Evaluating Current Swim Time Conversion Methods

AP Statistics

Alec Vercruysse

July 21, 2019

Abstract

This paper aims to evaluate current techniques of coverting race times between different pool lengths. It then proposes a novel model to more accurately estimate conversion times, based off a weighted least squares linear regression that takes into account both age and gender.

Contents

1 Introduction

Competitive swimming can be done in three different length pools: 50 Meter “Olympic Size” pools, 25 Meter Pools, and 25 Yard “NCAA Size” Pools. These lengths are known as Long Course Meters (LCM), Short Course Meters (SCM), and Short Course Yards (SCY), respectively. In theory, while there are differences between swimming a long course and short course race, mainly surrounding the reduced turn – and therefore underwater – count, times should be convertible, so that one could compare a swim in one type of pool to another time of another type of pool, such as a Short Course Yards to Short Course Meters. For example, high level competitive meets require a swimmer to qualify by beating a specific qualifying time in an event. If the qualification times are provided in short course, for example, but the swimmer swims a race in long course, often the swimmer wants to compare their performance with the qualification time in a different pool length. This is so common, in fact, that often, for mid-level qualification meets, the meet host will often allow converted times to count.

1.1 The De-facto Model

USA Swimming itself does not have an official model for time conversion. The vast majority of time conversion, however, is accomplished through a single model popularized by a number of big websites and companies, including TeamUnify [1], SwimSwam[1], and Colorado Timing [4].

These models use a simple linear regression that include the original time and number of turns in the swim. They do not specify different "m" coefficients (see table) for different events, only for some differing distances in freestyle. Specifically, some conversions go between 400M and 500Y, 800M and 1000Y, and 1500M to 1650Y. Different conversion factors are specified for each. Note that $\hat{y}$ is the time in Long Course Meters, and x is the time in short course yards. "distance", here and throughout the rest of the paper, refers to the nominal distance of the event. For example, a 200Y and 200M event both have a nominal distance of 200.

Table 1: Example SCY -> LCM Conversion ($\hat{y} = mx + b$)

	distance	m	b
fly	<= 200	1.11	.7 * distance / 50
back	<= 200	1.11	.6 * distance / 50
breast	<= 200	1.11	1 * distance / 50
free	<= 200	1.11	.8 * distance / 50
free	500/400	0.8925	6.40
free	1000/800	0.8925	12.80
free	1650/1500	1.02	24.00

[1]SwimSwam, which uses Colorado Timing’s model, outputs the same time as TeamUnify’s calculator, which publishes its model

This raises the question of whether or not factors such as stroke, age, and gender should be accounted for in an improved model.

2 Dataset Generation

In choosing the data to evaluate the model with, important considerations were made that limit the scope of the conclusions made. The data was chosen to be from races by USA club swimmers. USA club swimmers are the most likely people to use such a calculator, as they often switch between the different pool sizes. International swimmers are more likely to stick with the common Olympic pool length, and collegiate swimmers are more likely to stick to the NCAA pool length. Since USA Club swimmers are generally more likely to use 25 SCY pool lengths than International swimmers, it should be noted that the results of this model do not generalize Internationally.

The website used to gather data on USA club swimmers is swimmingrank.com. This website is created by the dad of two club swimmers, and grew from a database of just local swimmers that he was asked to create for his club. It arose from the fact that USA Swimming's time search is very hard to use. Other options were evaulated, such as fastlanetek.com, which hosts results for many meets that the company provides timing solutions to, but swimmingrank.com was chosen for the sheer amount of results it had, as well as for ease of scraping, and the fact that it only includes data on USA Club Swimmers.

Most people use the website, which contains the data for many club swimmers by scraping the results for each Local Swim Committee's (LSC) results on their website, through its search feature. To get everyone's data, a different method had to be used.

2.1 Taking Advantage of Regions

`https://www.swimmingrank.com/regions` provides a list of regions to narrow down by, so this was a good place to start. The links are this page are easily iterable with some simple and fast regex.

While often it is standard to use a dedicated comprehensive scraping tool such as beautiful-soup, finding links on this website is very easy with regex, so the additional overhead of a python package like beautifulsoup, which is known to run pretty slow, is not ideal.

Each region has its own folder: e.g. `https://www.swimmingrank.com/aft/`, for that of Ala-Fla-Ten, with an index.html in the folder. By navigating through the subdirectories of this folder, luck had it that going to the `/strokes/` subdirectory led the web server to give me an index of the entire directory. From there, there was one subdirectory for each LSC, as well as one subdirectory containing info on the clubs in the region. Each LSC subdirectory contains the html file for each swimmer and stroke, which was exactly what was needed.

It was then possible to simply iterate through each LSC directory in the strokes directory of each region.

Each one of these directories contains the links for every HTML page of every event for every swimmer in the LSC. Finding the link took simple regex. We chose to write to a file here to avoid processing this multiple times, and holding a large array of links in memory.

