Machine Learning Models to Predict 2028 Olympic Track and Field Performances

Clare Robson
Eberly College of Science
The Pennsylvania State University
Centre County, PA
clr5887@psu.edu

Abstract **– This paper's aim is to develop and evaluate machine learning models to predict both winning (first place) and medaling (top three) threshold times for the women's and men's sprint and mid-distance Track and Field events for the 2028 Olympics. Sprint events are defined to be the 100m, 200m, and 400m races. Mid-distance events are defined to be the 800m and 1500m races. Using a dataset of historic top three Olympic results spanning back to 1896, individualized log-linear regression, k-nearest neighbors (KNN), and autoregressive integrated moving average (ARIMA) models were developed and evaluated via normalized root mean squared error (RMSE) and mean absolute error (MAE). The creation of an optimal model for these predictive purposes furthers the field of Track and Field sports prediction, and provides useable metrics for athletes and coaching staff to consider while training.**

The KNN model presented with on average the lowest normalized RMSE and normalized MAE compared to the other considered models, at 0.0154199 and 0.01293649 respectively, demonstrating that on average the KNN model's predictions were approximately 1.129% away from the actual historical values. 2028 Olympic predictions and respective 95% confidence intervals were generated and plotted using the KNN model.

Keywords **– Olympic Games, Track and Field, performance prediction, machine learning, trend analysis, modeling**

I. INTRODUCTION

A. Background

The modern Olympic Games are the world's foremost sporting events, comprised of summer and winter competitions featuring thousands of athletes from across the globe [7]. This largely commercialized event is viewed by billions, with the 2024 Summer Olympics wracking up approximately 5 billion viewers [9]. Athletes and coaches face huge expectations from their home countries to earn medals, and nations invest heavily in training programs and sports analytics to optimize outcomes.

The Track and Field Athletics at the Olympic Games features a variety of running and field events. This work focuses on sprint and mid-distance events, for women and for men. Sprint events are defined to include the 100m, 200m, and 400m races, and mid-distance events are defined to include the 800m and 1500m races. Many of these races are among the most watched Olympic competitions, and are incredibly competitive, as medal placement can be determined by fractions of seconds.

These events were selected for this project in order to narrow the scope of the project, and to simplify finding the optimal machine learning model. The spread between the 100m sprint and the 1500m race is much narrower than between the 100m and 5000m, the next event in the lineup, and hurdle events, while also classified as sprint events, are different in strategy and final results time spread.

Each of these events have been around since either the first or second Olympics, providing 29 – 30 Olympics worth of data, whereas any other event at the longest has been a part of the Olympics for 26 games (besides the Marathon, which is technically apart of Track and Field at the Olympics). While this is not a huge difference, it would make a difference in the creation of each machine learning model. Therefore, events analyzed are narrowed to sprint and mid-distance events, excluding hurdles.

When training for these events, performance prediction is incredibly important for athletes, coaching staff, and support staff, to dually gain an insight as to what times can be anticipated from medalists (the top 3 finishers), and to develop goal-oriented training plans. Accurate predictive modeling can assist with resource allocation for athletes and teams, allowing those qualified for and competing in multiple events to focus their training on where they're likely to perform the best. For national teams and sports scientists, modeling supports long term strategic development and to help forecast realistic goals.

Despite advances in sports analytics, existing research into Track and Field performance prediction is limited, and research into Track and Field Olympics performance prediction is even more limited. As it stands, current research tends to focus on single events, oftentimes just the marathon or the 100m sprint, and within that, only on men's performances rather than women's.

Additionally, much current research uses environmental conditions or an athlete's attributes like season best, country of origin, and personal record for prediction, which, while incredibly useful, still leaves a gap for performance prediction accounting for long term historical patterns. This research aims for fill that gap by developing models based upon historical performance data, that can be applied to multiple events and both men and women's performances. The generalizability of these model's designs is purposeful, so that the optimal one can be easily used as a tool moving forward for this range of events.

B. Literature Review

In developing this work, much prior work in sports performance prediction was first examined to gain an understanding of useful methodologies. This research was helpful in the selection and evaluation of this work's machine learning models, because the pros and cons of each solution could be identified. Additionally, this prior work demonstrated where there was room for growth in the field of Track and Field performance prediction.

One paper examined is Szmygin et al. (2023) [1], wherein a multi-layer perceptron (MLP) model is proposed to predict men's 100m sprint times. Researchers analyzed race data from every historical man's 100m performance faster than 10.55 seconds. The model incorporates both race conditions, such as wind speed and location of the race, and athlete specific attributes, such as their birth country, age, and personal best record.

The approach demonstrates how deep learning can capture non-linear relationships in performance data. Additionally though, it also highlights the challenges that come from using detailed individual level features that are not consistently or easily available for all events or historical time periods. With the intent of focusing on women's races as well as men's, and on more events than the 100m, this specific research was designed to focus on aggregated performance data. These detailed level features were not always recorded for women, or for events longer than the 100m, therefore, generalizability is maintained by removing the MLP model from the list of potential models.

Efiong et al (2019) [3], is another paper that focuses on Track and Field performance prediction. Like Szmygin et al, it also focuses on the men's 100m sprint. Using past Olympic results, this paper proposes a MATLAB simulated model to predict likely Olympic gold medalist times for the men's 100m sprint for the next 5 Olympic competitions.

Unlike Szmygin's deep learning model, Eifong's approach is focused on long term trends, demonstrating accurate how predictive modeling can be with historical trend analysis. This finding influenced the inclusion of ARIMA models in this work, as ARIMA models can be used to analyze patterns and trend over time, much like the MATLAB simulated model was designed to do.

Crowley et al. (202) was one of the most important papers in the development of this work. While it doesn't focus on Track and Field, it does focus on predictive modeling in elite swimming, and is still useful for the development of Track and Field performance prediction. Swimming competitions are set up much the same as Track and Field, with events of different lengths, conducted in heats of 8, with the same events for women and for men, and the paper's overall goal is to develop result predictions based upon historical data, the exact same goal of this work.

The model discussed in the paper predicted winning times for the Tokyo 2024 Olympics using historical trends and event level features, and shows how historical trend modeling can yield robust predictions even with limited input features. Crowley et al. breaks down their predictions by performance categories (1^{st} – 3^{rd} place, 4^{th} – 8^{th} place, and 9^{th} – 16^{th} place), and then uses linear regression and forecasting models to examine trends and predictions.

This paper's work doesn't focus on trends in the same way that Crowley et al. does, but does take influence from the predictions based on performance categories, and creates predictions for first place performances and third place performances. Additionally, it takes influence from how Crowley et al. generated and plotted their 95% confidence intervals.

C. Objective Statement

Methods and insights from each of the above papers informed the modeling framework in this study. In particular, Crowley et al's approach of using previous times to predict performances by categories was useful to examine in the creation of a machine learning model to predict winning and medaling performances in Track and Field, by testing multiple modeling approaches. By using this cross-disciplinary approach, this project introduces a more generalizable framework for predicting Olympic medaling performances. Focus was also put on reproducibility.

Therefore, this paper aims to develop and evaluate several machine learning models to predict gold and bronze medal times for the Los Angeles 2028 Olympic Games, for the men's and women's 100m, 200m, 400m, 800m, and 1500m events, in order to select the most effective one. This model will then be able to generalize performance prediction across multiple events and genders, contributing to the advancement of predictive analytics in Track and Field.

II. METHODOLOGY

A. *Data Description and Preprocessing*

For this research, a dataset containing first, second, and third place results for all Track and Field Olympic events was used [6]. This dataset contains the variables gender, event, location, year, medal, athlete name, nationality, and result, and spans from 1896 to 2016. Women's results begin in 1928, when the women's100m and 800m races were added, with the other women's races being added in later years.

Results from the Tokyo 2020 Olympics [8] and the Paris 2024 [10] were manually added. Due to the Covid-19 Pandemic, the Tokyo games were postponed until 2021. Therefore, the year listed for those races is 2021, as opposed to 2020.

For the purposes of this work, unnecessary variables were removed from the dataset. Variables retained included Gender, Event, Year, Medal, and Result. Any results from a race resulting in a silver medal were removed from the dataset, as using bronze medal race results provides a medaling threshold. Predicting times needed to win a bronze medal will provide a medaling threshold, which upon hit is predicted to result in an Olympic medal. Events other than the listed sprint and mid-distance events were also removed from the dataset. The final large dataset included 458 observations.

Using this large dataset, smaller datasets for each of the sprint and mid-distance events were created. That way, events could be analyzed separately, as well as all together. From there, those smaller datasets were split even further into datasets containing either gold medal results or bronze medal results. This resulted in the creation of 20 individual datasets.

B. *Model Creation*

Using an 80/20 training testing split, training and testing datasets were created from each of the individual datasets to be used in the training of four machine learning models.

1. Model 1: General Log-Linear Regression

This model uses the log of race times, but it includes year, gender (male), medal type (gold or bronze), and event as predictors. It's applied to all races together.

log(Result) ~ Year + Male + Medal + Event

- Model 2: Individualized Log-Linear Regression

This model also uses the log of race times and year as the only predictor. It is applied separately to each race, so each gender and medal specific event is fitted independently from other events.

log(Result) ~ Year

- Model 3: Individualized K-Nearest Neighbors (KNN), k = 3

This model uses the log of race times and finds the closest neighbors to make predictions. Like the individualized linear regression model, it is applied separately to each race.

log(Result) ~ Year

- Model 4: Individualized Autoregressive Integrated Moving Average (ARIMA)

This time series model uses race times directly and captures trends over time. It is applied separately to each race to predict future times, and optimal ARIMA parameters are automatically selected based upon Akaike Information Criteria (AIC).

Result ~ ARIMA(p, d, q)

p = number of lagged observations included

d = number of differences to make the series stationary

q = number of past forecast errors included

Upon visually examining histograms of race results, it was determined that the majority were right-skewed and not normally distributed. Examples are shown in Figures A1 and A2 (other histogram examples can be found in Appendix B).

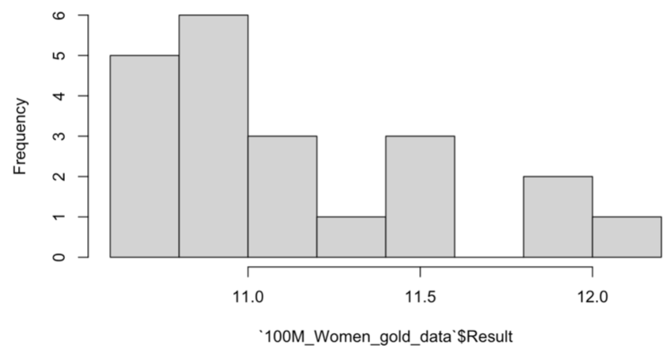

Fig. A1 Histogram of women's 100m gold medal Olympic finals race results

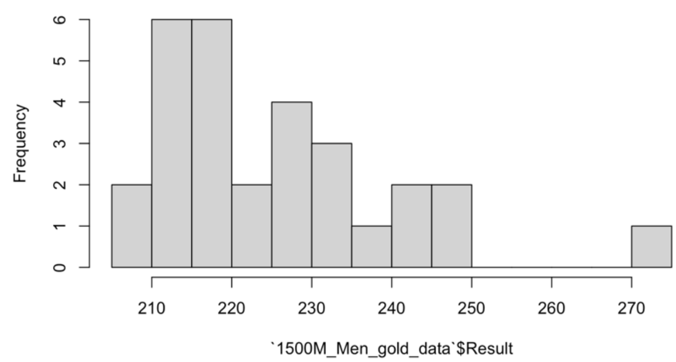

Fig. A2 Histogram of men's 1500m gold medal Olympic finals race results

Additionally, plots of historical gold medal and bronze medal times for each race display the improvement of times over the years at a decreasing rate. The relationship between result and year is non-linear. Plot examples are displayed in Figures A3, A4, A5, and A6 (remaining figures for each race can be found in Appendix C).

