# Optimizing Power in Cycling

[redacted]

Department of Economics, University of [redacted]

# Table of Contents

## Abstract

Cycling is an optimization problem where *power* is optimized over an interval of *time*. These variables can be quantified and predicted to make pacing during competition and training precise. Modeling optimal power output has many practical uses for endurance athletes who need to optimize the limited resource of anaerobic power. I use my own best-efforts of the past six months to show that modeling power as a function of time is both useful to the average cyclist and can be done easily with data from the popular social media platform Strava. In this paper, I outline three different models of power as a function of time. Additionally, I show that a simple two parameter model can form the intuition necessary to understand how fatigue works in the real world.

## Introduction

I have three goals in writing this paper. The first is to derive and explain a model that can predict a cyclist's *maximum average power output* at a given time *t*. The second is to apply this model to real world power data and make predictions. The overarching goal is to explain the practical uses for this model, mainly how it can give cyclists an edge in training and race strategy by *optimally allocating power*. The main insight from this model is that cycling is an optimization problem where the finite resource of anaerobic power is optimized.

There are three models that give accurate understandable predictions of maximum average power. The first is Monod and Scherrer (1965), which I will focus on. It features two parameters which form a simple basis for modeling maximum average power. The second is Morton (1996) which adds time constants to improve predictions at low values of *t*. The third is Alvarez (2002) which adds exponential decay terms to improve predictions at high values of *t*.

There are more iterations of this basic model that try to refine the predictions at extreme time intervals, but these models are much more complex. The trade-off for describing the world with math is elegant simplicity or maximalism for accuracy. Despite the difficulty of using math to model the world, the predictions models make have tremendous value.

The practical use of this model is very simple. Jones et al. (2010) states that, "Knowledge of an athlete's CP (or CV) and W' (or D') can be used to refine and monitor training protocols and to optimize competition pacing strategies." (1888) Take for instance, an athlete wants to do a 20-minute effort at 80% intensity. They can use this model to predict their 20-minute maximum power, take 80% for that prediction, then they know how many watts they must average over that interval. This adds precision to training, where one can do intervals at different percentages of their maximum average power.

For cycling, power is the finite resource that needs to be optimized. Good economics consists of optimizing finite resources. An informational edge is possible in cycling by modeling maximum average power for any given effort. It is used by smart cyclists to *pace their efforts* and *avoid unnecessary fatigue*. Since cyclists have a limited supply of power, using it inefficiently can be dire. Sub-optimally allocating power can not only lead to poor race results, but injuries, poor training patterns, and unnecessary risk.

## Deriving the Critical Power Model

The simplest and most intuitive model that predicts maximum work capacity is Monod and Scherrer (1965) which has two parameters: *W'* and *CP*. *W'* represents a *finite amount of energy* that is depleted. This is the kind of energy that is used in a sprint which is intense but not sustainable. *W'* is like any other scarce resource like money, time, or fossil fuels. *CP* or critical power represents the *"maximum rate [muscles] can keep up for a very long time without fatigue"* (P. 329). Today, critical power is referred to as the aerobic threshold. This is the point where the legs begin to fill with lactic acid and breathing becomes difficult. Threshold power is the power that can be held for one hour, which is commonly referred to as functional threshold power or "*FTP*". Monod and Scherrer (1965) define the relationship between these two parameters as $P(t) = W'/t + CP$. *P(t)* represents the maximum average power that can be held for a given value of *t* (in seconds). This relationship implies that *P(t)* is a concave up hyperbola that has end behavior resembling $\lim_{t \to 0} P(t) \to \infty$ and $\lim_{t \to \infty} P(t) \to 0$. More subtly, notice that *W'* is being discounted by *t*. *W'*, therefore, depletes over time.

