Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Forecasting Tour de France TV audiences: A multi-country analysis

Daam Van Reeth*KU Leuven - Faculty of Economics & Business, Warmoesberg 20, Brussels 1000, Belgium*

ABSTRACT

The Tour de France is the world's biggest cycling event. The race attracts up to 25 million TV viewers per stage worldwide. In this article, we forecast TV audiences for individual stages of the Tour de France for five European countries where cycling is popular: Belgium, Denmark, France, The Netherlands and Spain. The predictions follow from a multivariate ordinary least squares regression model that explains historical viewing habits for the Tour de France as a function of attributes of the individual stages, and contextual information such as TV channel and day. Although the accuracy of the forecasts changes from year to year and can be very different between TV markets, in most cases our predictions clearly outperform forecasts based on naive models. Our findings illustrate that a large part of the variation in TV viewership is determined by how the race route is designed by the race organizer, independent of actual race developments.

© 2018 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

TV viewership is a significant indicator of interest in major sports events like the Olympic Games, the Super Bowl or the Champions League. The organizers eagerly report – sometimes inflated – TV audience sizes, in order to emphasize the wide appeal and success of their competitions. However, TV audience data also have more profound economic consequences. For instance, the number of people that watch a live sports broadcast can be used to set prices for commercials, to evaluate sponsorship decisions, or to value future TV broadcasting rights for sports competitions. The problem is that the actual TV audiences cannot be determined precisely until after the event has taken place, and, thus, well after the TV commercial air time has been sold or the broadcasting contract negotiated. It is therefore extremely useful to have accurate audience forecasts long before the broadcast of the event, in order to avoid a mismatch between the estimated and true audiences. Of course, such forecasts can only be based on ex ante determinants of the viewership; i.e., characteristics of the broadcast that are known prior to the event, such as the type of game or race, the moment of the broadcast, or the broadcast channel.

This paper forecasts TV audiences for the world's biggest cycling event, the Tour de France. The race consists of 21 day-long stages over a 23-day period – with two days of rest – and is run yearly in July. The rider who takes the shortest time overall to complete all of the stages wins the competition. Individual Tour de France stages differ in length and route. The heterogeneous nature of the event has important implications for TV interest in it. There are roughly three types of stages: flat stages, mountain stages, and time trial stages. Flat stages are easy to ride and tend to end in a bunch sprint. They are of almost no importance to the overall Tour de France win, and public interest in these stages is usually rather limited. When the riders have to climb steep mountain passes, the race becomes much harder, and large time differences between riders are created. Thus, mountain stages are crucial to the overall Tour de France result and attract much larger audiences. While flat stages and mountain stages are raced collectively, time trial stages are ridden individually, or sometimes in teams. The absence of direct rider interaction makes time trial stages less spectacular to the TV public, but such stages are still very important to the overall Tour de France result, as they can result in significant time differences between riders (Van Reeth, 2013, p. 43).

We predict TV audiences for all 21 stages of the Tour de France. The forecasts are based on a model that uses

E-mail address: daam.vanreeth@kuleuven.be.

<https://doi.org/10.1016/j.ijforecast.2018.06.003>

0169-2070/© 2018 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

detailed historic TV audience data for the race as the dependent variable, and information that is made public many months before the event takes place as independent variables. We begin by forecasting TV audiences for the 2017 Tour de France, then, to better evaluate the validity of the initial findings, extend the analysis with forecasts of the 2016 and 2015 Tour de France TV audiences (based, of course, on a model estimated on data up to 2015 and 2014, respectively). The study focuses on five countries for which we were able to collect sufficient reliable and consistent historic data, namely Belgium, Denmark, France, the Netherlands and Spain. For Belgium, we split the language regions into two separate television markets, because TV viewership behavior differs greatly between Flanders (the Dutch-speaking Northern part of Belgium) and Wallonia (the French-speaking Southern part of Belgium).

The remainder of the paper is organized as follows. We continue with a survey of the literature in Section 2. Next, information about the data and the methodology is provided in Section 3. The covariates of the model are discussed in Section 4. Section 5 then presents the forecast results, and Section 6 concludes with a discussion of the main findings and policy implications.

2. Literature

There exists a substantial body of academic literature on forecasting TV audiences. For example, both Danaher, Dagger, and Smith (2011, pp. 1220–1221) and Yeh (2015, p. 834) present extensive overviews, and, combined, they document over 20 TV forecast studies published between 1980 and 2014. Many of the reported studies predict future viewership by applying traditional statistical methods to datasets that usually include historic TV viewing data as a dependent variable and a set of independent variables that are primarily time-, calendar-, or program-related. In addition, methods that involve the use of decision trees (e.g. Danaher & Dagger, 2012; Meyer & Hyndman, 2006), neural networks (e.g. Meyer & Hyndman, 2006; Nikolopoulos, Goodwin, Patelis, & Assimakopoulos, 2007; Weber, 2002) or nearest neighbour analysis (e.g. Nikolopoulos et al., 2007) have been introduced in this context as well.

Although the aforementioned studies are insightful from a methodological point of view, most forecast prime-time ratings for channels, rather than ratings for programs. More in line with our research context are a couple of recent studies that have focused on specific programs or program formats. Hunter III, Chinta, Smith, Shamim, and Bawazir (2016) develop an empirical model for predicting the audience of new dramatic television series. In particular, they examine the roles of previously untested predictors of the performances of new television shows, such as the originality of the show's concept, the track record of success of the show's creative team, and the size of the conceptual network that is created from the teleplay of the pilot episode (Hunter III et al., 2016, p. 13). All of these predictors are known prior to the decision to greenlight the pilot script. In recent years, online information has been used increasingly for forecasting TV ratings. Fukushima, Yamasaki, and Aizawa (2016) predict

the ratings of TV dramas by considering the cast and the staff involved with them, whose popularity was measured by counting the number of hits received by the Wikipedia pages of the actors and staff, as well as the numbers of tweets related to them. Yeh (2015) uses online information for predicting TV ratings for movies. Online word-of-mouth data and program variables are used as potential predictors. This method is different from the conventional statistical approach of using program variables alone as predictive variables (Yeh, 2015, p. 842). Zhang, Wu, Qiao, and Fang (2016) present a framework for forecasting the TV ratings of seasonal entertainment TV shows in China. As supplement to conventional data, the TEI@I methodology is used. The authors claim that the utilization of big data from the Internet results in noticeable improvements in the forecast accuracy (Zhang et al., 2016, p. 2905).