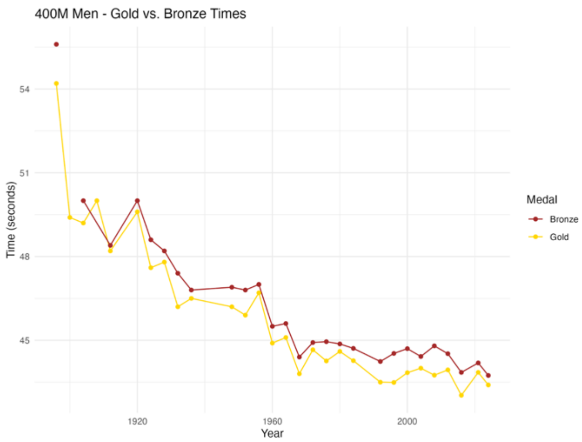

Fig. A3 400m men's Olympic historical trends

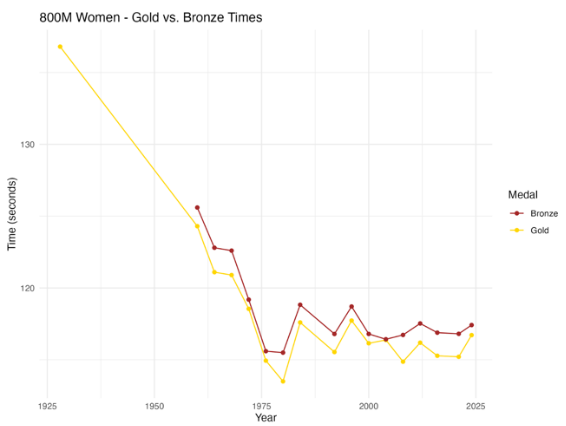

Fig. A4 800m women's Olympic historical trends

Fig. A5 200m women's Olympic historical trends

These factors limited the machine learning models that could be used, as well as how they could be implemented. Specifically, for the linear regression models and KNN model, a log transformation was applied to the dependent variable time, so that the impact of extreme values could be reduced, and model performance could be improved. The log transformation ensures that percentage improvements in time are treated consistently, and linearizes displayed trends.

The ARIMA model is built to handle non-normal data, and therefore the log transformation was not applied.

C. *Model Evaluation*

These models were evaluated via root mean squared error (RMSE), mean absolute error (MAE), normalized RMSE (NRMSE), and normalized MAE (NMAE). These normalized metrics are created by dividing RMSE and MAE by average race time. For methods 1 through 3, where the regression is based on the log transformation of the results, RMSE and MAE are calculated based on the original result values, rather than the log transformed ones. This way, the normalized values can be compared across all four methods.

This normalization is necessary not just for valid comparison across all four methods, but for comparison within each method from event to event. For example, a MAE of 1.00 seconds might be significant for the men's 100m race, where at an Olympic level this race is now run below 10 seconds, but not for the women's 1500m, which was most recently won in 231 seconds. Each value shows approximately what percentage, relative to average race time, each models' predictions are off by, with NRMSE being more sensitive to large prediction errors than normalized MAE.

Raw results for Models 2, 3, and 4 are displayed in Table A1, Table A2, and Table A3. Model 1's raw results ended up not being comparable, due to the resulting relatively large RMSE and MAE (3.00689 and 1.603443 respectively) and difficulty in normalizing each statistic for comparison. Therefore, these results are not displayed or compared, though the model summary can be found in Appendix E as Figure E1.

D. *Raw Results*

Raw results for each method include RMSE, MAE, NRMSE, NMAE, 2028 Olympic time predictions, and values for the lower and upper confidence interval bounds for each of the events' models. In each Medal column, "B" refers to bronze and "G" refers to gold. Tables for models 2, 3, and 4 are displayed in Tables A1 – 3.

TABLE A1

MODEL 2 RAW RESULTS (INDIVIDUALIZED LOG-LINEAR REGRESSION MODEL)

Race	Medal	LM_RMSE	LM_MAE	Normalized_LM_RMSE	Normalized_LM_MAE	LM_2028_Predictions	Lower_95_CI	Upper_95_CI
100M_Men	B	0.257167924852734	0.219206866079081	0.0246471079981535	0.021008900333437	9.5618983397633	9.2672156523184	9.8565810272082
100M_Men	G	0.144756460116676	0.115014396295956	0.0141755958984831	0.0112630386449443	9.54628854283743	9.26567404918832	9.82690303648653
100M_Women	B	0.195514319556856	0.160804713677737	0.0175545965932082	0.0144381336635454	10.5145669949238	10.1875132605335	10.841620729314
100M_Women	G	0.24904028103519	0.218284576169658	0.0220039124434697	0.0192864972759903	10.5592347966405	10.0149949543908	11.1034746388902
1500M_Men	B	4.42122027640097	3.86033779645916	0.0202148351013174	0.0176503515120783	206.694348627407	198.300916184413	215.087781070402
1500M_Men	G	5.22094155653914	3.68262284180747	0.0238586177439202	0.016828820956187	206.602941617946	196.8645654138	216.341317822093
1500M_Women	B	5.46721570493724	5.23263837348013	0.0228107966493912	0.0218320725427874	242.618119245031	231.956315980245	253.279922509817
1500M_Women	G	4.89462166506772	4.50623188735836	0.0200761064180188	0.0184830610219645	236.594236841733	232.007337181082	241.181136502384
200M_Men	B	0.210726813650311	0.191439851015374	0.0103216503551289	0.00937695195020444	19.3939115651676	18.963039229018	19.8247839013172
200M_Men	G	0.378423025619622	0.319723350216517	0.0184281970109385	0.0155696786080602	19.1133722935673	18.3009884960994	19.9257560910353
200M_Women	B	0.406423878313809	0.389407395070521	0.0180934392126348	0.0173358884839408	21.3448518958495	20.4613760355579	22.2283277561412
200M_Women	G	0.55472507890603	0.410510096244335	0.0244911734616349	0.0181240660593525	21.3354533591164	20.1106879503887	22.560218767844
400M_Men	B	0.765039326476235	0.666309780281411	0.0172027191598362	0.0149826807942393	42.582577028888	40.9204135669215	44.2447404908545
400M_Men	G	0.563439673275903	0.477350209093267	0.0122062320900326	0.010341209035816	42.2307160056692	41.048809226813	43.4126227845254
400M_Women	B	1.08461875348858	0.941772668223535	0.0216692611911677	0.0188153836219406	48.7845321341077	47.0572120665469	50.5118522016685
400M_Women	G	0.528913879973793	0.427263999038499	0.0108458075182596	0.00876139437536225	48.2402631353463	47.491879152356	48.9886471183366
800M_Men	B	2.9959721391005	2.83999621946232	0.0277524745794525	0.0263076287852373	99.0555181816166	94.1898099058937	103.92122645734
800M_Men	G	2.97376388446059	2.71098968322628	0.0282435547959027	0.0257478362923951	99.0854028380978	92.7307569838357	105.44004869236
800M_Women	B	1.18392320203718	1.13065367908969	0.0100810899356027	0.00962750067344767	115.06447845984	112.229554123464	117.899402796217
800M_Women	G	2.2510190393769	2.11324995906544	0.0193423904051634	0.0181585784113376	110.886450422546	109.131531427064	112.641369418028

TABLE A2

MODEL 3 RAW RESULTS (INDIVIDUALIZED KNN MODEL, K = 3)

Race	Medal	KNN_RMSE	KNN_MAE	Normalized_KNN_RMSE	Normalized_KNN_MAE	KNN_2028_Predictions	Lower_95_CI	Upper_95_CI
100M_Men	B	0.445808498558361	0.332469711991185	0.0427265189340963	0.0318640705377789	9.86990527220055	9.06341959196828	10.6763909524328
100M_Men	G	0.205453535733249	0.161897597625625	0.0201194910135383	0.0158541796271217	9.79999659863828	9.49084771376397	10.1091454835126
100M_Women	B	0.118063845341451	0.0930838781443084	0.0106005697276275	0.00835769949668314	10.9520445547334	10.7876775756741	11.1164115337926
100M_Women	G	0.237914885241841	0.206236368145297	0.0210209299559853	0.0182219798679358	10.6798842726171	10.16949559131	11.1902729539242
1500M_Men	B	1.95712388624085	1.73892880863034	0.00894842015549017	0.00795078211936725	215.456117963496	212.267707271272	218.644528655719
1500M_Men	G	4.77307946738427	4.20707913029571	0.0218119810843398	0.019225477186664	218.928643961599	208.89911598753	228.958171935668
1500M_Women	B	5.69337001899834	5.54387867824461	0.0237543775044087	0.0231306566273088	241.529119229364	229.833354397033	253.224884061695
1500M_Women	G	5.42560784739696	5.13780895023414	0.0222540347304397	0.0210735795938016	237.643451255749	233.457940410708	241.82896210079
200M_Men	B	0.164710644196389	0.137829874006633	0.00806772355977612	0.00675107141490172	19.9325668272224	19.6701047998858	20.195028854559
200M_Men	G	0.255874928140973	0.196247971424366	0.0124604299070354	0.00955675536519924	19.466280946153	18.9310391453671	20.001522746939
200M_Women	B	0.37138518693556	0.290368624854269	0.0165335642486616	0.0129268169105963	22.1165031690561	21.3631527662783	22.8698535718339
200M_Women	G	0.370978124108981	0.277117699198087	0.0163787251262243	0.0122347770065381	21.7129357361871	21.0004839417742	22.4253875306001
400M_Men	B	0.25315511012148	0.196749413979556	0.00569246065212897	0.00442411886084629	44.3510578312253	43.8345462980426	44.8675693644081
400M_Men	G	0.728645950610955	0.54139700016981	0.0157852242333396	0.0117287045097446	43.72936053515	42.2395237158806	45.2191973544194
400M_Women	B	0.535053194786558	0.524980811307212	0.0106896615900351	0.010488428569004	49.4287185944189	49.1806751658298	49.676762023008
400M_Women	G	0.411127053613861	0.295734798192441	0.00843049323883515	0.00606428157605827	49.0493122671432	48.1441206428733	49.9545038914131
800M_Men	B	1.07330731635357	0.99931888145097	0.00994232677410214	0.00925695252378469	103.260876357972	100.971230009641	105.550522706302
800M_Men	G	1.88887121827213	1.65363162862639	0.0179397019495881	0.0157054955705802	101.415279292206	97.4761330998081	105.354425484605
800M_Women	B	0.889746095077839	0.793222282172702	0.0075761758777064	0.00675427692585747	117.279666846518	115.161550769848	119.397782923188
800M_Women	G	0.892045549908214	0.833224788461841	0.00766510321933547	0.00715967251798536	115.620618765214	114.186613850382	117.054623680047

TABLE A3

MODEL 4 RAW RESULTS (INDIVIDUALIZED ARIMA MODEL)