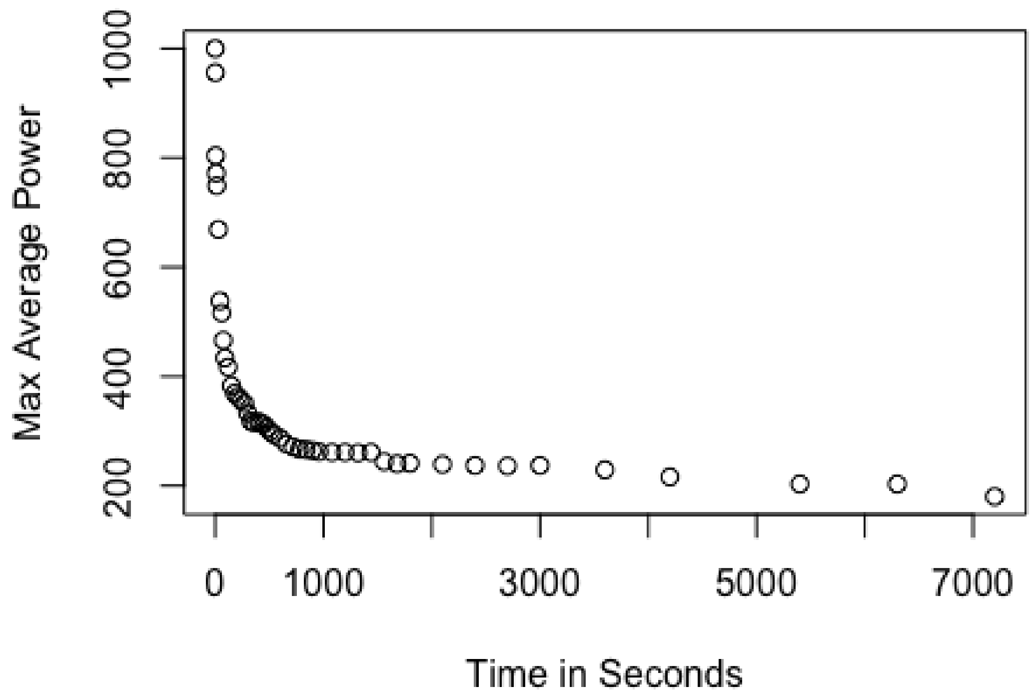
As mentioned earlier, the predictions at values less than three minutes and greater than thirty minutes are problematic for Monod and Scherrer (1965). To remedy the inaccuracies at the asymptotes, Morton (1996) purposes adding short duration limits: $P(t) = W' * [t -$

$\left(W'/(CP - Pmax)\right)\big] + CP$. These added terms improve the model's predictions at short time intervals. Alvarez (2002) adds exponential decay terms to improve the model's predictions at longer time intervals: $P(t) = Pmax * [tau1/t * \left(1 - e(-t/tau1)\right) - tau1a/t * \left(1 - e(-t/tau1a)\right) + CP * \left[tau2n/t * \left(1 - e(-t/tau2n)\right) - tau2a/t * \left(1 - e(-t/tau2a)\right)\right]$. For the purposes of this paper, I will be using Monod and Scherrer (1965).

## Real World Example

Does the model fit real world data? If the model has no connection to reality, then it is completely useless. The data I am using is my own maximum average power values from the past six months on Strava. This dataset, though only 48 observations, consists of my best efforts from over 150 hours of riding. The sample, therefore, is inclusive of hundreds of efforts but only the best efforts have been selected. There are two problems with this dataset. I used two different power meters where one is subject to drive train efficiency loss. This will negatively bias these observations, assuming fitness is constant. My fitness also improved over the six months, my VO2max rose from 65mL/kg/min to 69mL/kg/min, which will bias the data in a positive direction. I decided to keep the data as is because of these two opposing biases which I cannot precisely correct and will only bias the data about one to two percent. These biases will not affect the relationship that Monod and Scherrer (1965) define but may make predictions from this model inaccurate.

