

Efficiency in Results: Employing Data Envelopment Analysis to Determine the Relative Efficiency of Professional Cyclists

Nolan Willoughby
29 April 2024

Executive Summary

Comparing the performances of professional cyclists is difficult due to the complex nature of each cyclist's season objectives and schedule, as well as due to the subjective prioritization of the evaluator of certain results and races over others. To mitigate these issues, a Data Envelopment Analysis was conducted on the set of 84 unique riders who placed in the top ten of at least one 1.UWT race (one day races at the highest level of competition) in the 2024 season as of 24 April 2024. This compared riders according to the number of wins, podiums (top three finishes), and top tens scored in Monuments (the five oldest and most prestigious one day races) and non-Monuments (all other races considered) by both the individual rider and their respective teammates per the number of races entered by each individual rider. In this way, an efficiency score can be generated to determine which riders yielded the greatest results per their number of race days compared to other riders. This method is indifferent to riders' objectives and evaluators' preferences, instead focusing simply on the data and bypassing these typical complications to rider comparison.

Data Collection

In order to conduct this analysis, data must first be collected and organized. In order to do this, results for the 13 1.UWT races held to date in 2024 were referenced from their respective results pages on FirstCycling.com. A table of these races and their classification can be found in Table 1. Using these results, the highest ten finishers from each race were added to the Excel dataset along with their team affiliation at their first encounter and ignored on subsequent encounters. After collecting this list of riders, each individual rider's page on FirstCycling.com was consulted and their individual results from each 1.UWT race were added. These results were then color-coded to be pale green with green text if the rider's individual result was top ten or higher, green with black text if a teammate finished in the top ten but not the individual, white with black text if neither teammate nor individual finished top ten, and pale red if the rider did not attend the race. Race days were then calculated by summing the non-empty cells in each rider's corresponding row of results. The results of this collection can be found in the attached Excel spreadsheet "Willoughby, Nolan, DEA Project Data.xlsm" in the "Raw Results Data" worksheet, and a sample can be found in Figure 1.

Table 1: List of 1.UWT races held in the 2024 season as of 24 April 2024

Date	Race	Classification
28/01/2024	Cadel Evans Great Ocean Road Race	Non-Monument
24/02/2024	Omloop Het Nieuwsblad	Non-Monument
02/03/2024	Strade Bianche	Non-Monument

16/03/2024	Milano-San Remo	Monument
20/03/2024	Classic Brugge-De Panne	Non-Monument
22/03/2024	E3 Saxo Classic	Non-Monument
24/03/2024	Gent-Wevelgem	Non-Monument
27/03/2024	Dwars Door Vlaanderen	Non-Monument
31/03/2024	De Ronde van Vlaanderen	Monument
7/04/2024	Paris-Roubaix	Monument
14/04/2024	Amstel Gold Race	Non-Monument
17/04/2024	Flèche Wallonne	Non-Monument
21/04/2024	Liège-Bastogne-Liège	Monument

Figure 1: Sample Raw Rider Results

Rider	Team	CEGORR	Omloop	Strade	MSR	Brugge De Pi E3	Gent-Wevelg Dwars	Ronde	PR	AGR	Flèche	Liège	Race Days
Tadej Pogačar	UAE				1	3						1	3
Toms Skujinš	Lidl-Trek		27		2	32	8	28	10	39	12	48	9
Maxim van Gils	Lotto Dstny				3	7				20	3	4	5
Matej Mohorič	Bahrain		24		5	6	15	13	DNF				6
Benoît Cosnefroy	AG2R				6	20				16	4	16	5

This results data was then filtered by team to aid in counting each individual rider's relevant individual and team results. The results considered relevant were race wins, top three finishes (podiums), and top ten finishes. These were separated according to whether the race considered was a Monument or non-Monument race. Additionally, due to the team nature of the sport, these same results won by teammates on the same day the individual was present at the same race were also considered. These results were tallied manually for each rider, and can be found in the attached Excel spreadsheet in the "Data Envelopment Analysis" worksheet. A sample of these results can be found in Figure 2.