Race	Medal	ARIMA_RMSE	ARIMA_MAE	Normalized_ARIMA_RMSE	Normalized_ARIMA_MAE	ARIMA_2028_Predictions	ARIMA_95CI_Lower	ARIMA_95CI_Upper
100M_Men	B	0.527603235027224	0.507416666666778	0.0505657691227932	0.0486310778864077	10.5070833333339	9.04040929534934	11.9737573713184
100M_Men	G	0.239818837842808	0.191376811594251	0.0234847890820442	0.0187409967286683	10.3091304347828	9.17235265942155	11.445908210144
100M_Women	B	0.484020031222904	0.474062500000028	0.0434585886619891	0.0425645342312034	11.3231250000001	10.4793316709897	12.1669183290104
100M_Women	G	0.536526225477374	0.383875	0.0474046850571986	0.0339172115214703	11.109375	10.3038808695388	11.9148691304612
1500M_Men	B	9.62793053890388	8.99823061383817	0.0440211109249036	0.0411419781622905	226.919828651377	196.624874860707	257.214782442047
1500M_Men	G	14.1557423326821	13.2548550724638	0.0646888001980949	0.0605719326677553	227.224782608696	198.10480228428	256.344762933111
1500M_Women	B	5.58253729648687	5.34800000000001	0.0232919515033595	0.0223133944341683	242.894	232.253743207615	253.534256792385
1500M_Women	G	5.55451179083892	5.13888888888896	0.0227827557355201	0.0210780091421595	238.664444444444	227.986934132436	249.341954756453
200M_Men	B	0.717181842992643	0.624857142857132	0.0351284209929782	0.0306062472010742	20.8042857142857	19.0115713435132	22.5970000850581
200M_Men	G	0.915177578396674	0.848333333333332	0.0445667191817226	0.0413115818521223	20.55	18.6418574893635	22.4581425106365
200M_Women	B	0.627868616829336	0.518499999999949	0.0279518582895642	0.0230829159710402	22.861999999999	20.7797917595489	24.9442082404491
200M_Women	G	1.0846832367369	0.957992655453585	0.0478888846241457	0.0422954814769795	22.5144982258867	21.434806866578	23.5941895851953
400M_Men	B	2.27259637663959	2.23582608695654	0.0511017353984438	0.0502749165082871	46.7078260869565	41.458574434286	51.9570777396271
400M_Men	G	2.62946689583073	2.30666666666667	0.0569641875179967	0.0499711149624495	45.8665217391305	40.6591012360509	51.07394224221
400M_Women	B	1.14651646303051	1.12333333333333	0.0229058963045521	0.0224427277570591	50.25	47.7619613592663	52.7380386407337
400M_Women	G	0.992790679515658	0.943333333333243	0.020357977023561	0.0193438140806543	49.7099999999999	47.1542053244804	52.2657946755195
800M_Men	B	4.74331234667723	4.57968253968255	0.0439385445563876	0.0424227987990108	109.399047619048	90.1443282121668	128.653767025928
800M_Men	G	4.70104874625333	4.24283806600431	0.0446485777020926	0.0402966859721181	111.499667575102	98.1577508187458	124.841584331458
800M_Women	B	5.23756304655768	5.15999999999999	0.0445977779849939	0.0439373297002724	122.6	116.902397275788	128.297602724212
800M_Women	G	2.71755538095147	2.56480769231112	0.02335120947736	0.02203869040245	118.942307692311	106.833935422253	131.050679962369

III. RESULTS

A. Metric Comparisons

Mean NRMSE, minimum NRMSE, maximum NRMSE, NRMSE range, mean NMAE, minimum NMAE, maximum NMAE, and NMAE range are displayed in Table A4.

TABLE A4

SUMMARY OF NORMALIZED RMSE AND MAE METRICS ACROSS MODELS 2, 3 AND 4

	Mean Normalized RMSE	Min. Normalized RMSE	Max. Normalized RMSE	Range Normalized RMSE	Mean Normalized MAE	Min. Normalized MAE	Max. Normalized MAE	Range Normalized MAE
Model 2	0.0192	0.01008	0.02824	0.01816	0.0167	0.00876	0.02631	0.01755
Model 3	0.01542	0.00569	0.04273	0.03703	0.01294	0.00442	0.03186	0.02744
Model 4	0.03916	0.02036	0.06469	0.04433	0.03585	0.01874	0.06057	0.04183

Model 3 has both the lowest mean NRMSE and mean NMAE values, at 0.01542 and 0.01294 respectively. Model 2 has the smallest ranges of values, with a 0.01816 range for NRMSE, and a 0.01755 range for NMAE.

Mean values can also be compared for each race, between each model. These results are displayed in Table A5, with the highlighted values being the lowest per race between each model.

TABLE A5

NORMALIZED RMSE AND MAE COMPARISON FOR EACH RACE

Medal	Race	Normalized_LM_RMSE	Normalized_KNN_RMSE	Normalized_ARIMA_RMSE	Normalized_LM_MAE	Normalized_KNN_MAE	Normalized_ARIMA_MAE	lowest_RMSE	lowest_MAE
B	100M_Men	0.0246471079981535	0.0427265189340963	0.0505657691227932	0.021008900333437	0.0318640705377789	0.0486310778864077	LM	LM
G	100M_Men	0.0141755958984831	0.0201194910135383	0.0234847890820442	0.0112630386449443	0.0158541796271217	0.0187409967286683	LM	LM
B	100M_Women	0.0175545965932082	0.0106005697276275	0.0434585886619891	0.0144381336635454	0.00835769949668314	0.0425645342312034	KNN	KNN
G	100M_Women	0.0220039124434697	0.0210209299559853	0.0474046850571986	0.0192864972759903	0.0182219798679358	0.0339172115214703	KNN	KNN
B	1500M_Men	0.0202148351013174	0.00894842015549017	0.0440211109249036	0.0176503515120783	0.00795078211936725	0.0411419781622905	KNN	KNN
G	1500M_Men	0.0238586177439202	0.0218119810843398	0.0646888001980949	0.016828820956187	0.019225477186664	0.0605719326677553	KNN	LM
B	1500M_Women	0.0228107966493912	0.0237543775044087	0.0232919515033595	0.0218320725427874	0.0231306566273088	0.0223133944341683	LM	LM
G	1500M_Women	0.0200761064180188	0.0222540347304397	0.0227827557355201	0.0184830610219645	0.0210735795938016	0.0210780091421595	LM	LM
B	200M_Men	0.0103216503551289	0.00806772355977612	0.0351284209929782	0.00937695195020444	0.00675107141490172	0.0306062472010742	KNN	KNN
G	200M_Men	0.0184281970109385	0.0124604299070354	0.0445667191817226	0.0155696786080602	0.00955675536519924	0.0413115818521223	KNN	KNN
B	200M_Women	0.0180934392126348	0.0165335642486616	0.0279518582895642	0.0173358884839408	0.0129268169105963	0.0230829159710402	KNN	KNN
G	200M_Women	0.0244911734616349	0.0163787251262243	0.0478888846241457	0.0181240660593525	0.0122347770065381	0.0422954814769795	KNN	KNN
B	400M_Men	0.0172027191598362	0.00569246065212897	0.0511017353984438	0.0149826807942393	0.00442411886084629	0.0502749165082871	KNN	KNN
G	400M_Men	0.0122062320900326	0.0157852242333396	0.0569641875179967	0.010341209035816	0.0117287045097446	0.0499711149624495	LM	LM
B	400M_Women	0.0216692611911677	0.0106896615900351	0.0229058963045521	0.0188153836219406	0.010488428569004	0.0224427277570591	KNN	KNN
G	400M_Women	0.0108458075182596	0.00843049323883515	0.020357977023561	0.00876139437536225	0.00606428157605827	0.0193438140806543	KNN	KNN
B	800M_Men	0.0277524745794525	0.00994232677410214	0.0439385445563876	0.0263076287852373	0.00925695252378469	0.0424227987990108	KNN	KNN
G	800M_Men	0.0282435547959027	0.0179397019495881	0.0446485777020926	0.0257478362923951	0.0157054955705802	0.0402966859721181	KNN	KNN
B	800M_Women	0.0100810899356027	0.0075761758777064	0.0445977779849939	0.00962750067344767	0.00675427692585747	0.0439373297002724	KNN	KNN
G	800M_Women	0.0193423904051634	0.00766510321933547	0.02335120947736	0.0181585784113376	0.00715967251798536	0.02203869040245	KNN	KNN

As clearly shown in Tables A4 and A5, Method 3 has the smallest NRMSE and NMAE values on average. Additionally, Method 3 has the lowest NRMSE values for 15/20 of the races, and the lowest NMAE values for 14/20. Method 2 has the lowest NRMSE values for 5/20, and the lowest NMAE values for 6/20. There is no indication that any one model is better suited towards predicting shorter distance or longer distance events than any other. Method 4 has the largest mean values across the board.

B. *Model Selection*

In terms of model selection, it is important to consider both the average NRMSE and MAE values, as well as the range of NRMSE and MAE values. Lower means indicate better model fitting and therefore increased accuracy, while lower ranges indicate more consistent errors, meaning greater stability. Taking that into account, Model 3 presented as more accurate, while Model 2 presented as more stable.

In terms of accuracy, the Model 3 generally outperforms the Model 2, as it exhibited lower NRMSE and NMAE values across most events, suggesting that Model 3's predictions are closer to the actual race times on average. This is likely due to the KNN model's ability to show local patterns for each race, so it has greater flexibility in examining trends for individual races.

In terms of stability, Model 2 tends to be more stable, with less variation in its performance across different events. This is reflected in its smaller range of NRMSE and NMAE values. Model 3, while more accurate, shows slightly higher variability, which suggests it may be more sensitive to the specific data points in each event. Despite this, Model 3's overall accuracy makes it a more reliable choice for prediction, though Model 2's stability provides an advantage in scenarios where consistency across different events is more important.

Using NMAE, as MAE is less prone to influence from outliers than RMSE, Model 3's predictions on average reflect a 1.294% error rate.

C. *2028 Olympic Predictions*

Time predictions along with 95% confidence intervals generated using Model 3, the KNN model, are shown below in Table A6.

TABLE A6

2028 OLYMPIC TIME PREDICTIONS AND 95% CONFIDENCE INTERVALS

Race	Medal	Predictions (secs)	95% Confidence Interval
100M Men	B	9.8699	[9.0634, 10.6764]
	G	9.8000	[9.4908, 10.1091]
100M Women	B	10.9520	[10.7877, 11.1164]
	G	10.6799	[10.1695, 11.1903]
1500M Men	B	215.4561	[212.2677, 218.6445]
	G	218.9286	[208.8991, 228.9582]
1500M Women	B	241.5291	[229.8334, 253.2249]
	G	237.6435	[233.4579, 241.8290]
200M Men	B	19.9326	[19.6701, 20.1950]
	G	19.4663	[18.9310, 20.0015]
200M Women	B	22.1165	[21.3632, 22.8699]
	G	21.7129	[21.0005, 22.4254]
400M Men	B	44.3511	[43.8345, 44.8676]
	G	43.7294	[42.2395, 45.2192]
400M Women	B	49.4287	[49.1807, 49.6768]
	G	49.0493	[48.1441, 49.9545]
800M Men	B	103.2609	[100.9712, 105.5505]
	G	101.4153	[97.4761, 105.3544]
800M Women	B	117.2797	[115.1616, 119.3978]
	G	115.6206	[114.1866, 117.0546]

Confidence intervals were calculated using the following formula:

$$\hat{y}_{2028} \pm 1.96 \times \hat{\sigma}_{\text{residual}}$$

Individual plots were generated for each race, gender, and medal type, displaying historical results, 2028 predictions, and 2028 95% confidence intervals in the same line graph. One example is shown in Figure A6 (the remaining plots can be found in Appendix D).

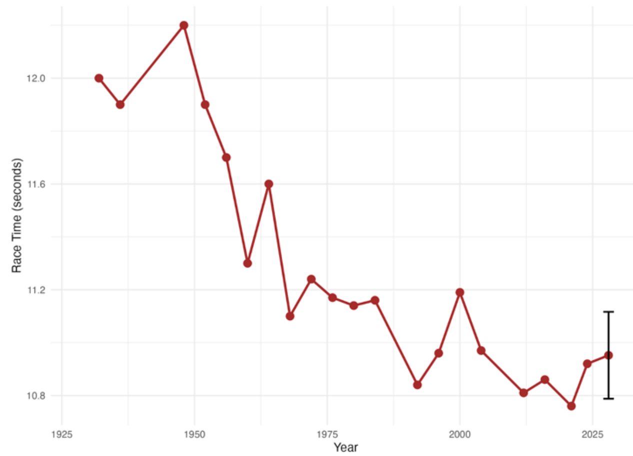

Fig. A6 Women's 100m bronze medalist race times vs year, including 2028 predictions and 95% CI's

III. RESULTS

Even though Model 3, the KNN model, demonstrated strong predicted performance based on NRMSE and NMAE, there are still several limitations to consider. One limitation is the distribution of residuals. Even though log(time) was used in the creation of the model, errors still skewed positive as shown by the QQ plots in Figure A7.