This produced a ~500 Megabyte file, with ~6 million links.

This brought the count down to approximately 5m links.

2.2 Event Page Format

Each event page contains contains very simple html, with all data stored neatly in tables. Pages include many points of data, such as season rankings, rankings by career best, upcoming championships meets, some interesting stats comparing one's performance to NCAA percentiles, and even a graph. Tables of interest include the demographic table at the top, containing `Age` and `Sex`, and the two bottom tables, detailing a history of all the swimmer's races sanctioned by the LSC. An example page can be found here. Sometimes a swimmer does not have LCM or SCM times in the event, such as in this case, which should be handled gracefully. While these pages could be ignored for the purposes of the more accurate swim time converter, in the interest of obtaining a complete data set, they were scraped.

Because of the heavy use of basic html tables, the package `lxml` was used, which supports html parsing. Unfortunately, everything is at the top level, and no CSS ids are used, so it was tricky to identify tables. Luckily, only the tables containing the event data are preceded with `<h3>` tags, and only the main demographic table is preceded with an `<h1>` tag, which allows for some identification.

2.3 Scraping Millions of Links

A `whois` lookup on swimmingrank.com revealed that they are hosted by Bluehost, has anti Denial-of-Service (DoS) attack protection. A DoS attack is when an attacker makes a lot of requests to a server at once, which often bogs down the server with work, so it is unable to process other users' requests. Since we are going to make a lot of requests to the swimmingrank server, it is important to make sure that we are not identified as someone who is trying to simply run a DoS attack against the server.

To do this, we rate-limited our requests, so that we did do not put too much load on the server. It is also considered good web-scraping etiquette to do so, as we do not want to use up too much of swimmingrank's computing resources. If we were to process one web-page a second, however, that would total to 57 days, so we had to keep it low. We chose to go with a delay equaling ~100ms to stay safe (everything is done in one thread). This method was also chosen for it's ease, since that is approximately the time it takes to process a single page.

Just in case our IP did get banned, we chose to use a VPN as well when sending our requests. We chose a location within the US in order to reduce latency as well as mask our traffc, since most traffic to a USA Swim Times website is most likely coming from the United states.

We also kept track of the last file scraped just in case this script needs to stop and restart multiple times.

Please note that for this revision of the paper, not all links have been scraped, in the interest of time. The data in this paper use approximately 1 million links. A large number of states and regions have had their swims completely parsed: Alabama, Florida, Tennessee, Alaska, Arizona, California, Nevada, Colorado, New Mexico, Utah, Wyoming, Connecticut, Maine, New England, D.C., Maryland, Virginia, Delaware, New Jersey, Pennsylvania, Georgia, North Carolina, South Carolina, Hawaii, Idaho, Montana, Oregon, and Washington. Notable omissions include Texas, Indiana, Michigan, and New York. It would be trivial, however, to simply scrape for longer.

In order to evaluate the current linear model for time conversion, a linear model was constructed, and the slope of best fit was compared to the slope provided by the model, which was the null hypothesis.

To create the model, input values need to be paired with their corresponding output values, e.g. SCY times with LCM times. There were multiple possible ways to accomplish this, each one with the possibility of leading different results, and with different assumptions in mind.

2.4 Match Swims by Closest Date

In order to match times, for each LCM time of a swimmer, their SCY time that was closest by date to the LCM time was found, and added as a column to the data. While this does not account for things such as a taper or just general meet conditions, due to the large amount of data available, outliers and edge case situations do not have much influence over the final result..

This has the potential to pair two times together that are relatively far, e.g. more than 1 year, apart. To mitigate the effect of this, the date of the yards swim was also included in the final dataset in order to be able to filter that "bad" data out.

While this method has the advantage of using the majority of the data available, it assumes that races swum at similar times of the year will have a similar performance. This is not true, when, for example, somebody swims a meet tapered, shaved, and with a fastsuit, and then later swims their start of season Long Course Meet, with just a regular brief and little LCM training

This is sometimes the case with high-school swimmers, as the LCM season generally starts in the beginning of the summer, after high-school championship meets. This model also fails when a swimmer might have a career in one course, and then switches to another course. For example, if a swimmer swims in LCM for a couple years before switching permanently to SCY, all LCM swims will "pair" with the first SCY swim they completed.

In order to proceed with caution, another model is also built, based on matching the best times of swimmers.

2.5 Match Swims by Best Times

Another method of matching times is to only get the best yards and best meters time for every swimmer. While this produces less data, it assumes that the data is more accurate if the swimmer is in similar condition and has similar meet conditions for their best SCY and LCM meets.