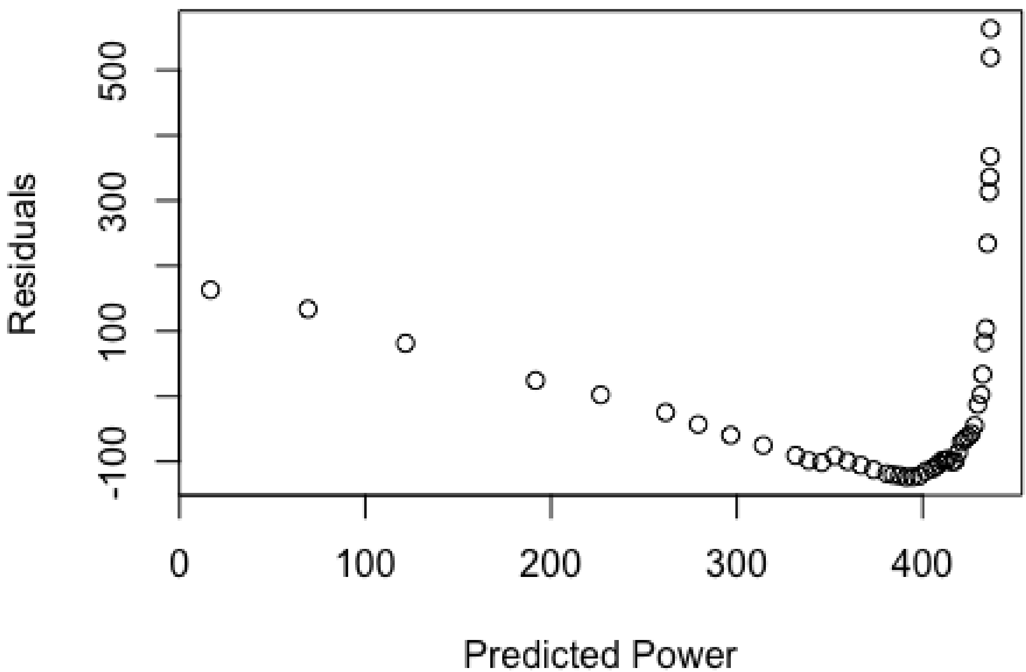
We can see that this simple model fits my best-efforts curve very well. Remember, Monod and Scherrer (1965) predict that we will see a concave up hyperbola that has end behavior resembling $\lim_{t \to 0} P(t) \to \infty$ and $\lim_{t \to \infty} P(t) \to 0$.



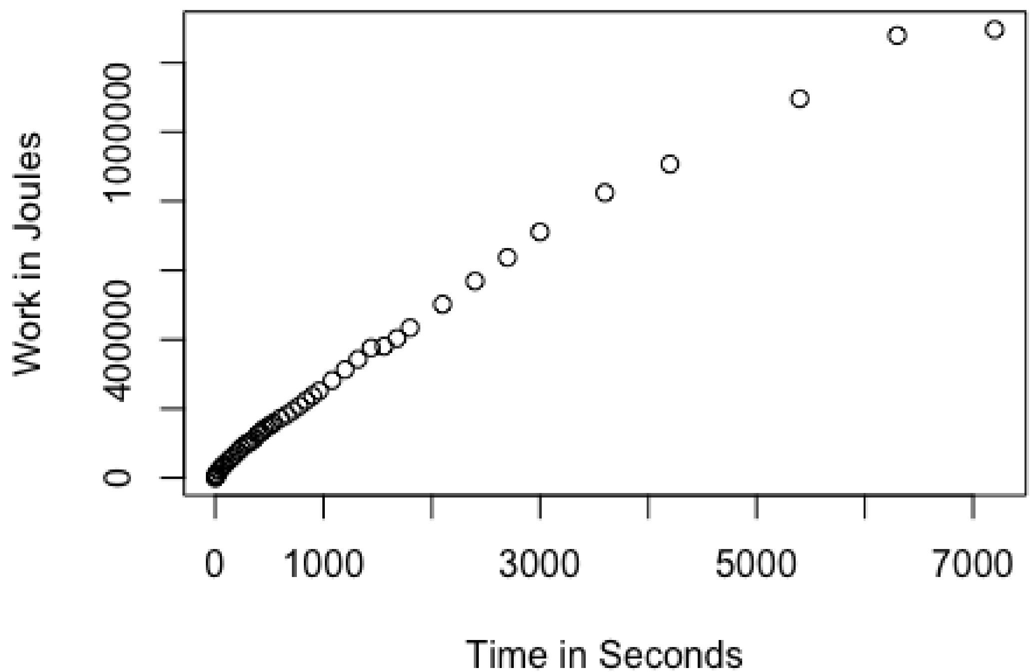

This concave up hyperbola is exactly what Monod and Scherrer (1965) predict. It has end behavior resembling $\lim_{t \to 0} P(t) \to \infty$ and $\lim_{t \to \infty} P(t) \to 0$. Let's see how well a linear regression models this relationship.

```
## lm(formula = power_watts ~ time_sec, data = power_curve)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -124.14 -102.14  -72.52   26.70  563.42
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 436.63630   30.67402  14.235  < 2e-16 ***
## time_sec     -0.05833    0.01507  -3.871  0.00034 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 171.5 on 46 degrees of freedom
```

```
## Multiple R-squared:  0.2457, Adjusted R-squared:  0.2293
## F-statistic: 14.99 on 1 and 46 DF,  p-value: 0.00034
```