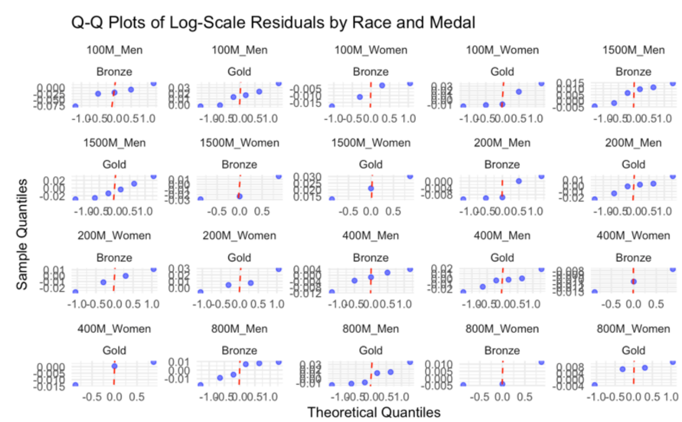

Fig. A7 QQ plots of log-scale residuals by race and medal

The positive skew suggests that the model tends to underpredict fast race times and overpredict slow race times. Therefore, the 95% confidence intervals generated and reported in Table A5 may not be reliable. To improve error distribution, a different transformation and/or additional transformations may be necessary in the future.

Another limitation of the KNN model is its reliance on local similarity, which can make it sensitive to variations in training data. Its predictions are influenced by nearby historical observations, which could lead to instability when predicting future results, like in the case of predicting times for the 2028 Olympics. Additionally, the choice of k = 3 introduces bias-variance tradeoff. While k = 3 was determined to be the optimal choice for this model, adjusting model parameters and transformations in future work may mean a different k value should be used to create a model that isn't too sensitive to noise, but also isn't reducing responsiveness.

IV. CONCLUSIONS

This paper describes the development of 4 machine learning models, designed to predict winning and medaling times for the sprint and mid-distance events at the 2028 LA Olympics. Using historical Olympic results, log-linear regression models, a KNN model, and an ARIMA model were built. After assessing each model via NRMSE, NMAE, and range of NRMSE and NMAE values, Model 3, the KNN model, was assessed to be optimal.

While this work is valuable, there are limitations. Model 3 relies solely on historical results, and does not include any outside factors such as environment or an athlete's training. Future work could expand upon these models by incorporating outside factors, and/or applying additional data transformations to limit positively skewed errors.

Taking into consideration the limitations of this work, future work could also include customizing the chosen k value per each race category. Building this into the model would be a complex solution to help improve accuracy with more regard to the differences for each smaller data set.

Overall methodology provides a framework by which to predict Olympic race performances, for multiple events and for both men and women. This framework can aid athletes, coaches, and analysts in understanding expected medaling and winning thresholds for the 2028 Olympic Games.

REFERENCES

[1] A. Szmygin, M. Wojtowicz, Ż. Świderska-Chadaj and R. Roszczyk, "Prediction of athletes' performance results using machine learning algorithms," 2023 24th International Conference on Computational Problems of Electrical Engineering (CPEE), Grybów, Poland, 2023, pp. 1-5, doi: 10.1109/CPEE59623.2023.10285142.

[2] Crowley, Emmet, Kwok Ng, Iñigo Mujika, and Cormac Powell. "Speeding up or Slowing Down? Analysis of Race Results in Elite-Level Swimming from 2011-2019 to Predict Future Olympic Games Performances." *Measurement in Physical Education and Exercise Science* 26,no. 2 (2022): 130–40. doi:10.1080/1091367X.2021.1952592.

[3] Efiong, John & Olajubu, Emmanuel & Dr.Aranuwa, Felix. (2019). Formulation of Sprint Time Predictive Model for Olympic Athletic Games. International Journal of Information Technology and Computer Science. 4. 33-43. 10.5815/ijitcs.2019.04.04.

[4] Wu, P. P. Y., Garufi, L., Drovandi, C., Mengersen, K., Mitchell, L. J. G., Osborne, M. A., & Pyne, D. B. (2022). Bayesian prediction of winning times for elite swimming events. *Journal of Sports Sciences*, *40*(1), 24–31. https://doi.org/10.1080/02640414.2021.1976485

[5] Mujika, I., Pyne, D. B., Wu, P. P., Ng, K., Crowley, E., & Powell, C. (2023). Next-Generation Models for Predicting Winning Times in Elite Swimming Events: Updated Predictions for the Paris 2024 Olympic Games. *International journal of sports physiology and performance*, *18*(11), 1269–1274. https://doi.org/10.1123/ijspp.2023-0174

[6] Ravaliya, Jay (2017). Olympic Track and Field Results. Retrieved from https://www.kaggle.com/datasets/jayrav13/olympic-track-field-results?resource=download

[7] NBC Olympics (2024). Track and field 101: Olympic history, records and results. Retrieved from https://www.nbcolympics.com/news/track-and-field-101-olympic-history-records-and-results

[8] Olympics (2024). Tokyo 2020 Athletics Results. Retrieved from https://olympics.com/en/olympic-games/tokyo-2020/results/athletics

[9] Olympics (2024). Around 5 billion people - 84 percent of the potential global audience - followed the Olympic Games Paris 2024. Retrieved from https://olympics.com/ioc/news/around-5-billion-people-84-per-cent-of-the-potential-global-audience-followed-the-olympic-games-paris-2024

[10] Runner's World (2024). Results from Track and Field at the 2024 Paris Olympics. Retrieved from https://www.runnersworld.com/races-places/a61776840/2024-paris-olympics-results-track-and-field/

APPENDIX A
ASSETS FROM MAIN BODY

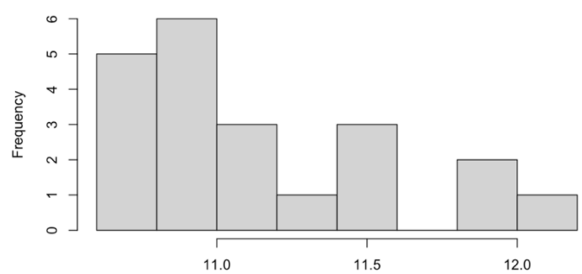

Fig. A1 Histogram of women's 100m gold medal Olympic finals race results

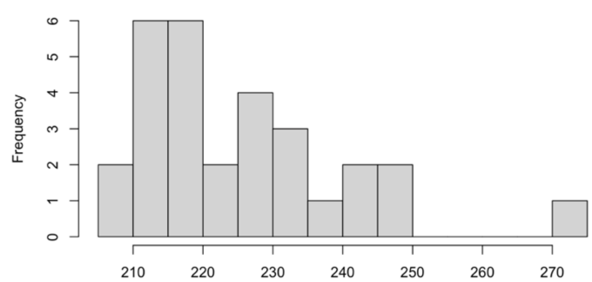

Fig. A2 Histogram of men's 1500m gold medal Olympic finals race results

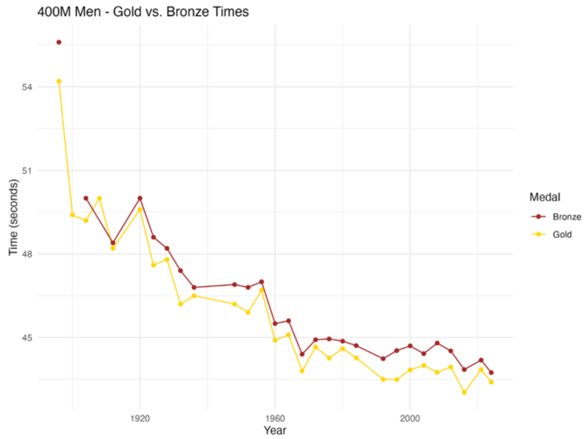

Fig. A3 400m men's Olympic historical trends

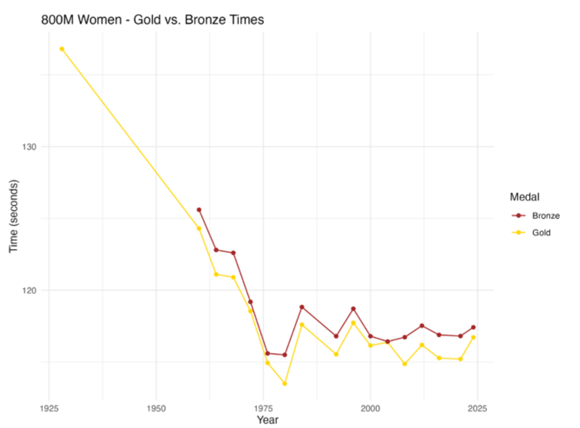

Fig. A4 800m women's Olympic historical trends

Fig. A5 200m women's Olympic historical trends

TABLE A1

MODEL 2 RAW RESULTS (INDIVIDUALIZED LOG-LINEAR REGRESSION MODEL)

Race	Medal	LM_RMSE	LM_MAE	Normalized_LM_RMSE	Normalized_LM_MAE	LM_2028_Predictions	Lower_95_CI	Upper_95_CI
100M_Men	B	0.257167924852734	0.219206866079081	0.0246471079981535	0.021008900333437	9.5618983397633	9.2672156523184	9.8565810272082
100M_Men	G	0.144756460116676	0.115014396295956	0.0141755958984831	0.0112630386449443	9.54628854283743	9.26567404918832	9.82690303648653
100M_Women	B	0.195514319556856	0.160804713677737	0.0175545965932082	0.0144381336635454	10.5145669949238	10.1875132605335	10.841620729314
100M_Women	G	0.24904028103519	0.218284576169658	0.0220039124434697	0.0192864972759903	10.5592347966405	10.0149949543908	11.1034746388902
1500M_Men	B	4.42122027640097	3.86033779645916	0.0202148351013174	0.0176503515120783	206.694348627407	198.300916184413	215.087781070402
1500M_Men	G	5.22094155653914	3.68262284180747	0.0238586177439202	0.016828820956187	206.602941617946	196.8645654138	216.341317822093
1500M_Women	B	5.46721570493724	5.23263837348013	0.0228107966493912	0.0218320725427874	242.618119245031	231.956315980245	253.279922509817
1500M_Women	G	4.89462166506772	4.50623188735836	0.0200761064180188	0.0184830610219645	236.594236841733	232.007337181082	241.181136502384
200M_Men	B	0.210726813650311	0.191439851015374	0.0103216503551289	0.00937695195020444	19.3939115651676	18.963039229018	19.8247839013172
200M_Men	G	0.378423025619622	0.319723350216517	0.0184281970109385	0.0155696786080602	19.1133722935673	18.3009884960994	19.9257560910353
200M_Women	B	0.406423878313809	0.389407395070521	0.0180934392126348	0.0173358884839408	21.3448518958495	20.4613760355579	22.2283277561412
200M_Women	G	0.55472507890603	0.410510096244335	0.0244911734616349	0.0181240660593525	21.3354533591164	20.1106879503887	22.560218767844
400M_Men	B	0.765039326476235	0.666309780281411	0.0172027191598362	0.0149826807942393	42.582577028888	40.9204135669215	44.2447404908545
400M_Men	G	0.563439673275903	0.477350209093267	0.0122062320900326	0.010341209035816	42.2307160056692	41.048809226813	43.4126227845254
400M_Women	B	1.08461875348858	0.941772668223535	0.0216692611911677	0.0188153836219406	48.7845321341077	47.0572120665469	50.5118522016685
400M_Women	G	0.528913879973793	0.427263999038499	0.0108458075182596	0.00876139437536225	48.2402631353463	47.491879152356	48.9886471183366
800M_Men	B	2.9959721391005	2.83999621946232	0.0277524745794525	0.0263076287852373	99.0555181816166	94.1898099058937	103.92122645734
800M_Men	G	2.97376388446059	2.71098968322628	0.0282435547959027	0.0257478362923951	99.0854028380978	92.7307569838357	105.44004869236
800M_Women	B	1.18392320203718	1.13065367908969	0.0100810899356027	0.00962750067344767	115.06447845984	112.229554123464	117.899402796217
800M_Women	G	2.2510190393769	2.11324995906544	0.0193423904051634	0.0181585784113376	110.886450422546	109.131531427064	112.641369418028