This model has the potential to pair swims of wildly different ability if the swimmer switches from one course to another. It instead assumes that one's best time in Yards is swam at a similar ability as their best time in Meters. This is not the case for every swimmer, and so it is important to proceed with caution. Particularly, USA swimmers are more likely to swim better in SCY than LCM, due to the fact that more swimming is generally done in SCY. However, since this model generalized to USA Swimmers only, having the model take into account such biases can be considered a good thing, as the accuracy will still increase for the average USA Swimmer.

This model is at a disadvantage over the previous one in that it contains less data, however, as only the best time for each swimmer is taken into account.

3 Evaluation of the Current Model

The current model for SCY to LCM conversion for all strokes where the distance is below 400 can be constructed as a formula with the equation:

$$y = ax + \frac{d * b}{50} \tag{1}$$

Where y is the time in meters, x is the time in yards, d is the nominal distance of the event, a is a general coefficient in the regression, and b is a stroke-specific coefficient. All times are formatted in centiseconds. This model, therefore, is constructed to only take account the stroke of the event when adding additional time to account for the decrease in turns.

Nevertheless, to evaluate the coefficients in the model, models with a similar formula are constructed. A separate linear model is therefore constructed for each stroke in order to account for different b values per stroke.

3.1 By Closest Date

Of the 3,896,490 swims currently in this dataset, the mean age of each swimmer is approximately 12.2 ($min = 5, Q1 = 11, Q3 = 14, Max = 18$). Approximately .443 of swims are from males.

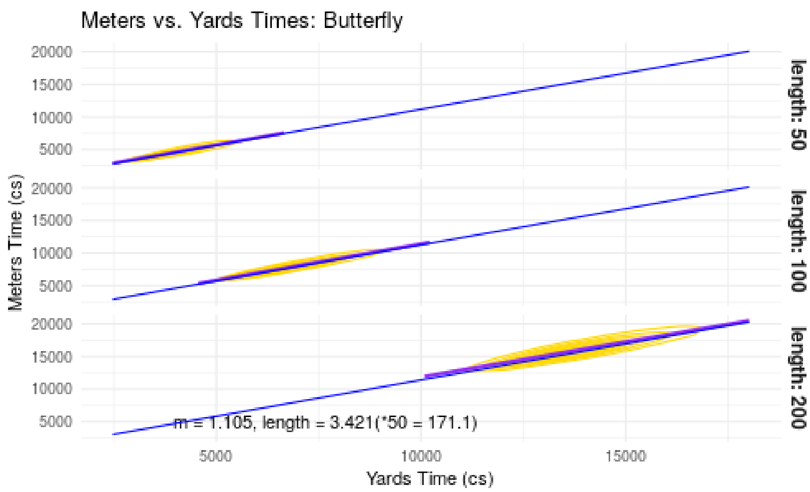

Figure 1: Butterfly Conversion with a basic linear model, and times matched by closest date. The purple line represents the linear model, while the blue line represents the de-facto model.

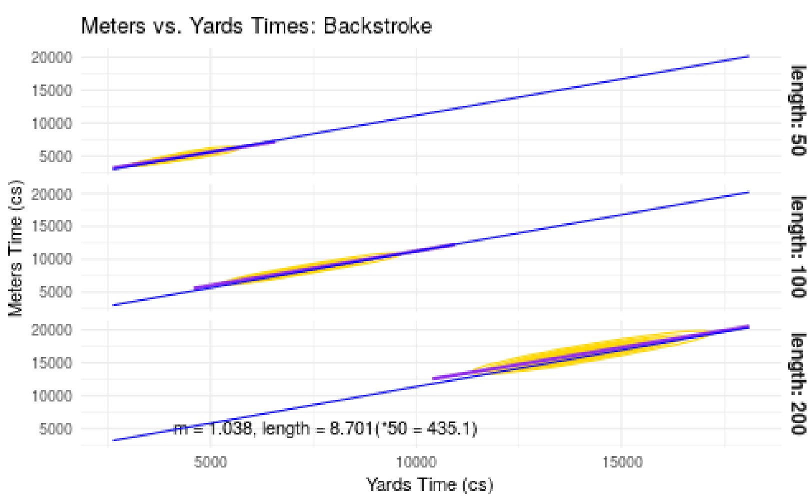

Figure 2: Backstroke Conversion with a basic linear model, and times matched by closest date. The purple line represents the linear model, while the blue line represents the de-facto model.