This is not a good model for the data, unsurprisingly. Obviously, linearity is violated here. The R-squared is very low and $S_{P(t) \mid t}$ is 171 watts, which would make any prediction with this model useless. Accuracy for this model should be tighter to have any practical use. A sensible transformation is to transform the response variable from watts to $total\ work$ in joules. Since a watt is one joule per second, we can transform the model into $TW(t) = W' + CP * t$. *TW(t)* should be a linear function of time, which makes sense considering that muscles put out more work over time. Let's plot it and see if it looks linear like Monod and Scherrer (1965) predict.



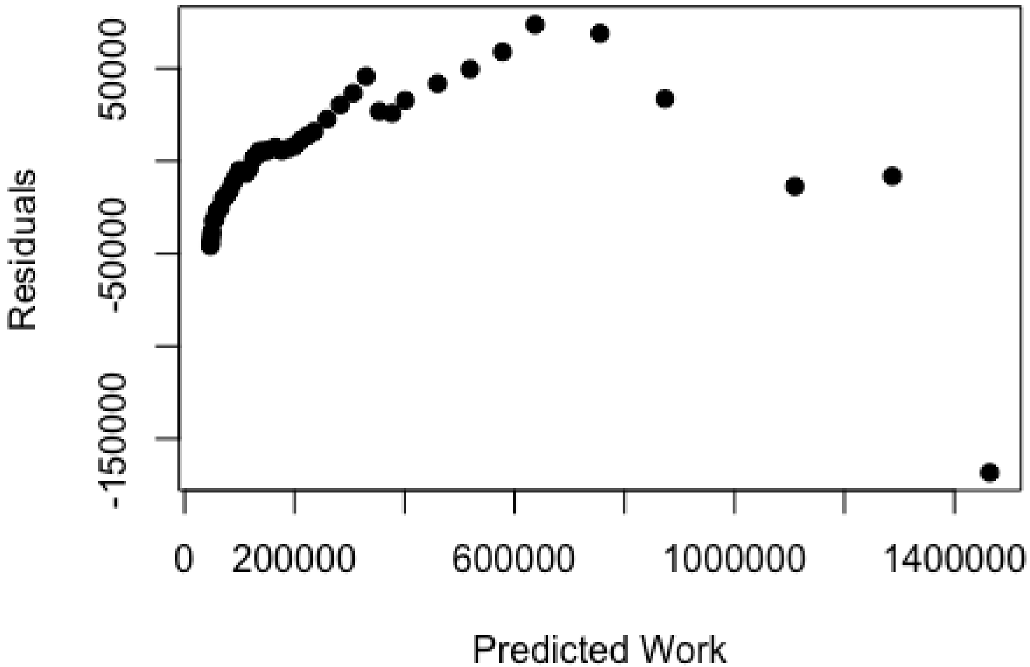

That looks like an excellent transformation. Let's be more precise and model it as a simple linear regression with time predicting total work.

```
## lm(formula = work_j ~ time_sec, data = power_curve)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -168265  -19056    3894   23430   73709
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 46595.205   6999.507   6.657 2.98e-08 ***
## time_sec      196.899      3.438  57.269  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 39130 on 46 degrees of freedom
## Multiple R-squared:  0.9862, Adjusted R-squared:  0.9859
## F-statistic:  3280 on 1 and 46 DF,  p-value: < 2.2e-16
```

The beta value for time is highly significant because it represents *CP*. The intercept is also highly significant and represents *W'*, the finite anaerobic work capacity. An important note here is that *W'* is in *joules*, while *CP* is in *watts* because the slope units of this relationship is in *joules/second* which is the definition of a *watt*.