Danaher et al. (2011, p. 1231) state that the forecasting of TV audiences for live sport is extremely challenging. The literature on forecasting TV audiences for live sports broadcasts is indeed sparse, although academic research into TV viewership for sport has grown substantially over the past decade. The seminal article by Forrest, Simmons, and Buraimo (2005) on TV viewership for the English Premier Football League inspired many similar studies for European football. At the same time, American sports economists started to publish research results on TV audiences for typical American sports like American Football and NASCAR racing (Van Reeth, 2015, p. 116). As a result, over 40 papers on TV audiences for sports were published in major academic (sports) journals between 2005 and 2018. However, only three of those articles discuss forecasts of TV audiences specifically. Feddersen and Rott (2011) analyse TV viewership behavior for games by the German national football team on German TV. They propose a model that achieves high explanatory and predictive power, and conclude from this that, although the conventional wisdom in media management practice that TV broadcasts follow their own rules, such that their success cannot be predicted well because of their highly complex nature, might be true for some entertainment products, it does not apply to the broadcasting of live football matches on German television (Feddersen & Rott, 2011, p. 366). Kim and Kim (2012) analyse TV viewership behavior for FIFA World Cup soccer games on Korean TV and demonstrate that the TV ratings for each game can be predicted very accurately in a model that employs only seven ex ante variables. They claim that such a simple model outperforms even the complex formula that is currently used for pricing air time for TV advertisement in Korea (Kim & Kim, 2012, p. 88). Finally, Nikolopoulos et al. (2007) forecast TV audiences for live sports broadcasts, although their approach is methodological (i.e., comparing the accuracy of various forecasting techniques) rather than managerial. They predict Greek TV audience shares for a heterogeneous set of 12 football and basketball games, and conclude that the multiple regression models perform relatively badly as a forecasting tool and are outperformed by methods that are much simpler conceptually, such as a bivariate regression model, nearest neighbour analysis or even human judgment (Nikolopoulos et al., 2007, p. 364). Note however that these findings should be treated cautiously, since the analysis is based on a small dataset of only 34 prior events.

Table 1
Dataset.

TV market	Number of observations	Number of years (period)	Selected for forecasts
Australia	124	6 (2010–2015)	No
Belgium – Flanders	378	18 (2000–2017)	Yes (2015, 2016, 2017)
Belgium – Wallonia	378	18 (2000–2017)	Yes (2015, 2016, 2017)
Denmark	330	16 (2002–2017)	Yes (2015, 2016, 2017)
France	112	6 (2012–2017)	Yes (2017 only)
Germany	125	6 (2006–2008, 2015–2017)	No
Italy	126	6 (2012–2017)	No
The Netherlands	188	9 (2007, 2010–2017)	Yes (2015, 2016, 2017)
Norway	83	4 (2014–2017)	No
Spain	185	10 (2008–2017)	Yes (2015, 2016, 2017)
Sweden	315	15 (2003–2017)	No
Switzerland – German speaking part	48	3 (2015–2017)	No
United Kingdom	161	8 (2010–2017)	No
United States	59	3 (2009, 2015–2016)	No

Notes: Each Tour de France consists of 21 stages. The number of observations is not always equal to the number of years times 21 because TV viewing data were not always available for each of the individual stages in each country.

Our paper is the first on the prediction of Tour de France TV audiences, but a different aspect of the race has already been the subject of a fascinating forecasting study. Hannas and Goff (2005) use concepts from physics to predict successfully the stage-winning times for the Tour de France in both 2003 and 2004. The model that the authors developed missed the actual sum of stage-winning times by just 1.77% (2003) and 0.05% (2004). Although the focus of their research (forecasting a rider's times) is very different from ours (forecasting TV audiences), their decision to model individual stages in order to obtain reliable forecasts for the entire Tour de France is very similar to the approach that we present below.

3. Data and methodology

Our forecasts are based on a self-developed dataset with historical average TV audiences – in absolute numbers – for broadcasts of individual Tour de France stages worldwide. The data were collected from various different sources over the years, and were retrieved from the websites of official TV audience measurement companies (e.g. www.barb.co.uk for the United Kingdom, www.cim.be for Belgium or www.kijkonderzoek.nl for the Netherlands) or media watchers (e.g. [https://www.davidemaggio.it](http://www.davidemaggio.it) for Italy or <http://ecoteuve.eleconomista.es> for Spain), or were received through direct contact with media companies or TV broadcasters (e.g. for Denmark). The entire dataset contains over 2500 observations from 13 countries (Table 1).

The data cover a period of at least 15 years for countries such as Belgium (2000–2017) and Denmark (2002–2017). Conversely, there are other countries for which limited information from a few years has been obtained so far (e.g. four years for Norway, three years for the United States). Here, we decided to only include countries with over 100 data points. Since there are 21 stages in each Tour de France, this corresponds to at least five years of observations, a period that we judged necessary for the production of reliable forecasts. The countries selected for the analysis are Belgium (divided into two language-based television markets, Flanders and Wallonia), Denmark, France, the Netherlands and Spain. Note that, although we also have access to TV viewing data for Sweden since 2003, the

country was not included in this study because the data refer to viewership for Eurosport only. The Swedish TV audiences for the sports-themed channel are too small for a sound statistical analysis. There were also four other countries with a sufficient number of observations that were not included in our analysis, for various reasons: Australia (missing viewership data for 2016 and 2017), Germany (a time gap of seven years between the initial observations and recent viewership information), Italy (heterogeneous viewership information because of different types of live broadcasts) and the United Kingdom (viewership data that refer not to the live broadcast, but to the primetime highlights show).

Our predictions will be based on historical viewing habits, the race route and different attributes of the Tour de France, and contextual information such as the TV channel and day. Therefore, we started by developing a multivariate regression model for each country in order to identify the determinants of TV viewership behavior. We found that developing a separate model for each country resulted in much better forecasts than having a combined model with country dummy variables. This is in line with the findings of Danaher et al. (2011, p. 1236) on the forecasting superiority of a separate model for each TV channel over a combined model with channel dummy variables. The decision to use a multivariate regression model was made after a comparison of findings in literature. The decision was based in part on a general conclusion that was drawn from past TV rating prediction efforts, namely that because traditional econometric methods often fit very well, with high R^2 values, a simple statistical model that involves a small list of independent variables can be used to explain TV ratings accurately (Yeh, 2015, p. 833) (Kim & Kim, 2012, p. 88). Another study claims that although there is still room for some improvement in forecast accuracy, it is difficult to achieve substantial gains from highly complex models (Danaher et al., 2011, p. 1237). A more fundamental reason for our selection of a multiple linear regression technique is that there are no theoretical formulations for interval forecasts for alternative approaches such as nearest neighbour methods or artificial neural network techniques. This may limit the usefulness of these alternative forecasting methods when decision makers require

information on the level of uncertainty that is associated with point forecasts (Nikolopoulos et al., 2007, p. 366). Finally, the results of linear regression techniques and its eventual policy implications are more succinct and easier to explain to interested stakeholders than more complex forecast models (Yeh, 2015, p. 833).