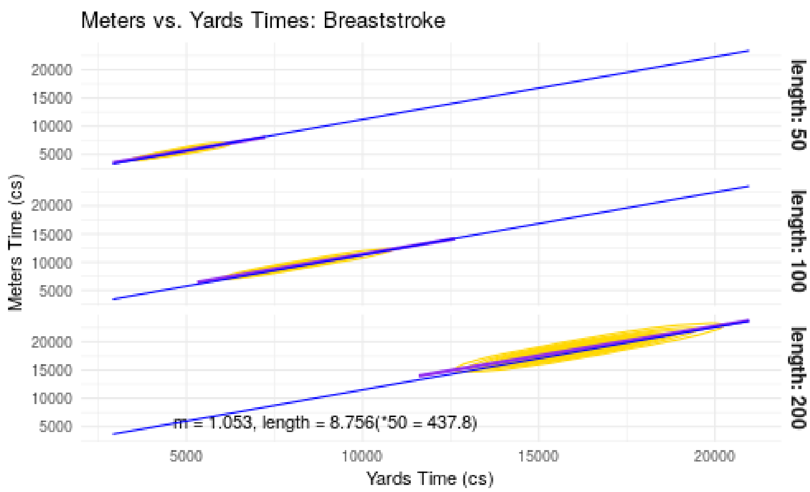

Figure 3: Breaststroke Conversion with a basic linear model, and times matched by closest date. The purple line represents the linear model, while the blue line represents the de-facto model.

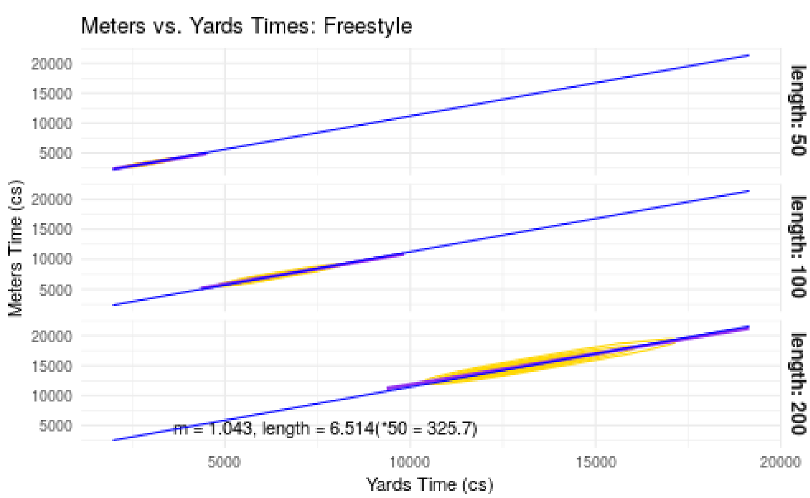

Figure 4: Freestyle Conversion with a basic linear model, and times matched by closest date. The purple line represents the linear model, while the blue line represents the de-facto model.

While the two linear models appear to be very similar, the b coefficients in each model, adding a constant to account for distance, seems to be very different for each model.

Table 2: Model coefficients at 95% confidence vs. de facto model

model name	model a	de-facto a	model b	de-facto b
Butterfly	(1.10 - 1.11)	1.11	(166 - 176)	70
Backstroke	(1.04 - 1.04	1.11	(432 - 438)	60
Breaststroke	(1.05 - 1.05)	1.11	(434 - 442)	100
Freestyle	(1.04 - 1.04)	1.11	(324 - 327)	80

Given that the coefficients provided are considered to be at 95% hypothesis, it is clear that the de facto model can be rejected. While this means that the de-facto model can be rejected, the provided model still does not look to be ideal.

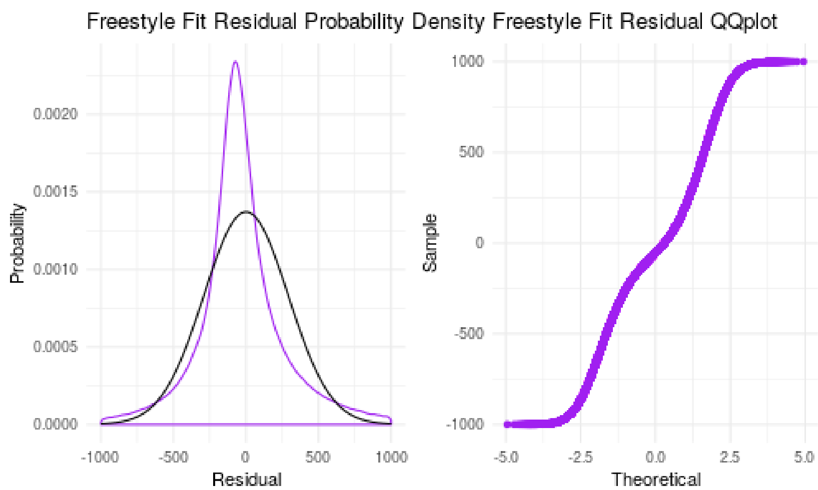

Figure 5: The QQplot and PDF function for the residuals of the freestyle fit. The corresponding normal distribution with the standard deviation of the residuals is shown in black.

The tails of the residuals are much more dense than that of a normal distribution, violating some assumptions required when interpreting a model.

3.2 By Best Time

Of the 688,868 swims currently in this dataset, the mean age of each swimmer is approximately 12.48 ($min = 5, Q1 = 11, Q3 = 14, Max = 18$), which is almost identical to the other dataset. Approximately .443 of swims are from males.