The predicted values indicate that my threshold power is around 197 watts. My anaerobic potential is 46595 watts, which is impossible. The best sprinters in the world can put out a max of around 2000 watts. Looking at this dataset, I have put out a max of 1000 watts at $t = 1$. This nonsensical value can be explained by the idiosyncrasies of the model. In real life, there is no power output at $t = 0$ but fitting a function to the data adds predictions that do not make sense. Morton (1996) adds time constants and additional terms to correct for these impossible predictions at low values of $t$. The standard error of the prediction for *CP* is three watts which is approximately a 1.5% error. This prediction can be used for practical purposes; however, it is not as good as an accurate FTP test. My actual FTP is around 250 watts. Estimating FTP from a critical power curve is notoriously inaccurate and usually leads to significant under-prediction even when using more accurate models.

Even though the predictor and intercept are highly significant, and the R-squared value is extremely high, I will still check for any OLS assumption issues with our data. I expect linearity to be violated, but homosckedasticity and normality to be ok. I expect to see a funky residual plot. Since the data is time series, there is probably autocorrelation.




Normally, a pattern in the residuals would indicate that linearity is violated, but I suspect that this pattern can be explained by autocorrelation. Homoscedasticity is ok, since there appears to be an equal number of residuals above and below zero. Normality will need to be tested by the Shapiro-Wilk test.

```
##  Shapiro-Wilk normality test
##
## data:  power_curve$work_j
## W = 0.77334, p-value = 3.41e-07
```

Normality is violated, since $H_o$ is rejected. This can be explained by either the data being from a non-*iid* sample or the autocorrelation. An *iid* sample is impractical because this model is for one specific rider and the data must be time series. There appear to be a few points at high values of *t* that are having a large impact on our model. Let's check for outliers and high leverage points.

```
power_curve$leverage = hatvalues(joules_cp_model)
```

The high leverage cutoff for this dataset is $h_i$ >0.083. According to this criterion, there are four points that are considered high leverage. They all occur at high values of *t*: (t = 7200, 6300, 5400, 4200). This result does not surprise me, since all these values were not maximum efforts for me. Monod and Scherrer (1965) is also known to have issues at $t > 1800$. Let's check for any outliers but calculating the studentized residuals for these points. I expect $t = 7200$ to be an outlier.

```
power_curve$stud_res = stdres(joules_cp_model)
qt(.025, 46)

## [1] -2.012896
```

I am concerned about the outlier at $t = 7200$ which came in way below the model's prediction. This can be explained as not being a maximum effort because this value happened to be on a fast ride but not until exhaustion. Larger residuals make sense at these longer intervals because I have only done max efforts at a maximum of 52 minutes. This model is also known to be inaccurate for any *t* value less than three minutes and greater than 30 minutes because it has not been adjusted to account for the end behavior of $P(t) = W'/t + CP$.

I doubt there will any serious problem with outliers. The critical value for a t distribution with 46 degress of freedom is 2.012. According to this criteria $t = 7200$ is considered an outlier, with a studentized residual of 5.13. I will remove this point to make our predictions more accurate, since it does not fit the criteria of a maximum effort. This may be problematic because I have a small sample size, but I expect *TW(t)* to be a robust relationship. Removing this point will not solve the autocorrelation. As I said before, *t* and *t-1* are probably positively correlated, but I will run a Durbin-Watson test to be precise.

```
durbinWatsonTest(joules_cp_model)

##  lag Autocorrelation D-W Statistic p-value
##    1         0.56772     0.4327108       0
##  Alternative hypothesis: rho != 0
```

The Durbin-Watson test is significant, which means there is a problem with autocorrelation. There appears to be positive autocorrelation which makes sense considering that *t-1* will be a good predictor for *t* and two of the regression conditions are violated. This does not necessarily need to be corrected, since the original linear adjustment fits the data extremely well. I will remove the outlier and rerun the regression.

```
## lm(formula = work_j ~ time_sec, data = fixed_power_curve)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -69780 -15522   6286  14559  48746
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 38231.147   4746.774   8.054 2.88e-10 ***
## time_sec      208.008      2.683  77.525  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25840 on 45 degrees of freedom
## Multiple R-squared:  0.9926, Adjusted R-squared:  0.9924
## F-statistic:  6010 on 1 and 45 DF,  p-value: < 2.2e-16
```