One potential drawback of regressions is that their effectiveness and validity are limited by a number of assumptions, such as normality and independence among the input variables (Yeh, 2015, p. 833). However, a Shapiro–Wilk test showed that normality is not a cause for concern in our dataset, and the regression results did not suffer from multicollinearity, as will be explained below.

4. Covariates

We turn to two academic studies on the determinants of TV audiences for professional road cycling in order to decide upon the covariates of our model. Van Reeth (2013) analyses TV audiences for Tour de France broadcasts in Flanders between 1997 and 2012, and Rodríguez, Pérez, Puente, and Rodríguez (2015) study TV audiences for a broad range of cycling races that were broadcast on Spanish television between 2007 and 2011. The covariates used in the two studies can be categorized into five groups: stage type covariates, calendar covariates, broadcast related covariates, competition related covariates and country-specific covariates. We describe how these covariates are operationalized in our models below, and refer to Table 2 for a complete overview of all of the variables and their definitions.

4.1. Stage type covariates

The aforementioned heterogeneous nature of Tour de France stages is highly relevant in understanding the TV viewership for the race. The intense bike action and the spectacular scenery in the mountain stages generate a much higher TV viewership than flat stages, in which sometimes little of interest happens until the last couple of kilometers. In our models, the reference type of stage is a flat stage, and five dummy variables are used to define the other types of stages: ‘high mountain stage’, ‘low mountain stage’, ‘individual time trial’, ‘team time trial’ and ‘mountain time trial’. An additional dummy variable (‘summit finish’) is introduced for stages that finish on a mountain top. Typically, the Tour de France consists of about ten flat stages, six to eight high mountain stages, about half of which finish on a mountain top, a couple of low mountain stages, and one or two individual time trial stages. Team time trial stages and mountain time trial stages are rather the exception in the Tour de France.

4.2. Calendar covariates

Time of day and day of the week are important exogenous covariates in explaining TV ratings (Danaher et al., 2011, p. 1236). However, time of day is largely irrelevant in this study, as Tour de France stages are broadcast at approximately the same time every day, namely between 2 and 6 pm CET. However, since Tour de France stages

are broadcast in the afternoon in European countries, it is important to draw a distinction between weekdays and weekends. Saturday and Sunday stages are expected to have much larger TV audiences because more people are at home on the weekend. Thus, we include ‘Saturday’ and ‘Sunday’ dummy variables in our models and define a weekday stage as the reference situation. Since we also expect an extra interest in both the first and final stages of the Tour de France, the dummy variables ‘opening stage’ and ‘closing stage’ were created. These variables capture any novelty effect of a fresh Tour de France on the opening day, and a potential parade effect of the ceremonial final stage, similar to the closing ceremony at the Olympic Games. We also expect TV audiences to increase as the Tour de France develops over the course of 23 days, in part because the battle for the overall win intensifies as the race nears its end, and in part because July, when the Tour de France takes place, sees increasing numbers of people at home for the summer break. We account for this growth in TV interest by introducing ‘week 2 stage’ and ‘week 3 stage’ dummy variables into our models, with the first week stages defined as the reference stages.

4.3. Broadcast related covariates

Broadcast related covariates are necessary for capturing TV viewership effects that result from the broadcasting company’s decisions of how to assign the channels. Obviously, there is no need for such covariates when the Tour de France live TV coverage remains on the same channel year after year. In our dataset, this was the case for France, the Netherlands and Denmark, with all broadcasts being on the major public channels France2, NPO1 and TV2, respectively. However, in the remaining countries, channel decision effects are important. In Spain, the majority of stages are broadcast on the sports-themed public channel Teledeporte, but a selection of stages are shown on the well-watched major public channel La1, or on the less-watched secondary public channel La2. In Belgium, the majority of the stages are broadcast on public channels EEN (Flanders) and La Une (Wallonia), but occasionally TV viewers have to switch to the smaller secondary public channels Canvas (Flanders) and La Deux (Wallonia). We introduced the dummy variables ‘smaller channel broadcast’ and ‘bigger channel broadcast’ to help model these channel effects properly. Neither of these variables is entirely exogenous, since the expected viewership may affect the decision as to how the channels should be assigned; however, they do not suffer from the problem of endogeneity, since the realized viewership cannot affect these decisions ex post (Kim & Kim, 2012, p. 83).

There was also a need for a ‘specific broadcast effect’ dummy variable in the models for Denmark and Spain. TV2, the channel that broadcasts the Tour de France in Denmark, became subscription-based in 2012, resulting in smaller TV audiences for all its programs. In Spain, TV viewers for the podium ceremony that follows the stage were included in the reported average TV audience for the stage until 2012, while such was not the case from 2013 on. It is important to control for this computational difference, as many TV viewers drop out once a cycling race has finished and

Table 2
Covariates.

Category	Variable	Definition	Ex ante	Ex post
Stage type covariates	High mountain stage	Dummy for mountain stages in the Alps, the Pyrenees or on the Mont Ventoux	X	
	Low mountain stage	Dummy for all other mountain stages	X	
	Summit finish	Dummy for stages with a mountain top finish	X	
	Individual time trial	Dummy for individual time trial stages	X	
	Team time trial	Dummy for team time trial stages	X	
	Mountain time trial	Dummy for mountain time trial stages	X	
Calendar covariates	Opening stage	Dummy for the first stage of the Tour de France	X	
	Closing stage	Dummy for the last stage of the Tour de France	X	
	Week 2 stage	Dummy for a weekday stage in the second week of the Tour de France	X	
	Week 3 stage	Dummy for a weekday stage in the third week of the Tour de France	X	
	First Sunday stage	Dummy for the second stage of the Tour de France	X	
	Mid Saturday stage	Dummy for a Saturday stage, excluding the opening stage (which is also a Saturday stage)	X	
	Mid Sunday stage	Dummy for a Sunday stage, excluding the first Sunday stage and the closing stage (which is also a Sunday stage)	X	
Broadcast related covariates	Smaller channel broadcast	Dummy for a broadcast on a smaller channel than usual (niche channel or secondary channel)	X	
	Bigger channel broadcast	Dummy for a broadcast on a bigger channel than usual (major channel)	X	
	Specific broadcast effect	Dummy to capture special broadcast effects, like a change to subscription TV (Denmark) or the broadcast of the podium ceremony (Spain)	X	
Competition related covariates	Doping	Dummy for the first stage following the release of doping news related to a Tour de France cyclist		X
	Dominance	Equals the number of stages already won in a given Tour de France by the Tour de France winner of the previous year		X
	Suspense	Dummy for third week mountain stages or time trial stages with at most 90 s of time difference at the start between the top two riders in the general classification		X
Country-specific covariates	National day	Dummy for stages held on a national day (here, July 21 for Belgium and July 14 for France), when not on a weekend	X	
	Regional day	Dummy for stages held on a regional day (here, only July 11 for Flanders), when not on a weekend	X	
	Local passage	Dummy for Tour de France stages crossing the country for which the forecast is made (excluding France)	X	
	National participants	Equals the number of participants (still) in the Tour de France of the country for which the forecast is made		X
	National success	Dummy for stages with a cyclist of the country for which the forecast is made either wearing a leadership jersey or in competition for a top-10 placing in the general classification		X
	Local temperature	Dummy for stages on a day with extreme temperatures in the country for which the forecast is made (extreme temperatures were defined as temperatures at least two standard deviations below or above normal)		X
	Local rain	Dummy for stages on a rainy day in the country for which the forecast is made (rainy days were defined as days with at least 20mm of rain)		X
	Year dummies	A separate dummy for each year in the dataset, to capture year-specific effects not covered by the other covariates		X

the result is known, and do not bother about the podium ceremony.