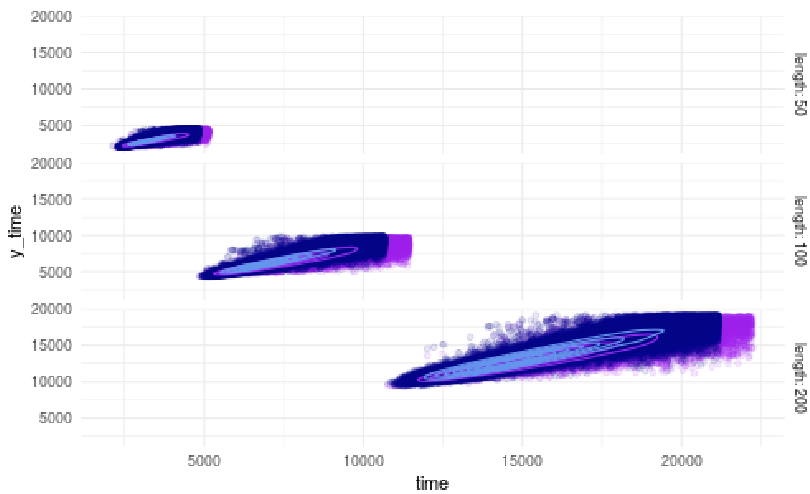

Figure 6: Comparison of the two datasets. Matched by best time is in blue, and matched by closest swim is in purple.

These models once again appear to very similar to the ones constructed with the other dataset, as the trend in data looks very similar. Once again, the b coefficients do not align with the de-facto model.

Table 3: Model coefficients at 95% confidence vs. de facto model

model name	model a	de-facto a	model b	de-facto b
Butterfly	(1.16 - 1.17)	1.11	(86 - 115)	70
Backstroke	(1.08 - 1.09)	1.11	(377 - 397)	60
Breaststroke	(1.08 - 1.09)	1.11	(440 - 465)	100
Freestyle	(1.12 - 1.13)	1.11	(177 - 190)	80

Once again, however, the residuals fail to be normally distributed.

The tails of the residuals are much more dense than that of a normal distribution, indicating that the model is not ideal. While the residuals are still unimodal and symmetric, they are unable to meet the requirements provided for some interpretation of the model. Most importantly, the results of this model cannot be interpreted in order to make prediction intervals, yet the model is still valid in order to make predictions and provide parameter estimates [3]. While it is possible to use other methods to normalize residuals, they will not be pursued in favor of building a new model.

Also note that the peak of the Probability Density Function Graph is not centered around zero, as it ideally would be. This means that there are a few influential points where the meters time is much slower than it should be, inflating the entire model.

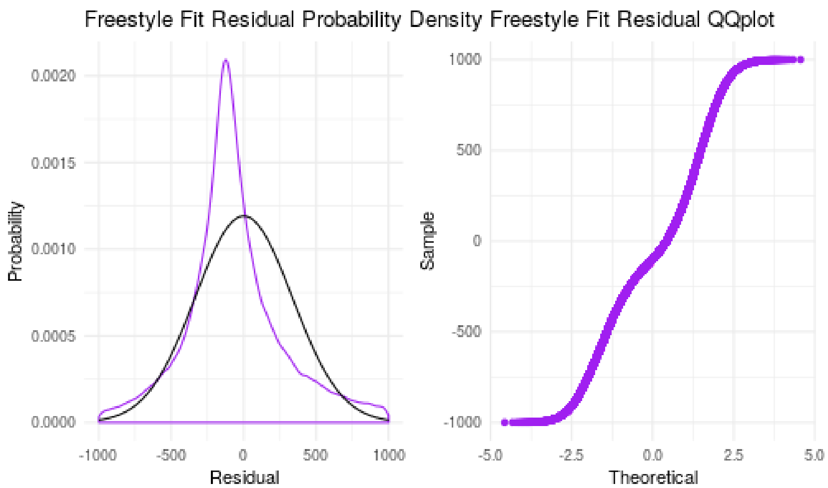

Figure 7: The QQplot and PDF function for the residuals of the freestyle fit. The corresponding normal distribution with the standard deviation of the residuals is shown in black.

4 An Improved Model

As seen in Figure 6, it is clear that the two data sets are not identical. From hereon out, we choose to use the data set grouped by best times since it seems to rid the dataset of some outer points that are not representative of most swimmers. We also choose to focus only on conversion between LCM and SCY times, but this model can easily be extended to provide conversions to SCM times as well.