TABLE A2

MODEL 3 RAW RESULTS (INDIVIDUALIZED KNN MODEL, K = 3)

Race	Medal	KNN_RMSE	KNN_MAE	Normalized_KNN_RMSE	Normalized_KNN_MAE	KNN_2028_Predictions	Lower_95_CI	Upper_95_CI
100M_Men	B	0.445808498558361	0.332469711991185	0.0427265189340963	0.0318640705377789	9.86990527220055	9.06341959196828	10.6763909524328
100M_Men	G	0.205453535733249	0.161897597625625	0.0201194910135383	0.0158541796271217	9.79999659863828	9.49084771376397	10.1091454835126
100M_Women	B	0.118063845341451	0.0930838781443084	0.0106005697276275	0.00835769949668314	10.9520445547334	10.7876775756741	11.1164115337926
100M_Women	G	0.237914885241841	0.206236368145297	0.0210209299559853	0.0182219798679358	10.6798842726171	10.16949559131	11.1902729539242
1500M_Men	B	1.95712388624085	1.73892880863034	0.00894842015549017	0.00795078211936725	215.456117963496	212.267707271272	218.644528655719
1500M_Men	G	4.77307946738427	4.20707913029571	0.0218119810843398	0.019225477186664	218.928643961599	208.89911598753	228.958171935668
1500M_Women	B	5.69337001899834	5.54387867824461	0.0237543775044087	0.0231306566273088	241.529119229364	229.833354397033	253.224884061695
1500M_Women	G	5.42560784739696	5.13780895023414	0.0222540347304397	0.0210735795938016	237.643451255749	233.457940410708	241.82896210079
200M_Men	B	0.164710644196389	0.137829874006633	0.00806772355977612	0.00675107141490172	19.9325668272224	19.6701047998858	20.195028854559
200M_Men	G	0.255874928140973	0.196247971424366	0.0124604299070354	0.00955675536519924	19.466280946153	18.9310391453671	20.001522746939
200M_Women	B	0.37138518693556	0.290368624854269	0.0165335642486616	0.0129268169105963	22.1165031690561	21.3631527662783	22.8698535718339
200M_Women	G	0.370978124108981	0.277117699198087	0.0163787251262243	0.012234777006538	21.7129357361871	21.0004839417742	22.4253875306001
400M_Men	B	0.25315511012148	0.196749413979556	0.00569246065212897	0.00442411886084629	44.3510578312253	43.8345462980426	44.8675693644081
400M_Men	G	0.728645950610955	0.54139700016981	0.0157852242333396	0.0117287045097446	43.72936053515	42.2395237158806	45.2191973544194
400M_Women	B	0.535053194786558	0.524980811307212	0.0106896615900351	0.010488428569004	49.4287185944189	49.1806751658298	49.676762023008
400M_Women	G	0.411127053613861	0.295734798192441	0.00843049323883515	0.00606428157605827	49.0493122671432	48.1441206428733	49.9545038914131
800M_Men	B	1.07330731635357	0.99931888145097	0.00994232677410214	0.00925695252378469	103.260876357972	100.971230009641	105.550522706302
800M_Men	G	1.88887121827213	1.65363162862639	0.0179397019495881	0.0157054955705802	101.415279292206	97.4761330998081	105.354425484605
800M_Women	B	0.889746095077839	0.793222282172702	0.0075761758777064	0.00675427692585747	117.279666846518	115.161550769848	119.397782923188
800M_Women	G	0.892045549908214	0.833224788461841	0.00766510321933547	0.00715967251798536	115.620618765214	114.186613850382	117.054623680047

TABLE A3

MODEL 4 RAW RESULTS (INDIVIDUALIZED ARIMA MODEL)

Race	Medal	ARIMA_RMSE	ARIMA_MAE	Normalized_ARIMA_RMSE	Normalized_ARIMA_MAE	ARIMA_2028_Predictions	ARIMA_95CI_Lower	ARIMA_95CI_Upper
100M_Men	B	0.527603235027224	0.507416666666778	0.0505657691227932	0.0486310778864077	10.5070833333339	9.04040929534934	11.9737573713184
100M_Men	G	0.239818837842808	0.191376811594251	0.0234847890820442	0.0187409967286683	10.3091304347828	9.17235265942155	11.445908210144
100M_Women	B	0.484020031222904	0.474062500000028	0.0434585886619891	0.0425645342312034	11.3231250000001	10.4793316709897	12.1669183290104
100M_Women	G	0.536526225477374	0.383875	0.0474046850571986	0.0339172115214703	11.109375	10.3038808695388	11.9148691304612
1500M_Men	B	9.62793053890388	8.99823061383817	0.0440211109249036	0.0411419781622905	226.919828651377	196.624874860707	257.214782442047
1500M_Men	G	14.1557423326821	13.2548550724638	0.0646888001980949	0.0605719326677553	227.224782608696	198.10480228428	256.344762933111
1500M_Women	B	5.58253729648687	5.34800000000001	0.0232919515033595	0.0223133944341683	242.894	232.253743207615	253.534256792385
1500M_Women	G	5.55451179083892	5.13888888888896	0.0227827557355201	0.0210780091421595	238.664444444444	227.986934132436	249.341954756453
200M_Men	B	0.717181842992643	0.624857142857132	0.0351284209929782	0.0306062472010742	20.8042857142857	19.0115713435132	22.5970000850581
200M_Men	G	0.915177578396674	0.848333333333332	0.0445667191817226	0.0413115818521223	20.55	18.6418574893635	22.4581425106365
200M_Women	B	0.627868616829336	0.518499999999949	0.0279518582895642	0.0230829159710402	22.861999999999	20.7797917595489	24.9442082404491
200M_Women	G	1.0846832367369	0.957992655453585	0.0478888846241457	0.0422954814769795	22.5144982258867	21.434806866578	23.5941895851953
400M_Men	B	2.27259637663959	2.23582608695654	0.0511017353984438	0.0502749165082871	46.7078260869565	41.458574434286	51.9570777396271
400M_Men	G	2.62946689583073	2.30666666666667	0.0569641875179967	0.0499711149624495	45.8665217391305	40.6591012360509	51.07394224221
400M_Women	B	1.14651646303051	1.12333333333333	0.0229058963045521	0.0224427277570591	50.25	47.7619613592663	52.7380386407337
400M_Women	G	0.992790679515658	0.943333333333243	0.020357977023561	0.0193438140806543	49.7099999999999	47.1542053244804	52.2657946755195
800M_Men	B	4.74331234667723	4.57968253968255	0.0439385445563876	0.0424227987990108	109.399047619048	90.1443282121668	128.653767025928
800M_Men	G	4.70104874625333	4.24283806600431	0.0446485777020926	0.0402966859721181	111.499667575102	98.1577508187458	124.841584331458
800M_Women	B	5.23756304655768	5.15999999999999	0.0445977779849939	0.0439373297002724	122.6	116.902397275788	128.297602724212
800M_Women	G	2.71755538095147	2.56480769231112	0.02335120947736	0.02203869040245	118.942307692311	106.833935422253	131.050679962369

TABLE A4

SUMMARY OF NORMALIZED RMSE AND MAE METRICS ACROSS MODELS 2, 3, 4

	Mean Normalized RMSE	Min. Normalized RMSE	Max. Normalized RMSE	Range Normalized RMSE	Mean Normalized MAE	Min. Normalized MAE	Max. Normalized MAE	Range Normalized MAE
Model 2	0.0192	0.01008	0.02824	0.01816	0.0167	0.00876	0.02631	0.01755
Model 3	0.01542	0.00569	0.04273	0.03703	0.01294	0.00442	0.03186	0.02744
Model 4	0.03916	0.02036	0.06469	0.04433	0.03585	0.01874	0.06057	0.04183

TABLE A5

NORMALIZED RMSE AND MAE COMPARISON FOR EACH RACE

Medal	Race	Normalized_LM_RMSE	Normalized_KNN_RMSE	Normalized_ARIMA_RMSE	Normalized_LM_MAE	Normalized_KNN_MAE	Normalized_ARIMA_MAE	lowest_RMSE	lowest_MAE
B	100M_Men	0.0246471079981535	0.0427265189340963	0.0505657691227932	0.021008900333437	0.0318640705377789	0.0486310778864077	LM	LM
G	100M_Men	0.0141755958984831	0.0201194910135383	0.0234847890820442	0.0112630386449443	0.0158541796271217	0.0187409967286683	LM	LM
B	100M_Women	0.0175545965932082	0.0106005697276275	0.0434585886619891	0.0144381336635454	0.00835769949668314	0.0425645342312034	KNN	KNN
G	100M_Women	0.0220039124434697	0.0210209299559853	0.0474046850571986	0.0192864972759903	0.0182219798679358	0.0339172115214703	KNN	KNN
B	1500M_Men	0.0202148351013174	0.00894842015549017	0.0440211109249036	0.0176503515120783	0.00795078211936725	0.0411419781622905	KNN	KNN
G	1500M_Men	0.0238586177439202	0.0218119810843398	0.0646888001980949	0.016828820956187	0.019225477186664	0.0605719326677553	KNN	LM
B	1500M_Women	0.0228107966493912	0.0237543775044087	0.0232919515033595	0.0218320725427874	0.0231306566273088	0.0223133944341683	LM	LM
G	1500M_Women	0.0200761064180188	0.0222540347304397	0.0227827557355201	0.0184830610219645	0.0210735795938016	0.0210780091421595	LM	LM
B	200M_Men	0.0103216503551289	0.00806772355977612	0.0351284209929782	0.00937695195020444	0.00675107141490172	0.0306062472010742	KNN	KNN
G	200M_Men	0.0184281970109385	0.0124604299070354	0.0445667191817226	0.0155696786080602	0.00955675536519924	0.0413115818521223	KNN	KNN
B	200M_Women	0.0180934392126348	0.0165335642486616	0.0279518582895642	0.0173358884839408	0.0129268169105963	0.0230829159710402	KNN	KNN
G	200M_Women	0.0244911734616349	0.0163787251262243	0.0478888846241457	0.0181240660593525	0.0122347770065381	0.0422954814769795	KNN	KNN
B	400M_Men	0.0172027191598362	0.00569246065212897	0.0511017353984438	0.0149826807942393	0.00442411886084629	0.0502749165082871	KNN	KNN
G	400M_Men	0.0122062320900326	0.0157852242333396	0.0569641875179967	0.010341209035816	0.0117287045097446	0.0499711149624495	LM	LM
B	400M_Women	0.0216692611911677	0.0106896615900351	0.0229058963045521	0.0188153836219406	0.010488428569004	0.0224427277570591	KNN	KNN
G	400M_Women	0.0108458075182596	0.00843049323883515	0.020357977023561	0.00876139437536225	0.00606428157605827	0.0193438140806543	KNN	KNN
B	800M_Men	0.0277524745794525	0.00994232677410214	0.0439385445563876	0.0263076287852373	0.00925695252378469	0.0424227987990108	KNN	KNN
G	800M_Men	0.0282435547959027	0.0179397019495881	0.0446485777020926	0.0257478362923951	0.0157054955705802	0.0402966859721181	KNN	KNN
B	800M_Women	0.0100810899356027	0.0075761758777064	0.0445977779849939	0.00962750067344767	0.00675427692585747	0.0439373297002724	KNN	KNN
G	800M_Women	0.0193423904051634	0.00766510321933547	0.02335120947736	0.0181585784113376	0.00715967251798536	0.02203869040245	KNN	KNN