4.4. Competition related covariates

It is well-known in sports economics that the more balanced the teams and the more uncertain the outcome of the competition, the higher spectator interest is likely to be. Thus, our models include competition-related variables that measure the competitive balance and outcome uncertainty in the Tour de France, as per Van Reeth (2013, pp. 46–47). The 'suspense' variable marks stages that appear to be crucial to the overall Tour de France win. Since the time differences between the top riders in mountain stages and time trial stages are usually well below two minutes at the finish, the outcome uncertainty with respect to the final Tour de France winner is only significant when the time differences in the overall standings are not too large. Thus, suspense stages are defined as third week mountain or time trial stages where the difference between the top two riders at the start of the stage is below 90 s. The number of stages already won in a given Tour de France by the Tour de France winner of the previous year acts as a second outcome uncertainty variable. If a defending Tour de France winner wins multiple stages, he shows that he is probably strong enough to reclaim the title. The impact of such a more predictable outcome is measured by the 'dominance' variable.

Since doping scandals are a potential threat to TV interest in cycling, a variable 'doping' was also included in the model. This covariate measures the instant reaction of TV viewers to the release of Tour de France-related doping news, i.e., the impact on TV audiences in the first stage after the breaking of the doping news.

4.5. Country-specific covariates

A last group of covariates captures country-specific effects, and includes variables that differ in value between countries. For example, a higher viewership is expected when an element of patriotism is involved in the broadcasting of a Tour de France stage. As a consequence, both the number of local riders ('national participants') and their success ('national success') are used as covariates in the models. Hosting the start or finish of a Tour de France stage will also increase national TV viewing for the event. This justifies the use of a dummy variable 'local passage' in the models for Belgium, Spain and the Netherlands; only Denmark did not host a Tour de France stage. Furthermore, the consumption of television broadcasts depends strongly on weather conditions (Feddersen & Rott, 2011, p. 359; Van Reeth, 2013, p. 48). Since we were unable to collect detailed daily weather data for Denmark, France or Spain, the variables 'local temperature' and 'local rain' were included in the models for Belgium and the Netherlands only. Finally, a festive national or regional day may potentially increase the TV viewership because more people are at home on such days. Thus, the covariates 'national day' and 'regional day' were used in the models for France, Flanders and Wallonia. Of the five countries in our analysis, both France (July 14) and Belgium (July 21) celebrate a national

day over the course of the Tour de France. In addition, the Belgian region of Flanders also celebrates a regional day on July 11.

Forecast models also often include linear and quadratic trend variables. However, after taking a closer look at the past Tour de France TV audiences in each of the countries selected, we decided not to include such variables in our analysis, as no clear increasing or decreasing trend in TV audiences is present in any of the countries, with the year-to-year changes in TV viewership seeming to follow a more erratic pattern. Therefore, we judged it better to include 'year dummies' in the model for each country instead, to control for any news or events that may affect Tour de France TV viewing in a specific year, but that are not covered by the other covariates.

The TV audience for the lead-in program is another variable that is used frequently in studies on TV ratings. We do not use such a variable since it was impossible to find the data for all years and all countries in our dataset. We do not think that this is a major problem, as Danaher et al. (2011) demonstrated that, although a lead-in variable can improve the model fit, it does not help to improve the audience forecast. This is not too surprising, since the lead-in variable must itself be forecast in future time periods, thus increasing the variation of one of the predictor variables (Danaher et al., 2011, p. 1236). This is a useful research finding, as the complications that arise from having to forecast the rating for the previous program too can be avoided, and all that is needed is a model that uses information for the actual program.

The potential determinants of TV ratings fall into two categories: ex ante variables, i.e., covariates that are known prior to the race, such as the type of stage or the channel that will broadcast the stage; and ex post variables, i.e., covariates that only become known once the Tour de France is underway, such as suspense or the weather. Table 2 shows that, ignoring the year dummies, seven determinants are only available ex post: three of the competition related covariates and four of the country-specific covariates.

With all of the variables grouped appropriately, the regression equation for each country can be formalized as follows:

$$\begin{aligned} TV_{i,t} = & \beta_0 + \beta_1 TYPE_{i,t} + \beta_2 CALENDAR_{i,t} \\ & + \beta_3 BROADCAST_{i,t} + \beta_4 COMPETITION_{i,t}^{EA} \\ & + \beta_5 COMPETITION_{i,t}^{EP} + \beta_6 COUNTRY_{i,t}^{EA} \\ & + \beta_7 COUNTRY_{i,t}^{EP} + \varepsilon \end{aligned}$$

The dependent variable $TV_{i,t}$ represents the TV audience for Tour de France stage i (in absolute numbers) in year t . Independent variables are grouped by similar meanings in vectors of variables. For example, TYPE is the vector of all stage type covariates. Three groups of covariates (TYPE, CALENDAR, COMPETITION) are identical for all countries, while the BROADCAST and COUNTRY covariates differ between countries. All of the TYPE, CALENDAR and BROADCAST covariates are known ex ante, but a distinction has to be made between ex ante and ex post variables for the COMPETITION and COUNTRY covariates, hence the use of the superscripts EA and EP.

5. Results

5.1. Regression results for the individual TV markets

We estimated a total of twelve equations, since two regression analyses were performed for each of the TV markets. First, an explanatory model was tested for each market. The 'complete model' includes all ex ante and ex post covariates that are formalized in the regression equation presented above. This model is used to gain the maximal insight into the determinants of Tour de France TV viewing for the different countries. However, since the ultimate goal of this study is to establish an econometric model for predicting TV viewership for Tour de France stages, the forecasting model should only contain covariates that are available before the live broadcast of the first stage of the Tour de France. Thus, only the 19 ex ante covariates are included in the second model. This reduces the formalized regression equation for the 'a priori model' to:

$$TV_{i,t} = \beta_0 + \beta_1 TYPE_{i,t} + \beta_2 CALENDAR_{i,t} + \beta_3 BROADCAST_{i,t} + \beta_4 COMPETITION_{i,t}^{EA} + \beta_6 COUNTRY_{i,t}^{EA} + \varepsilon$$

The equations were estimated using multivariate ordinary least squares regressions. Scatter plots of the residuals and an outlier analysis indicated no violation of the basic regression assumptions. The results were also checked for multicollinearity. Because the variance inflation factors were below three in all equations, the results were judged sound.