4.1 The Weighted Least Squares Regression

Neither dataset properly takes into account the possibility that two best time swims were swam at very differing times, and therefore the natural error is very high, possibly influencing the model. To counter this, a weighting could be added to every data point that could estimate it's value, where the shorter the time difference between the swims, the more the model takes this data point into account. This can be mathematically introduced into the model by weighting each squared error term. Data points with more weight will then have higher error terms, making it more beneficial to minimize those terms over others. This method of regression is known as Weighted Least Squares (WLS).

$$\text{WSSR}(b, w_1, ...w_n) = \sum_i^n w_i(y_i - x_i b)^2 \tag{2}$$

Equation 2 shows the weighted sum of squared residuals for a model given the candidate parameter vector b and weights $w_1 ... w_i$. This term is minimized to find the best candidate β.

In order to map distance by time to a weighting, a minimum a maximum cutoff time delta must be specified. This maximum time delta was specified to be 1.5 years, which gives each swimmer at least one year to achieve a best time in both their short and long course season. Any additional time, especially amongst younger swimmers, gives the swimmer too much opportunity to grow and improve.

Absolute time delta values between 0 and 1.5 years will be linearly mapped to values between 1 and 0, respectively.

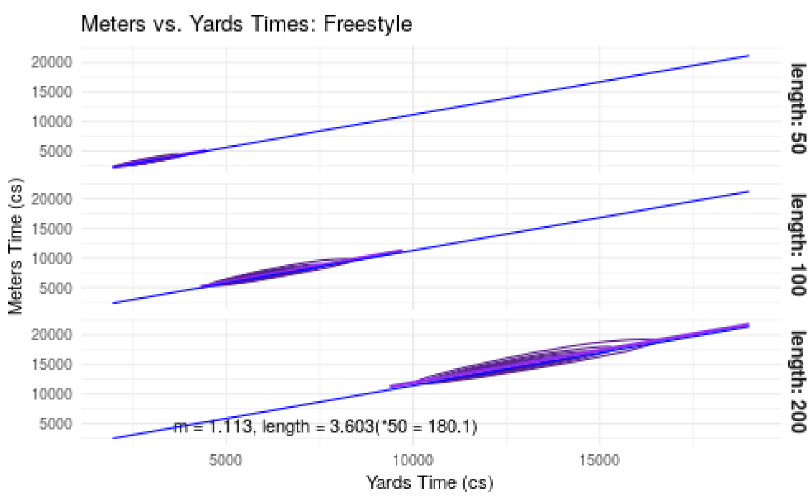

Figure 8: Weighted Least Squares Regression For Freestyle Events

Table 4: WLS model coefficients at 95% confidence vs. de facto model

model name	model a	de-facto a	model b	de-facto b
Butterfly	(1.17 - 1.18)	1.11	(10 - 34)	70
Backstroke	(1.08 - 1.09)	1.11	(337 - 353)	60
Breaststroke	(1.09 - 1.09)	1.11	(369 - 388)	100
Freestyle	(1.11 - 1.11)	1.11	(175 - 186)	80

The slopes of this model remains similar to the model constructed using Ordinary Least Squares (OLS), yet there is a general tightening of confidence intervals, as well as a lowering of model b coefficients. This indicates that when weighted by date, there is less variation of data, which means that the true distribution of times has less error than what was sampled. Furthermore, due to the general reduction of b coefficients, it is clear that some times were omitted that mapped a fast yards time to a slow meters time, which is often common among USA Swimmers swimming LCM rarely.

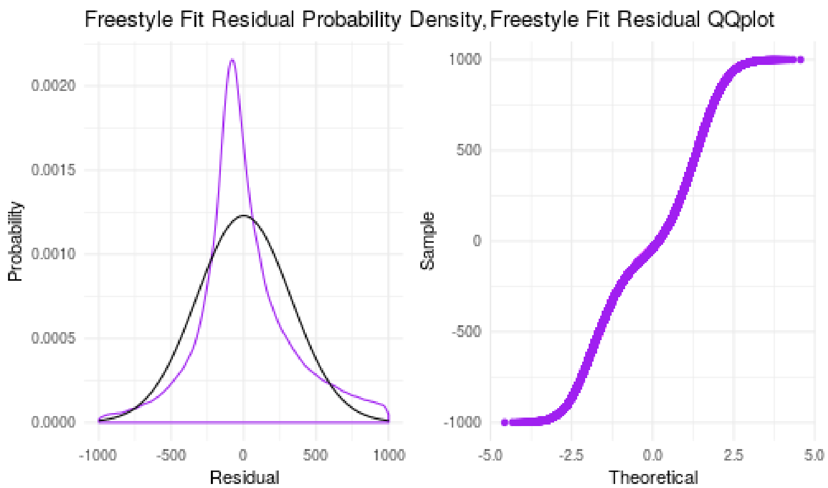

Figure 9: The QQplot and PDF function for the residuals of the WLS fit on Freestyle data. The corresponding normal distribution with the standard deviation of the residuals is shown in black.