TABLE A6

2028 OLYMPIC TIME PREDICTIONS AND 95% CONFIDENCE INTERVALS

Race	Medal	Predictions (secs)	95% Confidence Interval
100M Men	B	9.8699	[9.0634, 10.6764]
	G	9.8000	[9.4908, 10.1091]
100M Women	B	10.9520	[10.7877, 11.1164]
	G	10.6799	[10.1695, 11.1903]
1500M Men	B	215.4561	[212.2677, 218.6445]
	G	218.9286	[208.8991, 228.9582]
1500M Women	B	241.5291	[229.8334, 253.2249]
	G	237.6435	[233.4579, 241.8290]
200M Men	B	19.9326	[19.6701, 20.1950]
	G	19.4663	[18.9310, 20.0015]
200M Women	B	22.1165	[21.3632, 22.8699]
	G	21.7129	[21.0005, 22.4254]
400M Men	B	44.3511	[43.8345, 44.8676]
	G	43.7294	[42.2395, 45.2192]
400M Women	B	49.4287	[49.1807, 49.6768]
	G	49.0493	[48.1441, 49.9545]
800M Men	B	103.2609	[100.9712, 105.5505]
	G	101.4153	[97.4761, 105.3544]
800M Women	B	117.2797	[115.1616, 119.3978]
	G	115.6206	[114.1866, 117.0546]

Fig. A6 Women's 100m bronze medalist race times vs year, including 2028 predictions and 95% CI's