The regression results are reported in Table 3. We begin by turning our attention to the complete models. The explanatory power of these models varies between 67% for the Netherlands and 86% for France. This is close to the results obtained in studies of TV audiences for international football games (Feddersen & Rott, 2011; Kim & Kim, 2012, report adjusted R^2 values of 87% to 90%), and in line with the results from studies on TV audiences for professional road cycling (Rodriguez et al., 2015; Van Reeth, 2013, report adjusted R^2 values of 68% to 72%). Three covariates, namely, 'high mountain stage', 'week 3 stage' and 'closing stage', are significant determinants of Tour de France TV audiences in all six markets. All of the other covariates are also significant in one or more countries, except for 'opening stage' and 'doping', which are not significant in any of the models.

The forecast models that include only the variables that are known ex ante explain between 38% (Wallonia) and 79% (France) of the variation in TV viewership. The loss in explanatory power from removing the ex post variables is about 20% for most TV markets, but is smaller for France (7%) and larger for Wallonia (35%). In the a priori models, again three covariates are significant in all six TV markets: 'closing stage', 'week 3 stage' and 'summit finish'. Two covariates are not significant in any of the a priori models: 'mountain time trial' and 'opening stage'. We also notice a couple of remarkable differences between countries. Only one of the six stage type variables is significant for the Netherlands, while five of the six are significant for Flanders. In Spain, only two of the seven calendar variables are

significant, while in Flanders, France and the Netherlands, six of the seven are. The 'individual time trial' covariate is highly significant in the models for Flanders, Denmark and Spain, but with opposite signs. Individual time trial stages decrease the TV viewership in Flanders, but increase the TV audiences in Denmark and Spain, revealing a clear difference in viewership preferences between these markets. We also find that all of the broadcast related covariates are highly significant in both the complete and a priori models. This underlines the importance of using control variables for the choice of channel when analyzing TV audiences.

5.2. Forecast accuracy

We used the results of our a priori models to predict the 2015, 2016 and 2017 Tour de France TV audiences in Flanders, Denmark, the Netherlands, Spain and Wallonia. Because of the limited dataset, only forecasts for 2017 were made for France. We evaluated the accuracy of these 16 forecasts by comparing the estimated TV audiences not only to the actual audiences, but also to naive forecasts that we developed for validating the predictive power of our models. Such naive forecasts predict the TV audiences for the next Tour de France using only the TV audiences of previous years, without any additional information. Although various different alternatives were tested, below we report only the results of the two best naive forecasts. In the simple naive model, the forecasts are based on an unweighted average of the previous TV audiences for the same stage over the past five years (first stage, second stage, and so on). In the semi-naive model, the forecasts are based on an unweighted average of the previous TV audiences for the same type of stage over the past five years (e.g., mountain stage, flat stage, weekend stage, etc.).

The six panels of Fig. 1 display the actual TV audiences versus the forecasts per country for all stages in 2017 (similar graphs for the 2015 and 2016 forecasts are available upon request). A visual inspection of the graphs indicates that the variation in TV audiences for the 21 Tour de France stages is predicted accurately for Flanders (Fig. 1a), Denmark (Fig. 1c), Spain (Fig. 1e) and France (Fig. 1f), with no systematic over- or underestimation. However, the forecasts appear to have been a bit too pessimistic for Wallonia (Fig. 1b) and a bit too optimistic for the Netherlands (Fig. 1d), although the general pattern of the forecasts follows the pattern of the actual TV audiences for these markets too. We also observe that the semi-naive forecasts are closer to our predictions than to the actual TV audiences in all figures, and that the semi-naive forecasts are usually only slightly further from the actual TV audiences than the model predictions.

Four metrics are used to evaluate the accuracy of our forecasts in more detail. The overall prediction error (OVPE) is defined as the difference between the forecast audience and the actual audience, averaged over all 21 stages, as a percentage of the actual audience. Since over- and underestimations cancel each other out with this metric, a second measure based on absolute percentage errors was computed. The mean absolute prediction error (MAPE) is defined as the average of the 21 per-stage absolute percentage differences between the forecast audience and the

Table 3

Regression results for the 2017 forecast models.

Variable	Belgium: Flanders (n = 357)		Belgium: Wallonia (n = 357)		Denmark (n = 309)		France (n = 91)		The Netherlands (n = 167)		Spain (n = 166)	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
(Constant)	430,777***	382,250***	249,900***	171,295***	1,364	274,269***	5,900,612***	2,882,886***	663,721***	608,768***	187,298**	644,637***
Stage type												
High mountain stage	107,637***	116,008***	52,464***	49,740***	37,965***	52,670***	941,835***	768,752***	77,697*	53,899	133,421*	270,785***
Low mountain stage	58,208***	65,132***	38,379***	49,405***	37,653**	60,032**	638,995***	332,631*	30,758	-68,186	53,764	-81,584
Summit finish	36,605***	48,484***	8,021	16,259**	32,954***	34,180**	179,193	355,785**	57,979	118,667**	215,049***	391,349***
Individual time trial	-47,583***	-45,173**	-1,596	-5,083	72,146***	83,908***	-374,323**	-258,163	48,551	40,181	92,672	258,1307**
Team time trial	-56,988**	-53,134*	-17,531	-286	31,924	19,952	-953,994***	-767,895**	-98,889	7,013	9,054	80,630
Mountain time trial	13,979	-7,061	30,718	38,067	112,040**	56,673	-315,832	-740,024	323,629*	145,973	-392,720	-254,089
Calendar												
Opening stage	29,950	28,912	-4,852	5	116	-14,114	247,305	199,154	42,412	60,775	-168,592	-232,458
Closing stage	150,499***	156,148***	32,516***	45,879***	208,218***	213,649***	1,331,700***	1,699,364***	593,666***	649,107***	-439,509***	-254,895*
Week 2 stage	30,583**	28,325*	12,835**	16,746**	39,311***	35,361**	238,165*	311,886**	89,468	121,893**	-30,006	-6,856
Week 3 stage	37,516**	54,524***	13,816*	28,059***	72,007***	82,693***	352,311**	657,138***	117,204*	180,259***	202,940***	338,182***
First Sunday stage	86,498***	76,529***	20,785**	18,576	33,309	42,497	859,725***	917,206***	187,590**	174,146*	-109,127	-82,747
Mid Saturday stage	60,894***	71,665***	18,322***	27,424***	70,330***	77,243***	725,997***	894,317***	196,916***	228,971***	39,608	173,132
Mid Sunday stage	151,915***	158,322***	44,698***	50,006***	136,482***	142,656***	1,268,306***	1,459,018***	457,040***	500,511***	15,490	102,466
Broadcast related												
Broadcast on smaller channel	-199,573***	-91,320***	-44,272***	-37,797***								
Broadcast on bigger channel					42,104*	-82,646***						
Specific broadcast effect												
Competition related												
Doping	-7,726		59		-2,151		-321,625		74,782		182,272	
Dominance	-3,685		-2,515		23,541***		-29,399		-45,595		-114,327	
Suspense	71,137***		16,654**		73,367***				235,844***		290,456***	
Country-specific												
National day	79,000***	90,305**	18,306	13,455			941,729***	941,260***				
Regional day	41,123*	17,164										
Local passage	160,645***	180,215***	42,165***	48,063***					473,689***	559,931***	125,325	27,661
National participants	-6,097		-4,768**		57,671***		-78,054**		-1,757			
National success	7,121		-1,767		49,606***		217,539		-1,904			
Local temperature	44,742***		7,387						107,062**			
Local rain	53,311***		19,177***						108,954***			
Year dummies	Significant		Significant		Significant		Significant		Significant		Significant	
Adjusted R ²	0.75	0.56	0.73	0.38	0.70	0.49	0.86	0.79	0.67	0.42	0.84	0.65
Standard error	66,049	87,659	30,407	45,840	62,634	81,608	348,954	428,609	172,068	227,894	254,734	373,044