The model with $t = 7200$ removed looks a bit better than the previous one. $S_{TW(t)|t}$ is lower, the R-squared is higher, and the F-stat is more significant. Let's correct the autocorrelation now.

```
auto_joules_model = ar.ols(fixed_power_curve$work_j, order.max = 46, demean =
 F, intercept = T)
```

```
##
## Call:
## ar.ols(x = fixed_power_curve$work_j, order.max = 46, demean = F,     inter
cept = T)
##
## Coefficients:
##       1        2        3        4        5        6        7        8
##  0.8461   0.6269  -0.6448   0.2789   0.9096  -0.2291  -0.4420   1.4099
##       9       10       11       12       13       14       15       16
## -0.8392   0.7186  -1.2167  -2.0748   2.8380   0.4134  -1.1607   0.6881
##      17       18       19       20       21       22       23
## -0.9822   1.7949  -0.3478   2.2621  -1.6758  -2.0108  -0.9321
##
## Intercept: -114562 (2.224e-05)
##
## Order selected 23  sigma^2 estimated as  1.544e-11
```

It appears that the optimal order for the autoregression is 23 which is impractical. However, I will create the suggested lagged variables to make the model more precise.

```
for(i in 1:23) {
lag = NA
lag[(i+1):47] = fixed_power_curve$work_j[1:(47-i)]
fixed_power_curve = cbind(fixed_power_curve, lag)
}

colnames(fixed_power_curve) = c("power_watts", "work_j", "time_sec", "lag1",
"lag2", "lag3", "lag4", "lag5", "lag6", "lag7", "lag8", "lag9", "lag10", "lag
11", "lag12", "lag13", "lag14", "lag15", "lag16", "lag17", "lag18", "lag19",
"lag20", "lag21", "lag22", "lag23")
```

It seems like the larger lags are not significant because they correlate less with *t*. After playing around with partial F-tests I settled on this model with four lagged predictors. Adding more predictors made the residual standard error rise without adding much predictive quality. Unfortunately, cross validation is not applicable here because of the lagged variables and small sample size, so I cannot find the optimal model.

```

## lm(formula = work_j ~ lag1 + lag9 + lag10 + lag17, data = fixed_power_curv
e)
##
## Residuals:
##    Min      1Q Median      3Q     Max
```

```
## -33583  -4116    -44   5167  21872
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.144e+04  7.746e+03  -2.768 0.010470 *
## lag1         1.211e+00  5.124e-02  23.640  < 2e-16 ***
## lag9        -1.431e+00  3.831e-01  -3.736 0.000974 ***
## lag10        1.768e+00  4.480e-01   3.947 0.000567 ***
## lag17       -5.402e-01  2.347e-01  -2.301 0.030011 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12710 on 25 degrees of freedom
##   (17 observations deleted due to missingness)
## Multiple R-squared:  0.9985, Adjusted R-squared:  0.9983
## F-statistic:  4221 on 4 and 25 DF,  p-value: < 2.2e-16

## Analysis of Variance Table
##
## Response: work_j
##           Df     Sum Sq    Mean Sq    F value    Pr(>F)
## lag1       1 2.7243e+12 2.7243e+12 16858.5655 < 2.2e-16 ***
## lag9       1 1.6069e+09 1.6069e+09     9.9440  0.004164 **
## lag10      1 1.7392e+09 1.7392e+09    10.7626  0.003047 **
## lag17      1 8.5555e+08 8.5555e+08     5.2944  0.030011 *
## Residuals 25 4.0399e+09 1.6160e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$S_{TW(t)|t}$ is around 50% lower and the R-squared went up! Let's check the residual plot.

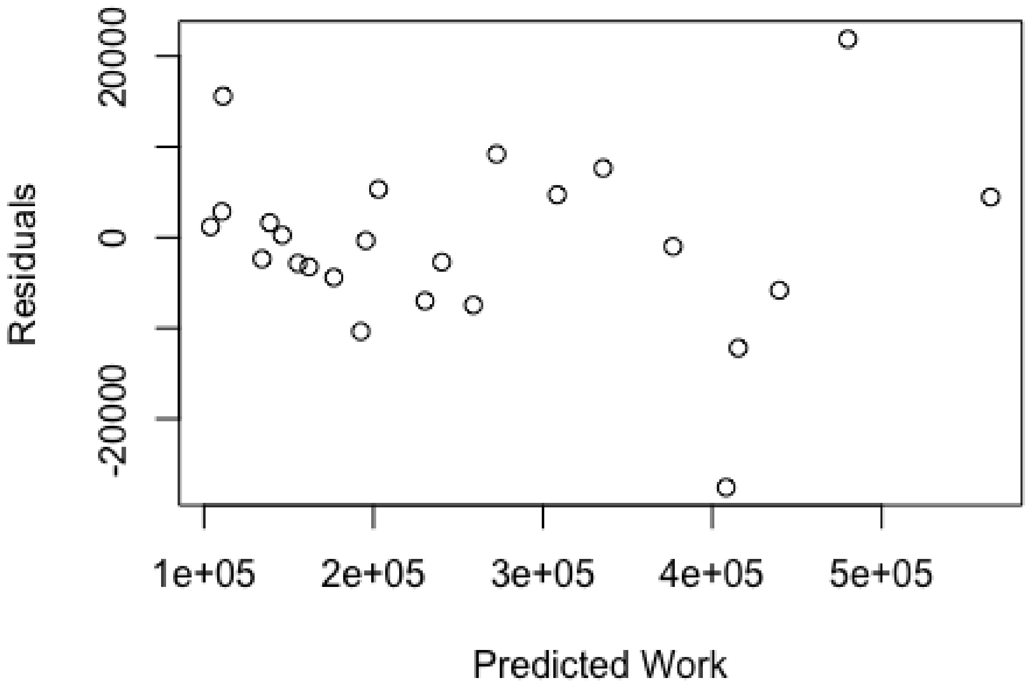
```
fixed_power_curve$auto_res[24:47] = residuals(auto_regression)