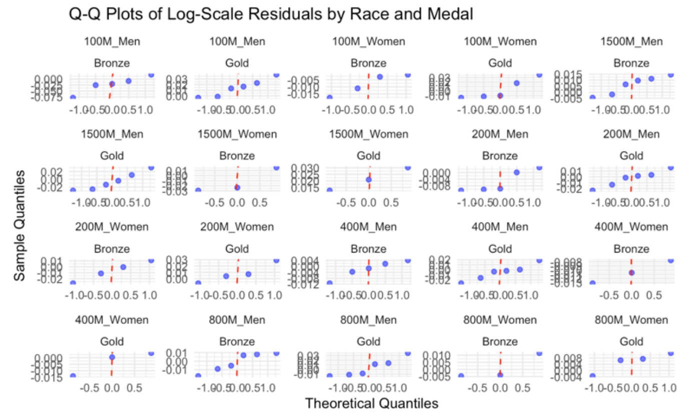

Fig. A7 QQ plots of log-scale residuals by race and medal

APPENDIX B

ADDITIONAL HISTOGRAMS OF HISTORICAL RACE TIMES, BY RACE, GENDER, AND MEDAL

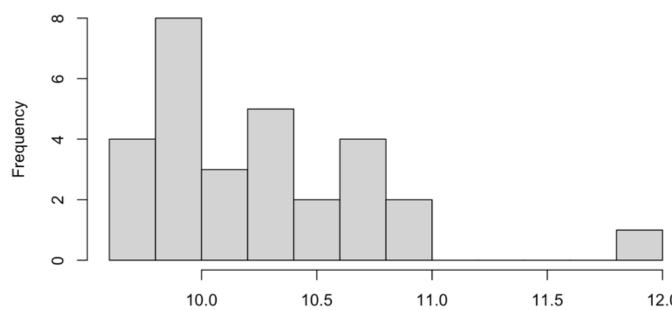

Fig. B1 Histogram of men's 100m gold medal Olympic finals race results

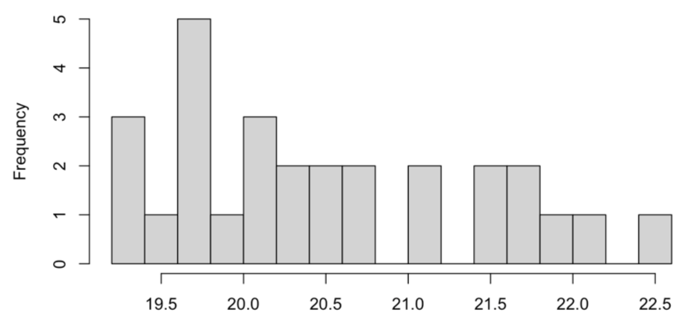

Fig. B2 Histogram of men's 200m gold medal Olympic finals race results

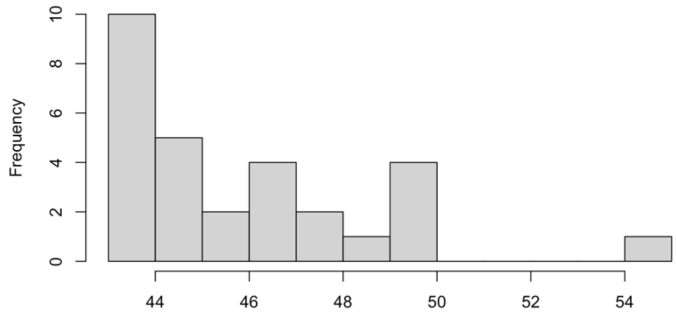

Fig. B3 Histogram of men's 400m gold medal Olympic finals race results

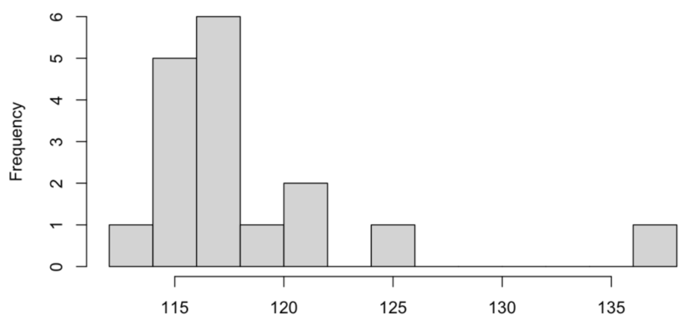

Fig. B4 Histogram of women's 800m gold medal Olympic finals race results

APPENDIX C
ADDITIONAL FIGURES DISPLAYING HISTORICAL PERFORMANCE PLOTS, FOR GOLD AND BRONZE MEDAL TRENDS

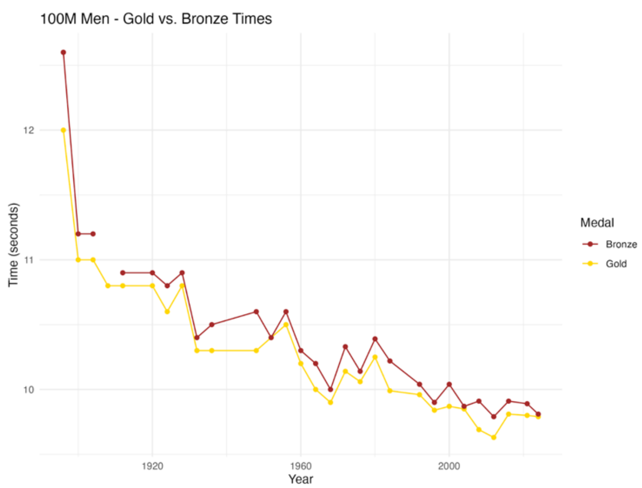

Fig. C1 100m men's Olympic historical trends

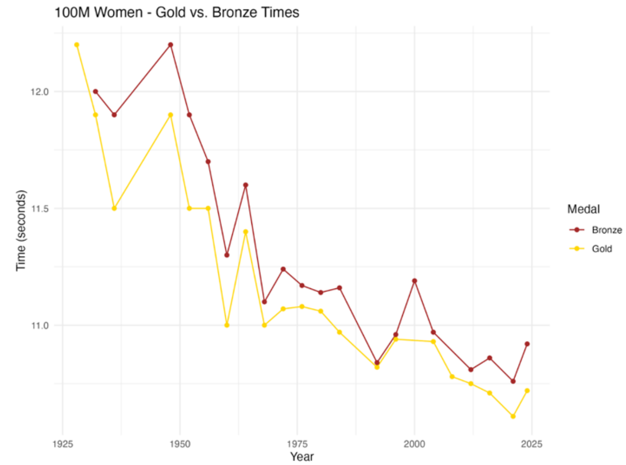

Fig. C2 100m women's Olympic historical trends

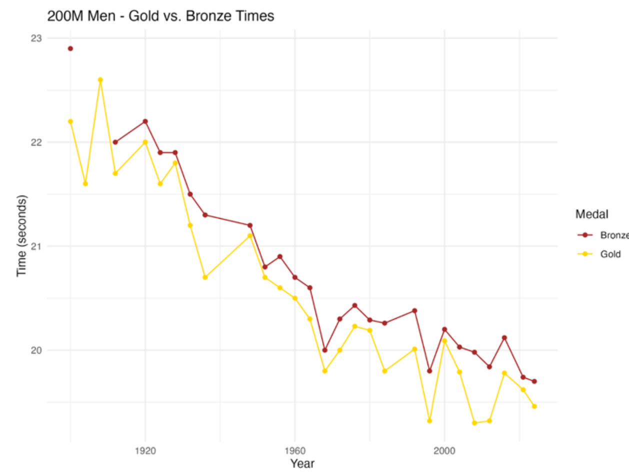

Fig. C3 200m men's Olympic historical trends

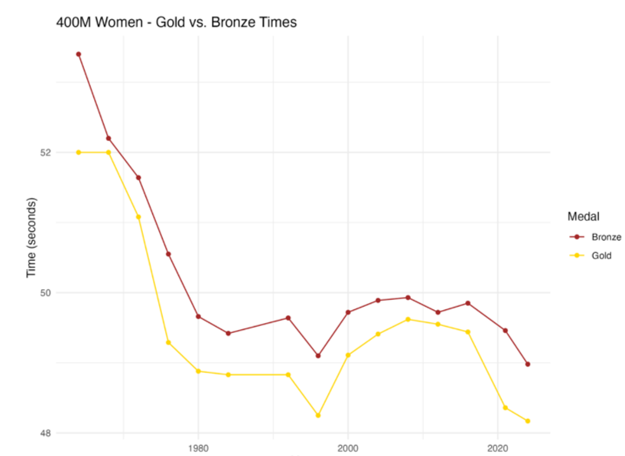

Fig. C4 400m women's Olympic historical trends

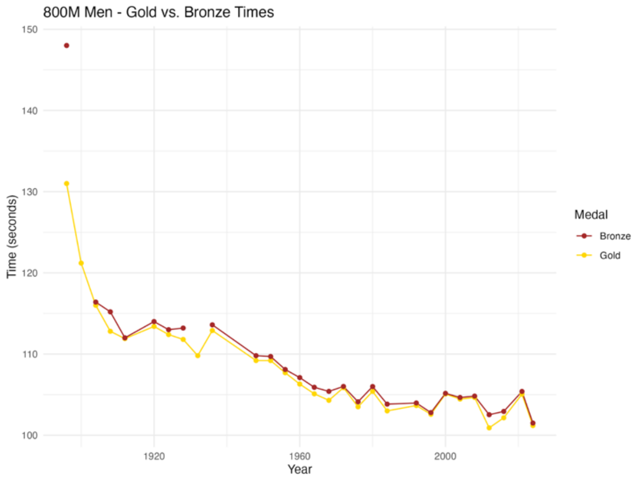

Fig. C5 800m men's Olympic historical trends

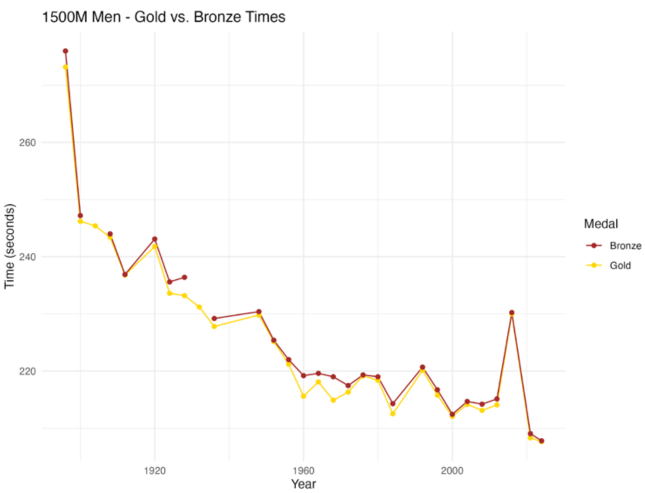

Fig. C6 1500m men's Olympic historical trends

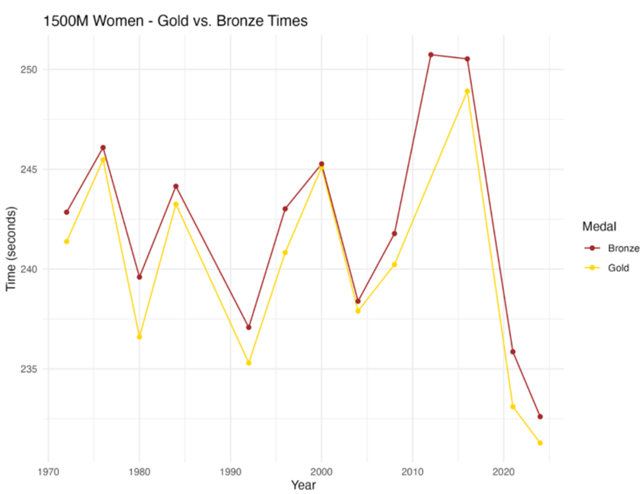

Fig. C7 1500m women's Olympic historical trends

APPENDIX D
ADDITIONAL FIGURES DISPLAYING HISTORICAL RESULTS, 2028 PREDICTIONS, AND 2028 95% CONFIDENCE INTERVALS

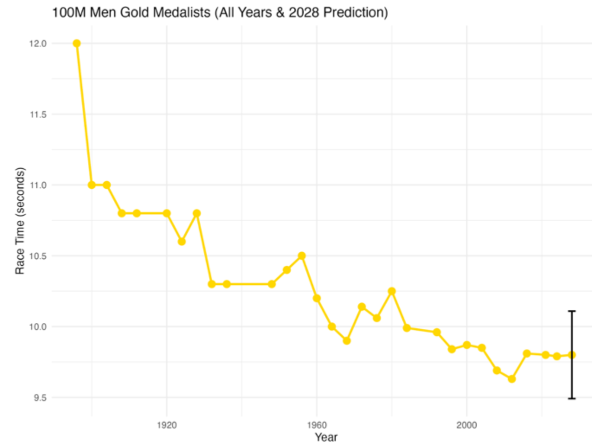

Fig. D1 Men's 100m gold medalist race times vs year, including 2028 predictions and 95% CI's

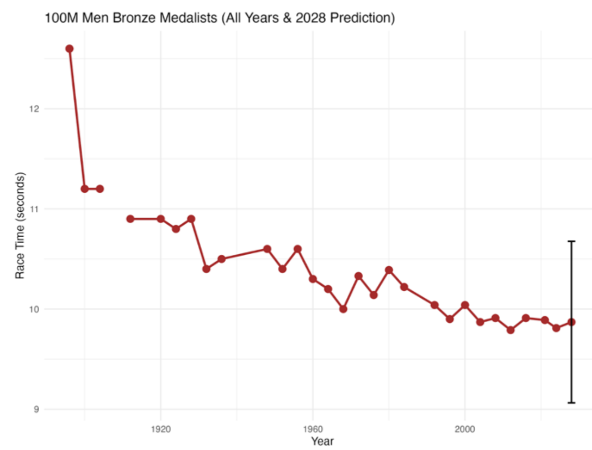

Fig. D2 Men's 100m bronze medalist race times vs year, including 2028 predictions and 95% CI's

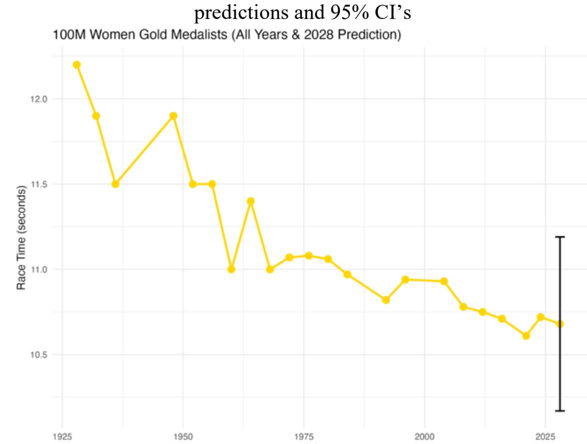

Fig. D3 Women's 100m gold medalist race times vs year, including 2028 predictions and 95% CI's

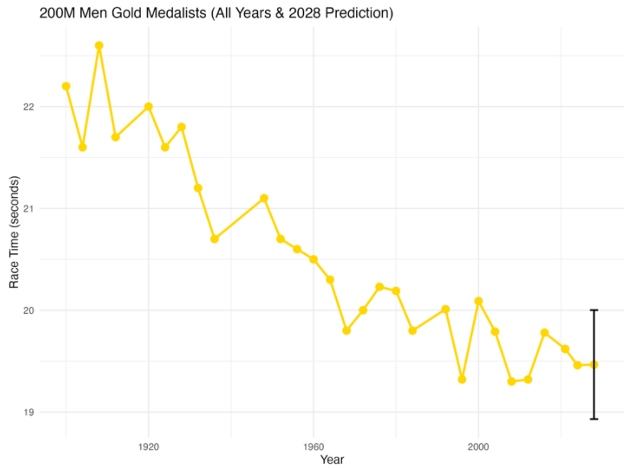

Fig. D4 Men's 200m gold medalist race times vs year, including 2028 predictions and 95% CI's

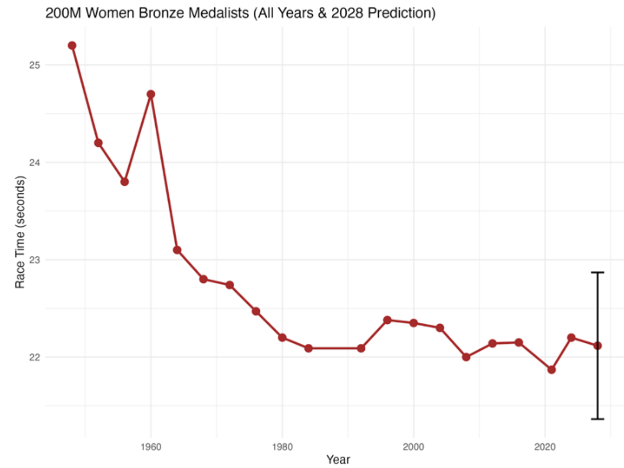

Fig. D7 Women's 200m bronze medalist race times vs year, including 2028 predictions and 95% CI's

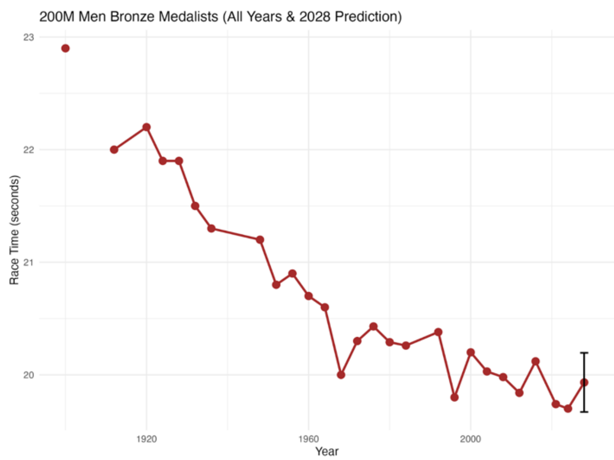

Fig. D5 Men's 200m bronze medalist race times vs year, including 2028 predictions and 95% CI's

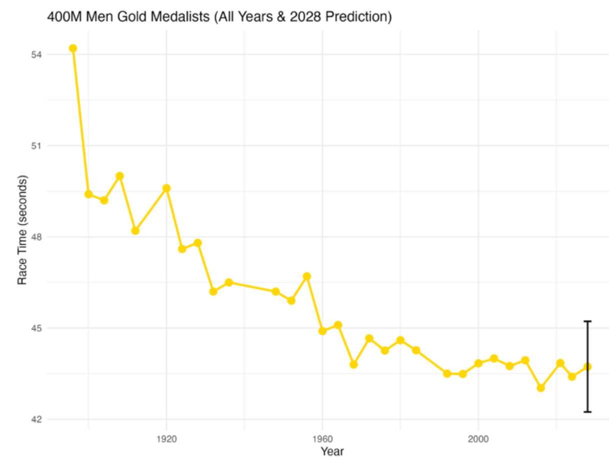

Fig. D8 Men's 400m gold medalist race times vs year, including 2028 predictions and 95% CI's

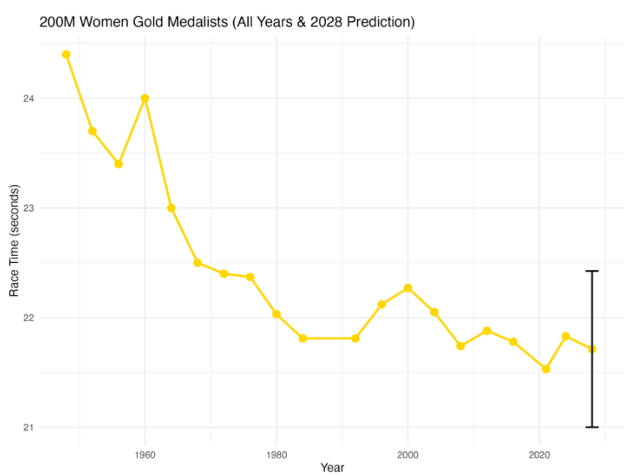

Fig. D6 Women's 200m gold medalist race times vs year, including 2028 predictions and 95% CI's

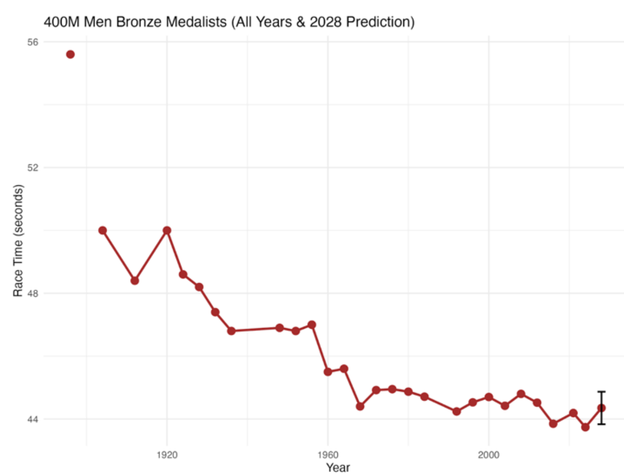

Fig. D9 Men's 400m bronze medalist race times vs year, including 2028 predictions and 95% CI's