Notes: (1) full model; (2) model with a priori variables only. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level. The results for 2016 and 2015 are available upon request.

Fig. 1a. Actual TV audiences versus forecasts for Tour de France stages in 2017 (Belgium –Flanders).

actual audience. This metric has been applied in several forecast studies, and we use the definition of Danaher et al. (2011, p. 1222). Our third metric is the *correlation coefficient* (CORR) between the forecast audience and the actual audience for all 21 stages. This measure provides additional information because it focuses on the relationship between the two variables, rather than on the difference in values. For example, if both CORR and OVPE are relatively high in the forecast, it is possible that the mechanisms that drive Tour de France TV viewership were identified correctly, but the magnitude of the impact was misjudged. In the opposite case, i.e., relatively low values of CORR and OVPE,

the forecast was probably a lucky shot rather than a valid analysis of the situation. The final metric is the *percentage of highly accurate forecasts* (PHAF), defined as predictions that differ from the actual audience by no more than 5%.

Table 4 reports the scores for the four forecast accuracy metrics. The results for the a priori model and the two naive models are compared, and the scores for the best performing model are presented in bold with a grey background. The OVPE of our a priori forecasts varies between a best score of a 0.7% underestimation for France in 2017, and a worst score of a 22.7% underestimation for the Netherlands in 2015. However, the OVPE of the a priori model is better

Fig. 1b. Actual TV audiences versus forecasts for Tour de France stages in 2017 (Belgium –Wallonia).

Fig. 1c. Actual TV audiences versus forecasts for Tour de France stages in 2017 (Denmark).

Fig. 1d. Actual TV audiences versus forecasts for Tour de France stages in 2017 (The Netherlands).

Fig. 1e. Actual TV audiences versus forecasts for Tour de France stages in 2017 (Spain).

Fig. 1f. Actual TV audiences versus forecasts for Tour de France stages in 2017 (France).

than the OVPE of the naive models even in the worst case. In fact, the a priori model performs best on this metric in nine of the 16 forecasts (56%), while the naive model has the lowest OVPE for four forecasts and the semi-naïve model for three.

The superiority of the forecasts from the a priori models over the naive forecasts is much more evident when the other metrics are used. In the a priori models, the MAPE varies between a best score of 10.1% for Flanders in 2016 and a worst score of 27.7% for Spain in 2016, while the corresponding MAPEs for the naive models are over 15% (Flanders, 2016) and over 65% (Spain, 2016). The a priori model performs best on this metric for 13 of the 16 forecasts (81%). The MAPE values that we find for our forecasts fit nicely with those reported in prior studies. In their survey, [Danaher et al. \(2011, p. 1222\)](#) concluded that the MAPE ranges between 12.1% and 45.7% across all TV forecasting studies. Our lowest MAPE score just beats the lowest score reported by [Danaher et al. \(2011\)](#), while, with the exception of the naive forecasts for Spain in 2016 and 2017, all of the MAPE scores in our study are smaller than 30, and thus, well below the upper limit reported for other studies.

For the a priori models, the correlation coefficients between the actual TV audiences and the forecasts vary between a minimum of 0.51 (the Netherlands in 2015) and a maximum of 0.94 (Spain in 2015). In 12 of the 16 forecasts (75%), the CORR values for the predictions based on the a priori model are higher than those for the naive models, while the differences in the four situations where the opposite is the case are small (0.08 at most).

The highest number of forecasts within 5% of the actual TV audiences was realized in the 2016 model for Wallonia: nine out of 21 (42.9%). In contrast, there were four cases (for the Netherlands and Spain in 2017, and for Wallonia and the Netherlands in 2015) where only two stages (9.5%) turned out to have highly accurate forecasts. In only four of the 16 forecasts do the naive models record a higher PHAF than the a priori models.

The results in [Table 4](#) also show that our a priori models clearly outperform the naive models on all four metrics in 2016 and 2017, while a more mixed picture emerges in 2015. The fact that the 2015 Tour de France was the best-watched Tour de France of the decade might explain why its audiences were more difficult to forecast. At the country

Table 4
Analysis of the forecast accuracy.

