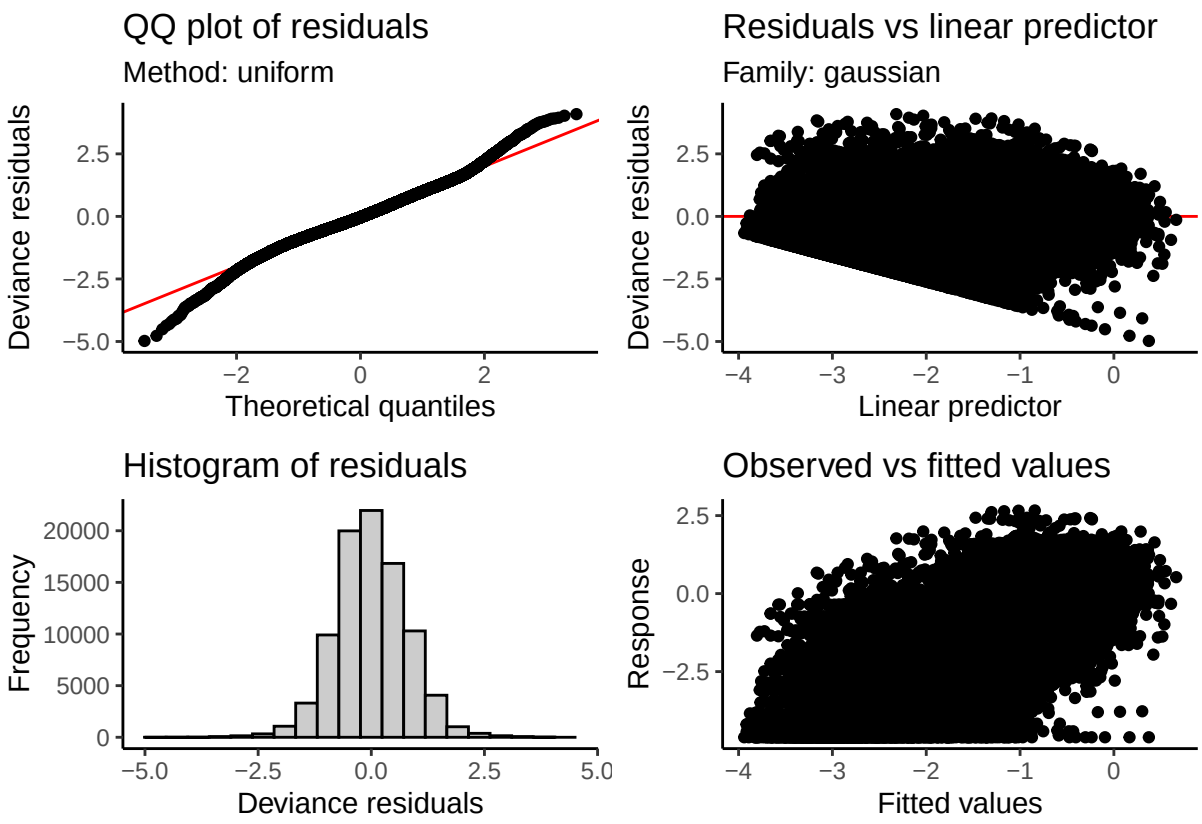
~, log_wis, s(Method, bs = "re") + s(CountryTargets, bs = "re") + s(Trend, bs = "re") + s(location, bs = "re") + s(time, by = location) + s(Horizon, k = 3, by = Model) + s(Model, bs = "re")

4.2 Model diagnostics

4.2.1 Cases



4.2.2 Deaths


QQ plot of residuals
Method: uniform
Deviance residuals
Theoretical quantiles
Residuals vs linear predictor
Family: gaussian
Deviance residuals
Linear predictor
Histogram of residuals
Frequency
Deviance residuals
Observed vs fitted values
Response
Fitted values