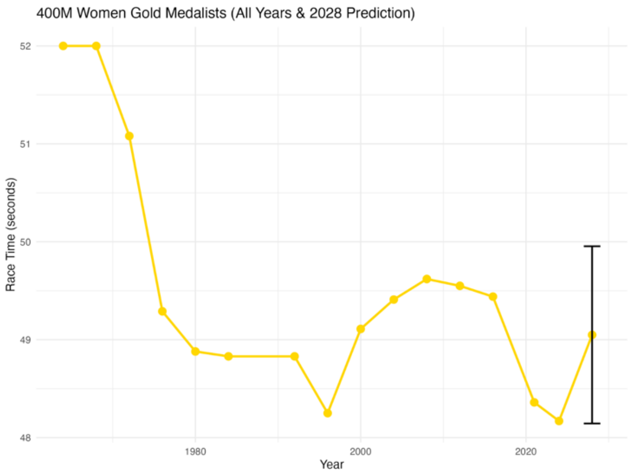

Fig. D10 Women's 400m gold medalist race times vs year, including 2028 predictions and 95% CI's

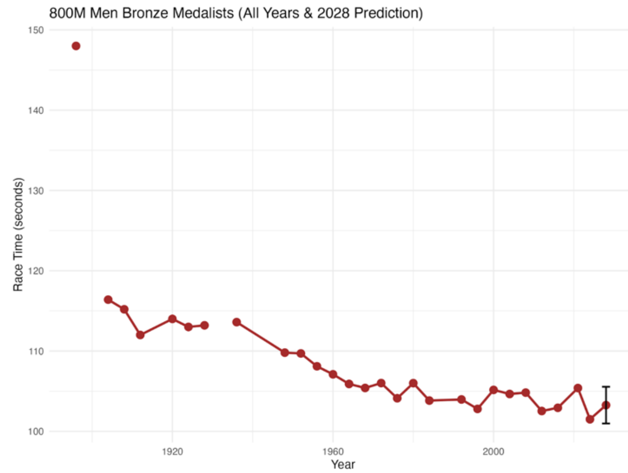

Fig. D13 Men's 800m bronze medalist race times vs year, including 2028 predictions and 95% CI's

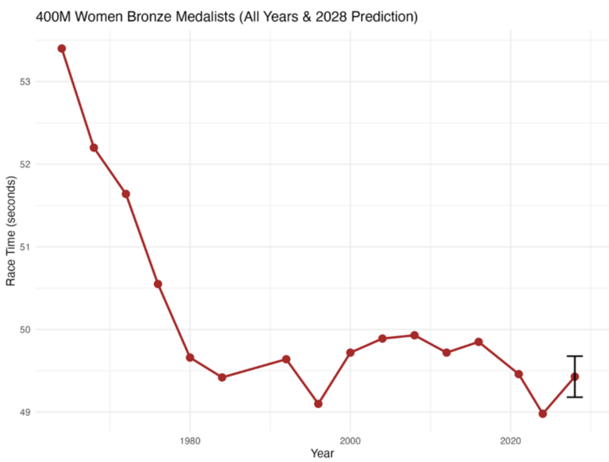

Fig. D11 Women's 400m bronze medalist race times vs year, including 2028 predictions and 95% CI's

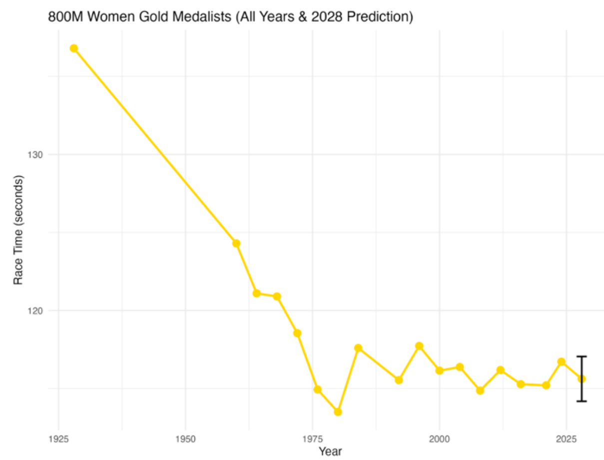

Fig. D14 Women's 800m gold medalist race times vs year, including 2028 predictions and 95% CI's

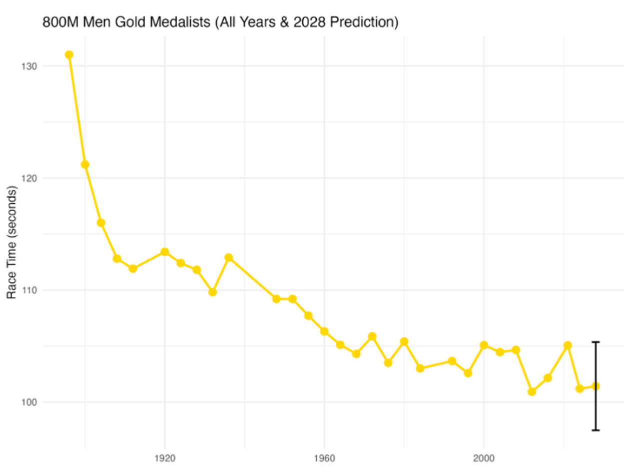

Fig. D12 Men's 800m gold medalist race times vs year, including 2028 predictions and 95% CI's

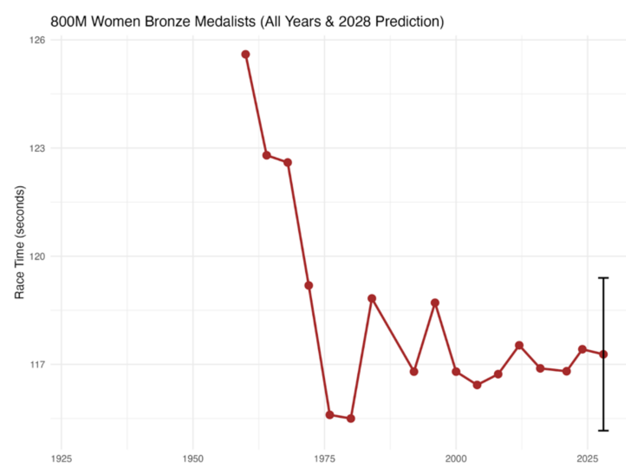

Fig. D15 Women's 800m bronze medalist race times vs year, including 2028 predictions and 95% CI's

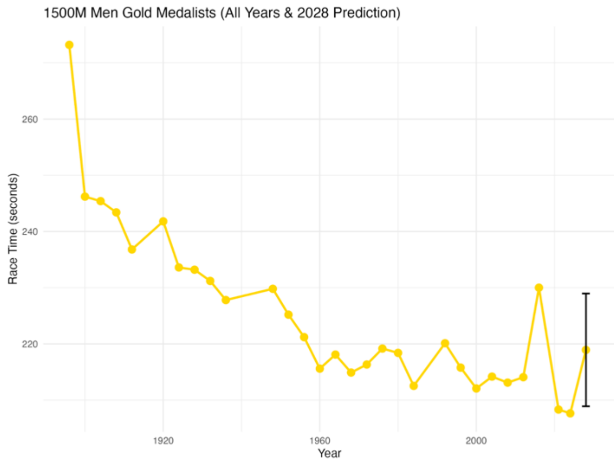

Fig. D16 Men's 1500m gold medalist race times vs year, including 2028 predictions and 95% CI's

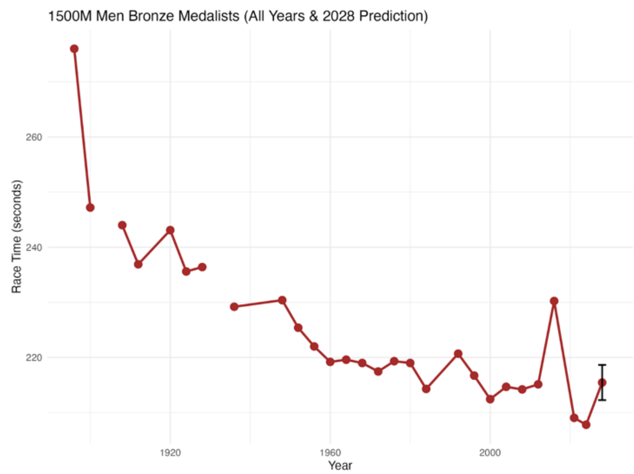

Fig. D17 Men's 1500m bronze medalist race times vs year, including 2028 predictions and 95% CI's

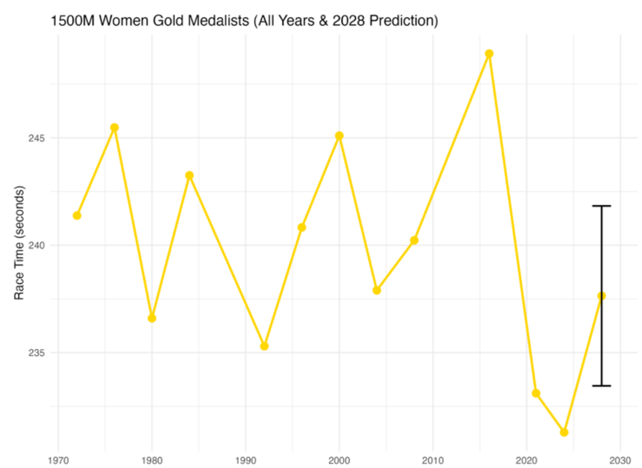

Fig. D18 Women's 1500m gold medalist race times vs year, including 2028 predictions and 95% CI's

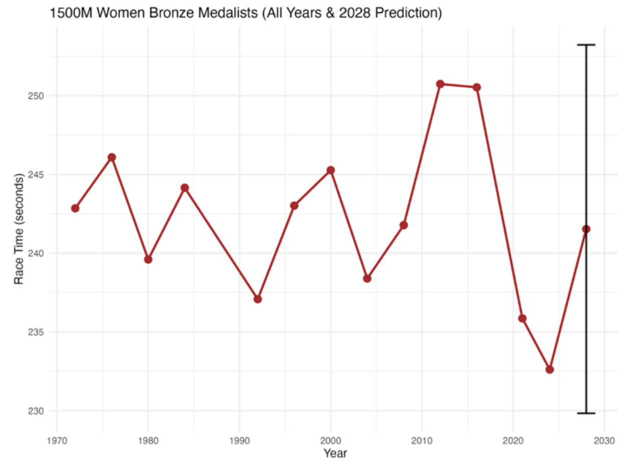

Fig. D19 Women's 1500m bronze medalist race times vs year, including 2028 predictions and 95% CI's

APPENDIX E
ADDITIONAL FIGURES

```
Call:
lm(formula = log(Result) ~ Year + male + Medal + Event, data = Train_olympics)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.046457 -0.014899 -0.002468  0.010242  0.218538 

Coefficients: (1 not defined because of singularities)
                   Estimate Std. Error  t value Pr(>|t|)    
(Intercept)       7.254e+00  8.520e-02   85.145  < 2e-16 ***
Year             -1.244e-03  4.275e-05  -29.100  < 2e-16 ***
male             -2.471e+00  6.808e-03 -362.933  < 2e-16 ***
MedalG           -1.255e-02  2.832e-03   -4.431 1.26e-05 ***
Event100M Women  -2.369e+00  6.686e-03 -354.290  < 2e-16 ***
Event1500M Men    3.087e+00  6.050e-03  510.209  < 2e-16 ***
Event1500M Women  7.197e-01  7.992e-03   90.056  < 2e-16 ***
Event200M Men     6.974e-01  6.027e-03  115.697  < 2e-16 ***
Event200M Women  -1.660e+00  6.919e-03 -239.864  < 2e-16 ***
Event400M Men     1.495e+00  5.847e-03  255.627  < 2e-16 ***
Event400M Women  -8.582e-01  7.261e-03 -118.202  < 2e-16 ***
Event800M Men     2.354e+00  5.803e-03  405.731  < 2e-16 ***
Event800M Women          NA         NA       NA       NA    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02666 on 346 degrees of freedom
  (8 observations deleted due to missingness)
Multiple R-squared:  0.9994,	Adjusted R-squared:  0.9994 
F-statistic: 5.356e+04 on 11 and 346 DF,  p-value: < 2.2e-16
```

Fig. E1 Model 1 summary