TV market	Model	2017				2016				2015				All years			
		OVPE	MAPE	CORR	PHAF	OVPE	MAPE	CORR	PHAF	OVPE	MAPE	CORR	PHAF	OVPE	MAPE	CORR	PHAF
Flanders	A priori	-2.7%	11.9%	0.77	28.6%	3.3%	10.1%	0.82	23.8%	-17.2%	17.2%	0.52	19.0%	-5.5%	13.1%	0.70	23.8%
	Naive	2.4%	15.1%	0.65	23.8%	9.6%	15.3%	0.68	14.3%	-11.9%	14.9%	0.31	19.0%	0.1%	15.1%	0.54	19.0%
	Semi-naive	3.1%	15.4%	0.73	19.0%	11.6%	15.6%	0.74	28.6%	-13.6%	15.1%	0.40	33.3%	0.4%	15.4%	0.62	27.0%
Wallonia	A priori	-13.7%	15.4%	0.77	14.3%	3.7%	10.5%	0.75	42.9%	-11.0%	15.1%	0.58	9.5%	-7.0%	13.7%	0.70	22.2%
	Naive	-10.6%	18.1%	0.44	4.8%	14.1%	18.8%	0.56	19.0%	-0.6%	15.1%	0.24	14.3%	0.9%	17.3%	0.41	12.7%
	Semi-naive	-10.2%	13.7%	0.69	9.5%	16.0%	18.2%	0.72	19.0%	-3.0%	12.6%	0.44	23.8%	1.0%	14.8%	0.61	17.5%
Denmark	A priori	6.9%	15.1%	0.85	19.0%	13.5%	21.3%	0.53	19.0%	-2.1%	17.3%	0.63	23.8%	6.1%	17.9%	0.67	20.6%
	Naive	7.4%	18.2%	0.80	23.8%	24.1%	26.5%	0.48	19.0%	16.7%	22.1%	0.71	4.8%	16.1%	22.3%	0.66	15.9%
	Semi-naive	6.8%	21.6%	0.51	19.0%	26.6%	29.0%	0.57	19.0%	17.0%	26.2%	0.32	19.0%	16.8%	25.6%	0.47	19.0%
The Netherlands	A priori	13.8%	20.4%	0.77	9.5%	13.4%	16.6%	0.77	23.8%	-22.7%	23.6%	0.51	9.5%	1.5%	20.2%	0.68	14.3%
	Naive	13.6%	21.7%	0.71	4.8%	16.0%	21.9%	0.70	14.3%	-27.6%	27.3%	0.44	4.8%	0.6%	23.6%	0.61	7.9%
	Semi-naive	13.5%	21.2%	0.63	9.5%	17.3%	26.5%	0.46	9.5%	-28.2%	28.9%	0.53	4.8%	0.8%	25.5%	0.54	7.9%
Spain	A priori	1.2%	24.4%	0.90	9.5%	16.4%	27.7%	0.90	14.3%	-15.2%	20.6%	0.94	14.3%	0.8%	24.2%	0.91	12.7%
	Naive	17.1%	54.7%	0.69	4.8%	30.3%	65.7%	0.52	9.5%	-2.0%	25.4%	0.67	14.3%	15.2%	48.6%	0.63	9.5%
	Semi-naive	18.1%	50.4%	0.92	9.5%	31.2%	67.3%	0.66	4.8%	-4.5%	29.4%	0.64	14.3%	14.9%	49.0%	0.74	9.5%
France	A priori	-0.7%	10.3%	0.93	38.1%	/	/	/	/	/	/	/	/	-0.7%	10.3%	0.93	38.1%
	Naive	1.3%	15.5%	0.70	28.6%	/	/	/	/	/	/	/	/	1.3%	15.5%	0.70	28.6%
	Semi-naive	1.3%	16.8%	0.76	19.0%	/	/	/	/	/	/	/	/	1.3%	16.8%	0.76	19.0%
All countries	A priori	0.8%	16.2%	0.83	19.8%	10.0%	17.2%	0.75	24.8%	-13.6%	18.8%	0.63	15.2%	-0.8%	17.3%	0.75	19.9%
	Naive	5.2%	23.9%	0.66	15.1%	18.8%	29.6%	0.59	15.2%	-5.1%	21.0%	0.47	11.4%	6.2%	24.8%	0.58	14.0%
	Semi-naive	5.4%	23.2%	0.71	14.3%	20.5%	31.3%	0.63	16.2%	-6.5%	22.4%	0.47	19.0%	6.4%	25.5%	0.61	16.4%

Notes: OVPE = overall prediction error, MAPE = mean absolute prediction error, CORR = correlation, PHAF = percentage of highly accurate forecasts. The best scores have grey backgrounds. In the naive model, the forecasts are based on an unweighted average of the previous TV audiences for the same stage over the past five years (first stage, second stage, and so on). In the semi-naive model, the forecasts are based on an unweighted average of the previous TV audiences for the same types of stages over the past five years (mountain stages, flat stages, time trial stages, weekend stages, weekday stages). The scores for the accuracy metrics in the "Average" row refer to the unweighted average of the individual scores of the six TV markets.

level, we see that, averaged over the years, our predictions outperform the forecasts from the naive models on all four metrics for Denmark, Spain and France, while the naive models perform better on the OVPE for Flanders, Wallonia and the Netherlands. The difference in OVPE is especially important in the case of Wallonia. This result may possibly be attributable to the strange results for the naive model in Wallonia in 2015, where both the lowest OVPE score (-0.6%) and the lowest CORR score (0.24) out of all naive forecasts were recorded. As was explained above, such a contradictory outcome is likely to be due to a lucky shot.

A final remarkable result is that the semi-naive model does not perform much better than the simple naive model. The forecasts from the semi-naive model do show a higher correlation with the actual TV audiences (0.61 versus 0.58), and there is a higher percentage of highly accurate forecasts (16.4% versus 14.0%), but the mean and absolute prediction errors are (marginally) smaller for the forecasts based on the simple naive model.

6. Discussion and conclusions

This paper has established a novel approach for the prediction of TV audiences for individual Tour de France stages. Our study has analyzed five countries that represent a global TV market of 150 million people, of which about six million watch the Tour de France. We learn from the results presented in Table 4 that the forecasts based on our a priori models clearly outperform the naive and semi-naive forecasts on all four metrics used. Averaged over the 336 individual stage forecasts, our overall prediction error is just 0.8% (versus about 6% for the naive models), while our mean absolute prediction error is equal to 17.3% (versus about 25% for the naive models). A total of 19.9% of the predictions (67 out of 336) were highly accurate (versus

about 15% for the naive models), and the overall correlation coefficient is 0.75 (versus about 0.60 for the naive models).

Several conclusions that are relevant to both media companies and to policy makers in sport can be drawn from our research findings. First, the relatively small difference in explanatory power between the complete model and the a priori model implies that a large part of the TV audience is determined by organizational variables; i.e., by characteristics of the race or the game that are entirely under the control of the event organizer. Competition-related variables, such as the competitive balance and outcome uncertainty, have only a minor impact on the global TV audience. In the case of the Tour de France, this means that accurate TV audience forecasts are possible by the end of October, eight months before the actual start of the race, when the profile of the next Tour de France is announced. Such information would be extremely useful in the price negotiations for commercial air time during live Tour de France broadcasts, for instance, which typically take place in the months leading up to the event.