## Warning in fixed_power_curve$auto_res[24:47] = residuals(auto_regression):
## number of items to replace is not a multiple of replacement length

fixed_power_curve$auto_pred[24:47] = predict(auto_regression)

## Warning in fixed_power_curve$auto_pred[24:47] = predict(auto_regression):
number
## of items to replace is not a multiple of replacement length

plot(fixed_power_curve$auto_pred, fixed_power_curve$auto_res, xlab = "Predict
ed Work", ylab = "Residuals")
```




This plot looks much better! Finally, let's see if the autocorrelation is fixed or is not a problem.

```
durbinWatsonTest(auto_regression)

##  lag Autocorrelation D-W Statistic p-value
##    1      -0.1065758      2.210763   0.968
##  Alternative hypothesis: rho != 0
```

After correcting for autocorrelation, it seems like the model with lagged terms is the best at predicting total work. Also, the autocorrelation has been eliminated, since the Durbin-Watson test has a very high p-value and a test stat close to two. However, it is not practical to use this model to make the quick predictions necessary to give cyclists an edge, since total work is a function of lagged variables rather than time.

## Predictions

Let's apply this analysis to predict points on my power curve that are not in the dataset with the joules_model. This simple model is not the most accurate, but we need to use it because it is a simple function of time. For more accuracy, the lagged model should be used. But for the purposes of this paper, I will use the simple two parameter model. Let's predict two points within the accuracy range for Monod and Scherrer (1965).

t = 1290, P(1290) = 260

t = 195, P(195) = 366