The residuals still seem to be slightly right-skewed. Although visually slight, the mode of the residuals is approximately -0.86 seconds, which is a large number in the world of competitive swimming. Although it is less than the previous mode of the OLS model of -1.13 seconds, which indicates a general improvement of the model, ideally the mode should be closer to zero. This is important because while the model should be as accurate as possible for every swimmer, that should not the at the expense of most swimmers getting a predicted time that is more than half a second off what it should be.

4.1.1 WLS with Extra Parameters

It is possible that some of the other parameters in the dataset can also contribute to the prediction. The two other parameters that can be easily added to the model are `age` and `sex`, where sex is one-hot encoded, where male is 1. After adding these parameters to the model, it is simple to check for both an improved model, as well as the significance of each coefficient.

Table 5: Extended WLS model estimated coefficients and respecting Std. Error. for Freestyle Conversion

name	estimate	Std. Error	p value
y_{time}	1.0973499	0.0008974	<2e-16
length	5.4081525	0.0684112	<2e-16
male	27.9759096	2.0396667	<2e-16
age	-9.1630760	0.1860438	<2e-16

It is clear that both coefficients have a significant effect on the model, with an addition of 27 centiseconds to male times, and a subtraction of 9 centiseconds for every year old that a swimmer is. These values intuitively make sense, since men are more likely to develop large power muscles that help pushing off walls but hinder endurance, and older swimmers are more likely to have more experience with swimming long course.

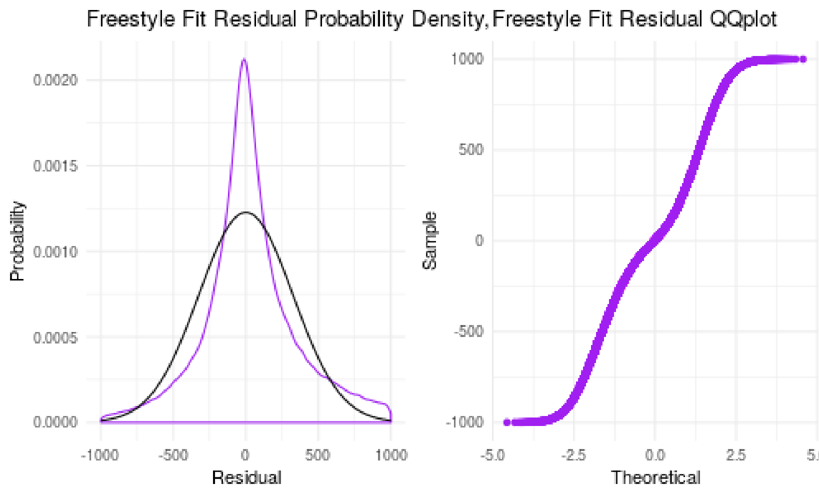

Figure 10: The QQplot and PDF function for the residuals of the extended WLS fit on Freestyle data. The corresponding normal distribution with the standard deviation of the residuals is shown in black.

This brings the mode residual down from -0.86 seconds to -0.11 seconds, which is well within the error that a general prediction should have, since there is so much random chance in the actual swim. The median unweighted residual is around positive tenth of a second off, and the mode residual is around a negative tenth of a second off, this model seems to be quite accurate, as long as no prediction intervals need to be made, since the residuals are not normal. The mode residual is important in this dataset because the residuals look to be more or less centered around the mode, so we choose to create an accurate model for the majority of swimmers. The mean residual, however, is still around 1 second, which means that there is a right skew to the distribution. This model could still be improved by trying to take into account other factors that might impact performance, such as by building a personalized model for each swimmer.

4.2 Fixed Effects Model

A fixed effects model takes into account the intrinsic performance of each swimmer before making a prediction about the swimmer. Since some swimmers are better or worse at Long Course Meters or Short Course Yards, this model would take into account their past performance, and then make a prediction for a conversion time.

This is possible with a Fixed Effects Model. Mainly used in econometrics, it counters omitted variable bias by performing regression within each group, in this case swimmer, instead of across each group [2].

Given N swimmers with S swims each, the classic random effects model which has been in use represents a conversion by the equation:

$$y_{is} = X_{is}\beta + \alpha_i + u_{is}, \text{ for swimmer } i \text{ and swim } s. \tag{3}$$

Where β is the parameter matrix, X is the corresponding coefficient matrix, and u is the error term. In this case, α is unique per swimmer, and represents the swimmer's innate ability. A classical random effects model is unable to account for this term, instead adding it as part of the error term.

Subtracting the mean performance per swimmer, however, is able to account for this:

$$y_{is} - \bar{y_i} = (X_{is} - \bar{X_i})\beta + (\alpha_i - \bar{\alpha_i}) + (u_{is} - \bar{u_i}) \tag{4}$$

Where $\bar{f}$ is the mean f for each swimmer. This reduces to:

$$\tilde{y_{is}} = \tilde{X_{is}}\beta + \tilde{u_{is}} \tag{5}$$

It also follows that the sex parameter can be omitted since its value is constant for each swimmer and is accounted for by the mean value subtraction of the fixed effects model. This also makes intuitive sense, since the sex of a swimmer would only correlated to the swimmer's inherent ability, and therefore accounted for by α.