Second, our findings have a huge economic value. Whether in euros or dollars, media companies worldwide spend billions on sports broadcasting rights deals, and sponsors and advertisers in turn pay similar amounts of money for TV commercials and airtime. Thus, since a change in TV audiences can be worth millions, reliable audience predictions are necessary for evaluating the effectiveness of this investment. Our results for the Tour de France support the conclusion by Kim and Kim (2012, p. 77) that the prediction of TV viewership for live broadcasts of sports events can be based on fairly simple models that include only basic factors such as viewing times and race characteristics. This finding is of considerable practical value to sports events organizers, media planners, sports rights companies, broadcasters, and advertisers, who would all benefit from a better understanding of the expected TV audiences.

Third, the broadcast channel related variables are highly significant in all of our models. The extreme importance of channel choice for TV audiences is relevant not only to media managers and programmers, but also to sports federations and sports event organizers, i.e., with respect to the sale of the TV broadcasting rights to their events. Allocating such media rights to the highest bidder might not always be optimal, and a trade-off must be made between the short-term money inflow from the sale of the rights and the long-term interest in the sport. While the Tour de France organizer ASO might get a much more lucrative deal from another French (private) broadcaster, it remains loyal to the public broadcaster France Télévisions because this guarantees that the race will have a much broader audience. In England, for instance, the BBC accounts for just 3% of sports broadcasts, but the public broadcaster accounts for 41% of all TV audiences for live sport (SportBusiness International, 2017).

Our analysis has some limitations, mainly relating to the nature of the data. We only had access to average TV audience data for complete broadcasts, not to TV audience data for specific parts of the coverage of a stage. Thus, it is impossible to forecast the evolution of TV audiences during the broadcast of a stage. Furthermore, live Tour de France broadcasts do not have the same duration in all countries. While football games, for instance, are broadcast in their entirety, cycling races usually only get partial coverage, and TV channels decide on the length of the broadcast individually. In some countries, only the final hour of a stage is broadcast, while in other countries the live broadcast of the same stage can last up to six hours. Thus, it was not possible to pool our data and we had to stick to country-specific estimates.

Our results for the 2015, 2016 and 2017 Tour de France TV audiences can be used for the further optimization of the models that are driving the predictions. For instance, we have not really explored yet what explains the observed differences in viewing behavior between countries. A better understanding of country-specific determinants will make forecasts for an increasing number of countries possible. Another research opportunity involves the development of a dynamic model that predicts TV audiences for Tour de France stages on a daily basis, i.e., on the eve of every broadcast of an individual stage. Such a dynamic model could take into account fresh information regarding race developments (e.g., on outcome uncertainty), as well as the most recent information on determinants that affect viewing behavior in general (e.g., the weather), and new information that results from the observed viewing behavior for earlier stages. In turn, this opens up the opportunity to develop a more dynamic pricing of commercial airtime on TV.

Nevertheless, forecasts will always be limited by variable characteristics that are largely unknown beforehand, such as the day's race developments, or characteristics of the race that can hardly be put into numbers, such as the quality of the racing commentary. However, our analysis does demonstrate that, in spite of these limitations, it is feasible to make accurate predictions of the TV audiences for individual Tour de France stages long before the race

actually takes place, and thus to determine better the economic value of the live broadcasts.

Acknowledgments

We thank two anonymous referees and the participants of seminars in Brussels (KU Leuven) and Düsseldorf (Arbeitskreis Ökonomie) for their useful comments and contributions.

References

- Danaher, P. J., & Dagger, T. S. (2012). Using a nested logit model to forecast television ratings. *International Journal of Forecasting*, 28, 607–622.
- Danaher, P. J., Dagger, T. S., & Smith, M. S. (2011). Forecasting television ratings. *International Journal of Forecasting*, 27, 1215–1240.
- Feddersen, A., & Rott, A. (2011). Determinants of demand for televised live football: features of the German national football team. *Journal of Sports Economics*, 12, 352–369.
- Forrest, D., Simmons, R., & Buraimo, B. (2005). Outcome uncertainty and the couch potato audience. *Scottish Journal of Political Economy*, 52, 641–661.
- Fukushima, Y., Yamasaki, T., & Aizawa, K. (2016). Audience ratings prediction of TV dramas based on the cast and their popularity. In *2016 IEEE second international conference on multimedia big data* (pp. 279–286).
- Hannas, B. J., & Goff, J. E. (2005). Inclined-plane model of the 2004 Tour de France. *European Journal of Physics*, 26, 251–259.
- Hunter, S. D., III, Chinta, R., Smith, S., Shamim, A., & Bawazir, A. (2016). Moneyball for TV: a model for forecasting the audience of new dramatic television series. *Studies in Media and Communication*, 4(2), 13–22.
- Kim, N., & Kim, Y. (2012). Predicting TV audience of international sporting events in Korea: the case of 2006 FIFA World Cup. *Journal of Economic Research*, 17, 77–88.
- Meyer, D., & Hyndman, R. (2006). The accuracy of television network rating forecasts: the effects of data aggregation and alternative models. *Model Assisted Statistics and Applications*, 1(3), 147–155.
- Nikolopoulos, K., Goodwin, P., Patelis, A., & Assimakopoulos, V. (2007). Forecasting with cue information: a comparison of multiple regression with alternative forecasting approaches. *European Journal of Operational Research*, 180, 354–368.
- Rodriguez, C., Pérez, L., Puente, V., & Rodriguez, P. (2015). The determinants of television audience for professional cycling: the case of Spain. *Journal of Sports Economics*, 16, 26–58.
- SportBusiness International. (2017). Free to air: BBC plots sports strategy for a changing world. Retrieved from <https://www.sportbusiness.com/sportbusiness-international/free-air-bbc-plots-sports-strategy-changing-world/>. (Accessed 25 2017).
- Van Reeth, D. (2013). TV demand for the Tour de France: the importance of stage characteristics versus outcome uncertainty, patriotism and doping. *International Journal of Sport Finance*, 8, 39–60.
- Van Reeth, D. (2015). TV viewing of road cycling races. In D. Van Reeth, & D. Larson (Eds.), *The economics of professional road cycling*. London: Springer.
- Weber, R. (2002). Methods to forecast television viewing patterns for target audiences. In A. Schorr, B. Campbell, & M. Schenk (Eds.), *Communication research in Europe and abroad – Challenges of the first decade*. Berlin: DeGruyter.
- Yeh, C. (2015). Online word-of-mouth as a predictor of television rating. *Online Information Review*, 39(6), 831–847.
- Zhang, Q., Wu, C., Qiao, H., & Fang, Y. (2016). Forecasting audience ratings of China's seasonal entertainment TV shows based on TEI@I methodology. *Systems Engineering – Theory and Practice*, 36(11), 2905–2914.

Daam Van Reeth is professor in economics at KU Leuven, Belgium. He teaches micro-economics and macro-economics at the Faculty of Economics and Business, Brussels campus. His primary research interests include TV viewership for sports, gender issues in sports, sports innovation, and doping. He has published several articles on road cycling and is the editor of the book *"The economics of professional road cycling"*.