```
predict(fixed_joules_model, newdata = new_time)
```

```
##         1
## 306561.08  78792.65
```

When we convert to watts, we get $\hat{P}$ = c(237.64, 404.06). These predictions are OK, the respective standardized residuals are 1.116 and 0.287. The reason why the standardized residuals are so low is because $S_{TW(t) \mid t}$ is high. To make this model more accurate, we need more statistical power and the more complex models of Morton (1996), and Alvarez (2002). Strava, Golden Cheeta, Cycling Analytics, and other cycling data platforms use some iteration of Monod and Scherrer (1965) - meaning that the original critical power model is still fundamentally useful.

## Competition

By knowing one's maximum average power at every second, power can be optimally allocated. This gives one an advantage over those who race by feel or race against other people. Being the strongest rider is useless if one's finite anaerobic potential W' is used inefficiently. For future research, it would be interesting to know how one can optimize W' for a race where the conditions vary, and individual efforts are not until exhaustion.

To speculate, if we can determine optimal pacing for each individual rider, competitive strategy will be completely different. The strongest rider will always win, like the strongest runner almost always does on the track. This will be the case because if all riders optimally allocate power, success is a function of strength, *ceteris paribus*. I predict, however, that race strategies will adapt, and cyclists will need to use game theory to determine the best strategy, given that all participants use optimal pacing. Perhaps, competitors will bait each other into expending more of their power and use other strategic tactics. If this development were to take place, modeling cycling will become a game theory problem, rather than the optimization problem it is now.

## Conclusion

In this paper, we derived the critical power model of Monod and Scherrer (1965), analyzed real world power data, and showed its potential to make training more precise. Though the analysis here is simple, there is a potential for greater precision with more sophisticated statistics and complex models.

Cycling is the ideal sport for data analysis, since every aspect of performance can be quantified and optimized. Economics, fundamentally, inquires into the problem of scarcity by seeking to optimize limited resources. In cycling, optimizing limited resources is optimally allocating power. The goal in cycling should not be to ride as possible or keep up with the fastest riders. The goal is to average the optimal amount of power for each effort.

Hopefully, this outline and analysis of critical power has posed more questions than it has answered. Though we do not have a theory of everything that explains our universe, we can have a theory of everything for cycling.

## References

Alvarez-Ramirez, J. (2002). An improved Peronnet-Thibault mathematical model of human running performance. *European journal of applied physiology*, *86*(6), 517-525.

Alvarez (2002) describes aerobic power as reaching its max at $t$ = [300, 400], then plateaus before leveling off at $t$ = 5000. They model this by adding exponential decay rates that adjust to account for a finite store of CP which is different from the assumption of infinite CP in Monod and Scherrer (1965).

Jones, A. M., Vanhatalo, A., Burnley, M., Morton, R. H., & Poole, D. C. (2010). Critical power: implications for determination of VO2max and exercise tolerance. *Med Sci Sports Exerc*, *42*(10), 1876-90.

Jones et al. (2010) describe how Monod and Scherrer (1965) can be used to model critical power in many different species and different kinds of muscles. They describe the history and different iterations of the critical power model, but also go into detail on the empirical evidence for this relationship. They say that the two parameter CP model can be used to predict time to exhaustion from t = [2min, 15min]. They describe how knowing CP and W' can be used to optimize pacing strategies.

Monod, H., & Scherrer, J. (1965). The work capacity of a synergic muscular group. *Ergonomics*, *8*(3), 329-338.

Monod and Scherrer (1965) define maximum average power as a function of W', time, and CP. This relationship forms a concave up parabola with end behavior resembling $\lim_{t \to 0} P(t) \to \infty$ and $\lim_{t \to \infty} P(t) \to 0$. They describe the difference between dynamic and static work. They define CP as a level of power that can be held for a long time without fatigue. They define W' as a finite store of energy that is instantly available.

Morton, R. (1996). A 3-parameter critical power model. *Ergonomics*, *39*(4), 611-619.

Morton (1996) introduces a new parameter to describe maximum power more accurately. They describe how Monod and Scherrer (1965) overestimates CP and underestimates W'. Morton (1996) relaxes the assumption of an asymptote at $t = 0$ to make more accurate predictions at lower values of $t$.

They also introduce a third parameter of maximal instantaneous power (Pmax). Pmax ameliorates the problem of high illogical values of W' in Monod and Scherrer (1965).

Vandewalle, H., Vautier, J. F., Kachouri, Mongi ,K., Lechevalier, J. M., & Monod, H. (1997). Work-exhaustion time relationships and the critical power concept. *A critical review. J Sports Med Phys Fitness*, *37*(2), 89-102.

Vandewalle et al. (1997) describe how Monod and Scherrer (1965) applies to the real world. They describe how it is difficult to predict exhaustion because of the hyperbolic curve. This relationship makes W' and CP prone to bias and inconsistent when transformed. However, they say that critical power is correlated with VO2max and Lactate Steady State which are real biological factors and not abstract terms like W' and CP.