Figure 2: Sample Relevant Results Tally

Rider	Team	Race Days	Mon Wins	Mon Pods	Mon T10s	NM Wins	NM Pods	NM T10s	Team Mon Wins	Team Mon Pods	Team Mon T10s	Team NM Wins	Team NM Pods	Team NM T10s
Tadej Pogačar	UAE	3	1	1		1								1
Toms Skujinš	Lidl-Trek	9			1		1	1					1	4
Maxim van Gils	Lotto Dstny	5			2		2							
Matej Mohorič	Bahrain	6			1			1						
Benoît Cosnefroy	AG2R	5						2			1			2

With this data, the definition of the components of the Data Envelopment Analysis was relatively straightforward. Namely, "Race Days" was considered as the riders' sole input, while Monument Wins, Podiums, and Top Tens; Non-Monument Wins, Podiums, and Top Tens; Teammate Monument Wins, Podiums, and Top Tens; and Teammate Non-Monument Wins, Podiums, and Top Tens were all considered as individual rider outputs. The riders are then considered as "Decision Making Units" within the context of the Data Envelopment Analysis.

Results and Findings

With this dataset, it is possible to perform Data Envelopment Analysis to obtain efficiency scores for every rider on a scale of zero being completely inefficient to one being totally efficient. Due to the selection process of riders, no rider can be found to have zero efficiency due to the fact that all riders present in the dataset were chosen based on their scoring of at least one result (output in the Data Envelopment Analysis framework). That said, certain riders were certainly found to be more efficient than others.

28 riders were found to perform totally efficiently, earning efficiency scores of one. Their race-to-race result data and summed result types can be found in the attached Excel spreadsheet in the worksheets titled “Efficient Riders’ Raw Results” and “Efficient Riders’ Result Types.” The list of 28 efficient riders and their teams are shown in Table 2.

Table 2: List of Efficient Riders Alphabetically

Rider	Team
Alex Kirsch	Lidl-Trek
António Morgado	UAE Team Emirates
Aurélien Paret-Peintre	Décathlon AG2R La Mondiale Team
Chris Hamilton	Team dsm-firmenich PostNL
Christophe Laporte	Team Visma Lease a Bike
Davide Formolo	UAE Team Emirates
Dorian Godon	Décathlon AG2R La Mondiale Team
Gianni Vermeersch	Alpecin-Deceuninck
Jan Tratnik	Team Visma Lease a Bike
Jason Tesson	TotalEnergies
Jasper Philipsen	Alpecin-Deceuninck
Jasper Stuyven	Lidl-Trek
Jonathan Milan	Lidl-Trek
Lenny Martinez	Groupama-FDJ
Mads Pedersen	Lidl-Trek
Marc Hirschi	UAE Team Emirates
Mathieu van der Poel	Alpecin-Deceuninck

Matteo Jorgensen	Team Visma Lease a Bike
Mikkel Bjerg	UAE Team Emirates
Natnael Tesfazion	Lidl-Trek
Nils Politt	UAE Team Emirates
Phil Bauhaus	Bahrain Victorious
Romain Bardet	Team dsm-firmenich PostNL
Santiago Buitrago	Bahrain Victorious
Simone Consonni	Lidl-Trek
Tadej Pogačar	UAE Team Emirates
Tim Wellens	UAE Team Emirates
Wout van Aert	Team Visma Lease a Bike

Several interesting insights can be gleaned from these riders' data. All six riders who participated in only one 1.UWT race received an efficiency score of one. This is partially due to the nature of the dataset forcing all riders considered to have at least one result to consider. Coincidentally, all riders considered with only one race day were the only riders from their respective teams to finish in the top ten on their single race day. Had a single of these riders had such a teammate, the other five would have received efficiency scores less than one and be found inefficient.