An inherent flaw of this model, however, is that it is customized for each swimmer, and therefore swimmers with more past swims will have a more accurate model. Other implications of a customized model are also discussed later.

name	estimate	Std. Error	p value
$\mathrm{y}_{\mathrm{time}}$	0.959	0.000428	<2e-16
length	13.384	0.0316	<2e-16
age	-36.777	0.287	<2e-16

Note that the dataset used in this case was the one where every meters time was matched with a yards time, since a fixed effects model required multiple observations for each swimmer. For this reason, non-duplicate swimmers were also removed since they would unfairly compute a residual of 0, since the demeaned values would also be zero. The residuals have a mode of around -8.2 centiseconds, a mean of 0, and median of -1.8, with an IQR of around 3 seconds. This fit therefore simply does not provide much of a benefit over the extended WLS regression, and due to its extended data requirements, is not ideal.

5 Conclusion

5.0.1 The Best Model

While the Fixed Effects Model (see 4.2) provides slightly increased performance over a general weighted least squares model with extra parameters (see 4.1.1), it is more beneficial in context to use the extended weighted least squares model.

A truly customized model such as a fixed effects model is unique to every swimmer, and technically accounts for that swimmer's ability completely independently. This is not ideal, for example, when a meet director chooses to provide converted qualifying times for an event, because such a converter would be unable to provide a general conversion time. While the extended WLS regression also suffers from a similar drawback, it is still more general and the meet director can choose values for both sex and age that make most sense in context.

Furthermore, the fixed effects model relies on a large quantity of individual swimmer data in order to be able to produce the α term and function properly. This is incredibly inconvenient as a swimmer would somehow have to find a way to provide all of this information. Due to the nature of the fixed effects model, the coefficients also do not make much intuitive sense, such as with the `y_time` coefficient in table 4.2, which seems like it should be above 1.

5.0.2 Extending to Model to All Strokes

Using the linear model:

$$\hat{\text{time}}_{\text{meters}} = \beta_1 \text{time}_{\text{yards}} + \beta_2 \text{length} + \beta_3 \text{male} + \beta_4 \text{age} \tag{6}$$

and the inverse taken for $\text{time}_{\text{yards}}$,

$$\hat{\text{time}}_{\text{yards}} = \frac{\text{time}_{\text{meters}} - \beta_2 \text{length} - \beta_3 \text{male} - \beta_4 \text{age}}{\beta_1} \tag{7}$$

The coefficients for the regression are computed to be:

Table 6: Best computed coeffecients for all strokes, distance < 400

stroke	β_1	β_2	β_3	β_4
Butterfly	1.180	0.3363	42.86	-2.016
Backstroke	1.084	6.729	9.446	1.250
Breaststroke	1.091	7.046	29.66	2.949
Freestyle	1.097	5.408	27.98	-9.163

5.0.3 Next Steps

The most important next step is to generate this model over more data, which was not collected in the interest of time. While many major states were represented in the data used to generate this model, more data could be collected to improve the model. More than just collecting more data through swimmingrank, it could also be useful to collect more information about each swim, and make a model using more parameters. Although over fitting is a concern, the sheer amount of recent data points available mitigate this issue.

One possible direction of analysis would be to analyze conversion times by region, to see which regions might be better at long course vs. short course, comparatively. This is something that would be possible using swimmingrank as a data source, but the script used to scrape the website would have to be altered to include region information as well. It can be further extended by analyzing international conversion times, which might lead to more accuracy, since conversion times for US swimmers are more likely to bias towards slower LCM times, especially for younger and more inexperienced swimmers.

Secondly, different machine learning or regression techniques could still be explored in order to further train the model. Apart from varying the method used to weight the residuals in the extended WLS model chosen by this paper, other, more advanced models that do not assume linearity could also be employed.

Lastly, the coeffecients of this model need to be computed for Individual Medley events, as well as distances longer than 400 yards/meters. These coefficients also need to be computed for SCY to SCM conversion, as well as LCM to SCM conversion. These computations were omitted in the interest of time.

References

[1] Course conversion of times/factoring of times. https://support.teamunify.com/en/articles/260-course-conversion-of-timesfactoring-of-times.

[2] Joshua Blumenstock. Mgmt 486: Fixed effects models. http://www.jblumenstock.com/files/courses/econ174/FEModels.pdf. Class Notes.

[3] Thomas Lumley, Paula Diehr, Scott Emerson, and Lu Chen. The importance of the normality assumption in large public health data sets. *Annual Review of Public Health*, 23(1):151–169, 2002.

[4] Brian Stanback. Swim time converter. https://swimswam.com/swimming-times-conversion-tool/.