Additionally, several riders with memorable results from the spring are missing from the list of efficient riders. The most glaring of these are Michael Matthews and Luca Mozatto, second in Milano-San Remo and De Ronde van Vlaanderen respectively, earned respective efficiency scores of 0.154 and 0.182. While second place in a Monument would, in most analyses, indicate a successful spring, these riders were punished by their high number of race days and lack of results, especially when compared to the riders joining them on the podiums. The spring campaign of Tadej Pogačar specifically, winner of both Strade Bianche and Liège-Bastogne-Liège and third at Milano-San Remo in three race days, meant that anyone podiuming a Monument would need to either have done so in less than three races, matched Tadej in results, or have helped teammates achieve a result that same day. Similarly, Maxim Van Gils, top ten twice in Monuments and top three twice in other races over five race days (he finished 20th that fifth day), earned an efficiency score of just 0.2. This score comes courtesy of Natnael Tesfazion's second in the Cadel Evans Great Ocean Road Race and Aurélien Paret-Peintre's fifth at Liège-Bastogne-Liège. With weights of 0.5 for each of these two, one race day yields 0.5 non-Monument podiums and 0.5 Monument top tens, while Van Gils only manages 0.4 of each.

Lastly, the case of riders with relatively anonymous spring campaign's whose efficiency equaled one should be noted. The most notable of these is the case of Alex Kirsch, whose tenth at the E3 Saxo Classic earned him his spot in the analysis. However, his only other individual result of even remote consideration was 20th at Omloop Het Nieuwsblad, while the rest of his spring was spent failing to finish at Gent-Wevelgem and Dwars Door Vlaanderen, and finishing an anonymous 102nd and 103rd at Milano-San Remo and Classic Brugge-De Panne respectively. However, Kirsch is neither known nor paid for his individual results. Rather, his role is to aid his teammates in achieving their objectives, which is how his efficiency was earned. Alex Kirsch helped Jasper Stuyven to seventh at Omloop, eighth in San Remo, and second at E3; Mads Pedersen to fourth in San Remo and a win at Gent-Wevelgem; Jonathan Milan to fifth at Gent-Wevelgem and seventh at Dwars Door Vlaanderen; Toms Skujiņš to eighth at E3; and Simone Consonni to fifth at Brugge-De Panne. This meant that any time Alex Kirsch has started a 1.UWT race in 2024 thus far, he has helped at least one teammate to at least a top ten finish. Many riders made the list in a similar way, however these riders generally had greater individual success than Kirsch.

Conclusion

While some interesting results have been noted that could signify a shortcoming in the application of Data Envelopment Analysis to this case, the concept serves as a solid foundation from which to conduct rider analysis. Maxim Van Gils, Michael Matthews, Luca Mozatto, and others had excellent springs for themselves, their teams, and their sponsors but failed to garner recognition from this analysis. Meanwhile, Alex Kirsch narrowly found entry into the dataset and managed to score total efficiency based almost entirely on the accolades of his teammates. While parameters could be added, adjusted, or removed to reward the former group and/or to punish the latter, the point of Data Envelopment Analysis is to identify "Decision Making Units" whose performance cannot be exceeded by some equivalent combination of other "Decision Making Units" at the cost of equivalent effort and resources. Simply put, the spring campaigns of the former group could not match the equivalent campaigns of the very best riders in the peloton. Additionally, the latter group, while seemingly paling in comparison to the former, were incredibly efficient in delivering glory to their teams and their sponsors. In this way, the Data Envelopment Analysis conducted gave a comprehensive outlook on the relative efficiency of the 84 professional cyclists who finished tenth or better in the 13 World Tour-level one day races held to date in 2024.

References

Results retrieved from [FirstCycling | Complete and up to date Cycling Database](#)
Linear Programming performed in [Excel](#) using [Opensolver](#)